

Philosophy and Medicine

**P&M126**

Michael Quante

# Personal Identity as a Principle of Biomedical Ethics

 Springer

# Philosophy and Medicine

Volume 126

## Founding Co-Editor

Stuart F. Spicker

## Senior Editor

H. Tristram Engelhardt, Jr., Department of Philosophy, Rice University,  
and Baylor College of Medicine, Houston, TX, USA

## Series Editor

Lisa M. Rasmussen, Department of Philosophy, University of North Carolina  
at Charlotte, Charlotte, NC, USA

## Assistant Editor

Jeffrey P. Bishop, Gnaegi Center for Health Care Ethics, Saint Louis University,  
St. Louis, MO, USA

## Editorial Board

George J. Agich, Department of Philosophy, Bowling Green State University,  
Bowling Green, OH, USA

Nicholas Capaldi, College of Business Administration, Loyola University,  
New Orleans, LA, USA

Edmund Erde, University of Medicine and Dentistry of New Jersey (Retired),  
Stratford, NJ, USA

Christopher Tollefsen, Department of Philosophy, University of South Carolina,  
Columbia, SC, USA

Kevin Wm. Wildes, S.J., President, Loyola University, New Orleans, LA, USA

More information about this series at <http://www.springer.com/series/6414>

Michael Quante

# Personal Identity as a Principle of Biomedical Ethics

 Springer

Michael Quante  
University of Münster  
Münster, Germany

ISSN 0376-7418

ISSN 2215-0080 (electronic)

Philosophy and Medicine

ISBN 978-3-319-56867-6

ISBN 978-3-319-56869-0 (eBook)

DOI 10.1007/978-3-319-56869-0

Library of Congress Control Number: 2017937750

© Springer International Publishing AG 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer International Publishing AG

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface to the English Edition

The German edition of this book elicited, upon its release 15 years ago, a favorable and at the same time critical echo. The favor consistently referred to the attempt of combining fundamental questions of practical and theoretical philosophy with central questions of biomedical ethics. The thus targeted level of analysis allowed on the one hand to differentiate between problems examined in biomedical ethics and current questions and advancements in the natural sciences and new possibilities of action made possible through technology, so that the considerations submitted here must not be viewed as out of date (leaving aside that one might find them wrong or implausible). On the other hand, it allows for the fulfillment of the requirement of relating the selected questions of biomedical ethics to a more general philosophical foundation, in crossing over to the discussion of this basis again and again. While doing so, as the reader will note, many central questions of practical and theoretical philosophy had to be addressed. The argumentation carried out here accepts premises from these areas without they themselves being exhaustively justified. The argumentation assumes certain burdens of proof that cannot be met in a single study. I have tried to fulfill some of these burdens in publications that have appeared in the last 15 years (now and again, I will refer to these in the following two sections). I am painfully aware of this inadequacy, but I tried in every instance to explicate the terms I used, to lay bare my premises, and, where detailed justifications or connected questions could not be dealt with, to identify the burdens of proof my argumentation entails.

A leading force in this study was the question of how far, if at all, and in what sense a conception of personal identity could be shown to be the principle of biomedical ethics. Due to this question, what needed to be clarified was whether assumptions regarding the diachronic identity of human persons are assigned an orientating or justifying function in everyday contexts or in philosophical considerations of biomedical ethics. From an evaluative point of view, it needed to be justified whether, how far, and in which sense a plausible philosophical conception of personal identity as the principle of biomedical ethics should be acknowledged and utilized. It is obvious that the central question not only lays the focus on the question of biomedical ethics discussed in this book but also calls for an explication of

the whole argumentation that is the basis of the conception of personal identity. In the first part of this introduction, I will roughly depict the structure of this book gained by these questions and concerns. Subsequently, I will, in the second part, address central criticism that has been brought forward against the proposal developed here. Furthermore, I would like to use this opportunity to address current research contributions that have a certain proximity to my thoughts concerning diachronic identity of human persons. Both will, I hope, contribute to a better understanding of the developed argumentation and facilitate a better categorization.

Shortly after the appearance of the German edition, I developed the wish to also make my contribution to the discussion available in English. Not only was it sensible from the point of view of overcoming the language barrier that is increasingly impeding the reception of German research contributions, but as the frame of reference of the developed argumentation in this study was analytical philosophy, especially the literature concerning personal autonomy on the one hand and the basis of biomedical ethics on the other, it made sense to address the target directly.

When David DeGrazia then published his book *Human Identity and Bioethics* in 2005, I decided to realize the plan of an English version of my book. Apart from some small differences concerning details and a fundamental difference regarding the ontology of organisms, we came to almost exactly the same results entirely independently of one another. The basic decision to differentiate strictly between the question of diachronic identity of the human individual on the one hand and the question regarding the narratively or biographically composed personality of human individuals on the other is to be found in both studies. Due to this, it is not particularly surprising that also the questions we chose regarding biomedical ethics discussed in both studies coincide to a great degree. Naturally, the concrete argumentation of two philosophers will differ in details, even if they have a similar basic conception and pursue a similar central question. As these differences are philosophically illuminating, I did not draw the conclusion from David DeGrazia's book that an English version of my own study has thus become redundant. On the contrary, it has encouraged me in my conviction and my plan to engage in the English debate with this English version of my study. That it has taken so long from the appearance of my book to reach this insight is due to many things that John Lennon called life: what happens to you while you are busy making other plans.

## I.

This book's goal is confirming that man's ability to lead a personal life and its forms of realization can and should play a role in biomedical ethics. The central thesis is that with the help of diachronic identity of human persons, we have an ethical principle of medium profundity, which de facto guides our everyday ethical intuition in many contexts and should also guide an adequate form of biomedical ethics. At the same time, this thesis aids the fundamentally skeptical thesis developed in this book. The philosophical question with regard to diachronic personal identity is not aimed

at a consistent phenomenon in the case of human persons and is thus misleading. The question must be broken down into four complexes of questions and respectively their answers; that being done, these can then be put into relation to one another. Thereby, a whole complex of philosophical explication of the personal life form of human individuals emerges that can do the complexity of the respective phenomenon justice. The four questions that must be kept separate in my study concern person-making characteristics that are needed for the analysis of personhood, the question of the narrative or biographical structure or personality (meant as the individual form the human individual gives his personality in life), as well as the synchronous unity and diachronic persistence conditions of human persons. The conception that is developed in its basic features in this book consists in answering the question relating to unity and persistence conditions with the help of the concept of human individual in the sense of a biologically specific organism and forgoing the concept of person here. The concept of person is then replaced in a second step by the two concepts of personhood and personality, which I then draw on to analyze the personal life form of human beings. In other words, in my overall conception, there are human individuals (organisms of a specific biological type) due to which they are recognized as having personhood, which they realize in their life in the form of personality.

Two fundamental objections which are based on misunderstandings have been brought forward against this conception: on the one hand, this is the *first* one; it has been criticized that with such a conception, the human person is naturalized and “person” reduced to a purely biological concept. But this is by no means the case. That the concept of the person, due to the characteristics that are named in this study, is not suitable for the answering of questions regarding synchronous unity and diachronic persistence is a thesis – that human individuals do not have personhood or personality or if so only in a naturalistic sense is an entirely different claim. Not only is this latter claim not made in the book, but it is also in fact implicitly rejected through the differentiation into the four questions. The aspects of personhood and personality and their respective features are part of our social life form and are to my mind as ontologically robust as those aspects of the human life form that can be captured in the natural sciences. Only if one assumes that these dimensions of human existence can be ontologically graded can there be any misunderstanding of the submitted conception. It is correct that the following argumentation regards the strict division of these two dimensions and the four referred-to questions as important. But this does not happen with the intention of eliminating one of the two dimensions or reducing one to the other. On the contrary, its goal is to philosophically explicate the complex internal connection between human- and personhood of the human individual.

On the other hand, this is the *second* fundamental objection; it is argued that my skeptical thesis expresses that I do not recognize the concept of person as having ontological or metaphysical dignity and do not believe it to be able to assist in answering basic metaphysical questions. The allegation amounts to the point just identified that ontological questions can only be answered along the lines of unity and persistence. But in my opinion, that is neither a necessary consequence nor are there good reasons for such a narrowing of ontology and metaphysics. The explica-



tion of personhood and personality in the sense of socially constituted aspects of our life form neither denies its existence nor its philosophical significance. *Or* this objection makes use of the limited conception of metaphysics that restricts the latter to an ontology using only the model of things and properties. That a human person cannot be captured adequately by such ontology or metaphysics is however a thesis that indeed is fundamental for my considerations. But from this only follows that such a narrow frame of such an ontology should not be accepted. This insight can be found, unfolded, and justified by such different philosophers as Hegel, William James, or the later Wittgenstein. It is true that the hereby-named philosophical tradition remains in the background in the book (the argumentation is already very complex and would have been thereby overloaded). It is unfounded that my skeptical thesis expresses a metaphysically or ontologically deflationary concept of person or indeed denies the ethical significance of the personal life form. For human existence, the opposite is true: leading the life of a person significantly matters! It is true however that the concept “person” is not suitable for the answering of certain ontological questions (as is the case the other way around for some concepts of the natural sciences that are unsuitable for the answering of other questions). But these questions (concerning synchronous unity or diachronic persistence) do not exhaust philosophical ontology, much less philosophical metaphysics.

The first chapter provides further distinctions, next to the differentiation of the four central questions that are needed for the explication of the whole conception and the answering of the four questions. The further path of the argumentation follows imperatively on my view from the just sketched basic setting of the course. I illustrate this argumentation in the fourth section of the first chapter, and it needs no repetition here. This is why I wish to pass on to the central objections raised against some of the theses in this book that do not concern the overall position of this book but the execution thereof.

## II.

The first fundamental *Weichenstellung* (in the sense of a strategic decision in direction) of this book’s conception was made in differentiating the classical question regarding personal identity into four distinct questions. The second *Weichenstellung* was made in showing that these four questions can be divided into two problem areas. The questions regarding synchronic unity and those regarding the diachronic persistence of human persons are answered through the use of the biologically understood concept of “human organism” from the epistemic-methodological point of view. The questions relating to being a human person are answered using the concepts “personhood” and “personality” from the epistemic-methodological point of view of a participant/agent.

For this reason, for the central question of this investigation into two epistemic-methodological perspectives, the following is the consequence: “personal identity” has a double role in biomedical ethics when being used as a justifying principle.

When understood as a synchronic entity or as the diachronic persistence of the human organism, *indirect* ethical consequences arise wherever ontological questions become ethically relevant. This is the case at the beginning of life and in the case of death; this function is indirect because the question concerning the beginning of life and those concerning death can be viewed as ontological problems. If, for instance, an ethical difference is acknowledged between abortion and contraception, ethical consequences arise from this criterion for the beginning of the existence of a human individual. And as long as the ethical status of a living, albeit dying, human is different to that of a corpse, this criterion will also have indirect ethical meaning in defining a human being's death.

Understood as personhood and personality, the principle of "personal identity" gains *direct* ethical relevance. This is de facto the case in our ethical practice in which the status of personhood, the integrity of personality, and also a human being's personal autonomy are acknowledged as ethically relevant. In every ethical (and metaethical) conception, which attributes relevance to being a human person, this aspect of our ethical practice can be philosophically justified. In this book, I explicate the connection between the narrative condition of the human personality and the principle of respect of autonomy. Beyond that, I show what the material content of this conception consists of, using the fundamental questions regarding the right to a self-determined death, the ethical bindingness of advance directives, and the possibility of justifying a certain form of paternalism in the context of medical action.

In the previous section, I rejected two fundamental objections, whose goal was the aforementioned *Weichenstellung*, on the grounds of them being misinterpretations of my conception. In this section, I wish to address some further fundamental objections that cannot simply be removed by unmasking them as misunderstandings. They can be placed in two groups: the first point of criticism regards the first *Weichenstellung*. The other objection is aimed, in contrast, at assumptions that I lay claim to in the context of answering the questions concerning synchronic unity and diachronic persistence of human organisms. The two groups are not entirely independent of one another, but to better identify the burden of proof, it is helpful to distinguish the two. I will begin with the latter group, because its objections are to constitute the argumentative points, which undercut my choice of course. In their monograph, *Menschliches Leben*, Sebastian Knell and Marcel Weber depict a biological approach that I outline in this book, and they then subject it to an in-depth critique. First of all, I wish to state that their depiction is factually correct and fair in every respect. They repeatedly say that my argument was not particularly developed and that central conceptions remain underdetermined. Saying that, I neither, and this is true, address in detail the concept of functional integration, which is central to my conception, nor do I explicate of which type the biological laws that I presume are to be. Their central point of criticism, if I understand Knell and Weber correctly, can be summed up thus: for an analysis of the concept of a biological organism, the concept of functional integration is indispensable. This though, and both authors agree with me on this point, leaves room for a range of different grades of being functionally integrated. This range is a distinguishing feature of the biological world, which causes a blurring of both the borders to the existence of individual organisms and –

here Knell and Weber refer to John Dupre's analyses in *The Disorder of Things* – those of species. This however causes, such is the conclusion Knell and Weber come to, the ontological condition of the biological to be such that the ontological burden, which my conception of diachronic persistence is to carry, becomes too heavy. The authors delineate their position clearly and unambiguously: whether a biological organism belongs to a specific species depends on pragmatic criteria, as much as the decision regarding certain criteria for the beginning and the ending of the existence of an individual organism. Thereby, however, my first *Weichenstellung* becomes misleading, because it – as I wish to put it – presupposes a distinction between the biological and the social, which is not tenable. In other words, “The concept ‘life’ does not refer to a natural kind; it is permeated by practical and evaluative intricacies of meaning” (p. 54, my translation). I have no retort to bring forward against the fact that, up to this point, there only exists a rough sketch of a thesis. But it is possible however to offer an explication of why I do not believe Knell and Weber's argumentation to be inescapable (leaving aside the fact that the burden of proof ascribed to me has not yet been fulfilled). This attestation is however helpful for the understanding of my own conception, as it helps in identifying some of the premises of Knell and Weber's criticism which I do not share. Firstly, they work with a strong conception of essentialism and natural types that I make no claim on. Neither am I obliged to assume that the determination of essential properties or natural types is wholly independent of epistemic efforts nor does it follow from the vagueness of the biological species that the borders between the biological and the social need to be dissolved. If reality is vague, then we must indeed draw on pragmatic criteria, if we wish to use our concepts unambiguously in all cases. It does not follow from this that there are no unambiguous cases and the normative aspects that go along with the use of concepts and with the constructing of theories can be, as I do, clearly acknowledged. From this, it cannot be deduced that the difference between the interest in cognition and methodological approach is dissipated. In fact, the opposite is true as a central argument for my first *Weichenstellung* lays in a criteria-guided and, therefore, pragmatic catalogue of requirements for the analysis of the conditions for the diachronic persistence of organisms. If there is vagueness in my suggested theory frame concerning the exact definition of the beginning or end of the existence of a (human) organism, then I can allow for such isolated cases or even types of cases. As I acknowledge with regard to the question of ascription of potentiality, these do not work in all contexts. To deduce from this fact however that they are useless in all cases is neither necessarily the case nor plausible. Hard cases make bad laws!

My impression is that Knell and Weber's critical argumentation rests on the assumption that my suggested first *Weichenstellung* is only sustainable if it is supported by a metaphysical realism and a concept of natural types and one of biological organisms that is free of any vagueness. Such a conception, this is at least how I understand Knell and Weber, can only carry the required metaphysical weight, if it forgoes any pragmatic reasons for the use of concepts, epistemic achievements, and theory building in its development. It is obvious that one should not accept these strong requirements. However, I have never done so (although it can be attributed to my very brief sketch of my conception, as it invites critics to fill in the gaps with their own

preconditions). With John Dupre, I am of the opinion that one can develop a conception of species that can function without the essentialism assumed by Knell and Weber.

The rejection of the criticism brought forward by Knell and Weber remains a defensive one, because it rests on rejecting certain preconditions, without which their objection cannot be formulated. The conceptual space thus held open must be filled by my own contribution of a theory, so as to justify the conception I have brought forward. As long as this has not been achieved, the objection raised by Knell and Weber remains aimed at problems and gaps in my own conception on a very fundamental level. This is different in the case of the criticism directed at my justification of the criterion for brain death. In contrast to the suggested criterion for the beginning of existence, my defense of the criterion for brain death was criticized in the context of organ donation. This debate, which is just now flaring up in Germany again, is one I do not wish to go into here. Instead, I will leave it at the indication that arguments that are brought forward against these two specific theses (the criterion for the beginning of existence and the criterion for the end of existence of a human organism) are not automatically objections against this book's conception of persistence of biological individuals. Neither is it the case that these criteria can only be justified on the grounds of my conception nor that my conception collapses, if the suggested criterion for the beginning and ending of existence cannot be supported. As I admitted above, this would be different for the objections brought forward by Knell and Weber. The burdens of proof, on which I went into detail on in the study submitted relating to a conception of the (human) organism, have as yet not been delivered (this would require a further book). I have named the reasons for which I am optimistic, despite the challenges to my conception that they can still be delivered. Anything further remains to be seen in the course of further development.

In view of the conception of personal life forms of the human being and in a narrower sense bioethical questions that are addressed in this book, the situation looks different. On the one hand, I have tried to fulfill those burdens of proof in my book *Person* which appeared in German in 2007 (an English edition of the monograph is in preparation at the moment, so that I wish to delay the discussion of the objections that need to be addressed to a future occasion). On the other hand, I have participated in the last two decades in bioethical debates on assisted suicide, euthanasia, and advance directives and was part of many publications on medical paternalism. The objections brought forward against my theses in these contexts do not, analogously to those against the criteria for the beginning and ending of existence, seep through to my suggestion of the personal life form of the human being. For this reason, they need not be addressed here.

### III.

In the introduction to the German edition to this book, I thank colleagues who supported me during that time in Münster. As my gratitude is still as current as ever, I wish to repeat my thanks.

The years in which I was able to study and do research at the philosophical seminar in Münster were informed by a cooperative atmosphere, a lively exchange of thoughts, and a constant willingness from all to have conversations and discussions. The productive environment ensured that the execution of my research never became a lonely undertaking.

Special thanks go to my academic teacher Ludwig Siep. Without his compassionate help and assistance, I could not have put my project into effect. Many friends and colleagues commented on this book in its different stages and in doing so freed it of many deficiencies: Johann Ach, Kurt Bayertz, Tom Beauchamp, Dieter Birnbacher, Jürgen Blühdorn, Birger Brinkmeier, Bernward Gesang, Christoph Halbig, John Harris, Martina Herrmann, Angela Kallhoff, Jussi Kotkavirta, Helga Kuhse, Arto Laitinen, Johanna Macher, Barbara Merker, Sibille Mischer, Rosemarie Rheinwald, Peter Rohs, Erzsébet Rózsa, Hans-Christoph Schmidt am Busch, Dieter Sturma, Christian Suhm, Manfred Wetzel, Marcus Willaschek, and Andreas Vieth. To all of them, I wish to extend my thanks.

This study was developed as a *Habilitation* paper at the Department of History/Philosophy at the Westfälische Wilhelms-Universität of Münster. When writing this introduction in September of 2001, I knew I would have to leave my *Heimat*-University. At the time, I could not have foreseen that I would be returning in 2009 as my teacher Ludwig Siep's successor. It gives me all the more pleasure to be able to say that the philosophical seminar's character has not changed, although many individuals and my role have of course done so.

This only leaves me to thank those who significantly contributed to the realization of the English edition. First and foremost, I wish to thank David Schweikard and Faith Puleston Jones for the translation of the manuscript. I also wish to thank Anna Maj Blundell for the editorial revision of this text and the incorporation of my updates. My thanks also go to Tris Engelhardt for accepting my book into his series and thereby giving me the opportunity of fulfilling my wish of finally realizing an English edition. I too wish to thank the anonymous reviewers for their helpful suggestions for improvement. And finally, I thank the Springer Publishing Company for attending to this project over the last years with precisely the right amount of pressure and patience.

Münster, Germany  
February, 2017

Michael Quante

# Contents

<b>1</b>	<b>Introduction</b> .....	1
1.1	The Concept of Person: Kinds of Usage.....	3
1.2	Personal Identity: Four Different Questions .....	4
1.3	Divergent Epistemological-Methodological Approaches .....	7
1.4	The Structure of the Study .....	9
<b>2</b>	<b>Human Persistence</b> .....	11
2.1	Easy Solutions to a Complex Problem?.....	13
2.1.1	The Charge of Inadequacy.....	14
2.1.2	With the ‘Weapons’ of the Simple View .....	15
2.1.3	The Problem .....	16
2.1.4	Easy Alternatives? .....	20
2.2	Human Persistence: A Complex View .....	23
2.2.1	Causality: First Group of Presuppositions .....	24
2.2.2	Sortal Dependence: Second Group of Presuppositions.....	25
2.2.3	The Biological Approach.....	29
2.2.4	Conclusion.....	32
<b>3</b>	<b>The Beginning of Life</b> .....	35
3.1	Targets.....	36
3.1.1	Preconditions .....	36
3.1.2	Targets .....	36
3.2	The Beginning of Life: Human Persistence (I).....	37
3.2.1	When Did This Human Organism Begin to Exist? .....	37
3.2.2	Objections.....	44
3.2.3	The Graduality of Integration: The Hard Problem.....	56
3.3	Persistence, Potentiality and Personhood .....	58
3.3.1	The Bioethical ‘Guillotine’ .....	59
3.3.2	The Debate on the Potentiality Argument .....	61

3.4	The Ethical Relevance of the Beginning of Life.....	68
3.4.1	Contraception Vs. Abortion: Possible Vs. Potential Persons .....	68
3.4.2	Gene Therapy, Persistence and Personality .....	71
<b>4</b>	<b>Death</b> .....	77
4.1	Targets.....	78
4.2	The Fight over the Definition of Death: Some Clarifying Considerations.....	79
4.2.1	The Background .....	79
4.2.2	Criteria of Death: Three Candidates.....	79
4.2.3	Three Levels of the Problem .....	80
4.2.4	Three Strategies.....	81
4.2.5	Further Differences.....	82
4.3	Death: Human Persistence (II).....	83
4.3.1	Definition.....	83
4.3.2	Explanations and Modifications .....	84
4.3.3	The Criterion of Death for Human Organisms.....	86
4.4	Criticism of Alternative Approaches .....	87
4.4.1	The Moral Strategy.....	88
4.4.2	The Subject of Human Death .....	91
4.5	Objections and Open Questions.....	98
4.5.1	Objections.....	98
4.5.2	Open Questions .....	103
4.6	Persistence and Biomedical Ethics: A Brief Conclusion.....	104
<b>5</b>	<b>Personality and Autonomy</b> .....	105
5.1	Targets.....	105
5.2	Personality .....	106
5.2.1	Is Identity Important?.....	106
5.2.2	Personality and the Question of the Identity of the Person Over Time.....	111
5.2.3	The Double Meaning of Identification .....	114
5.3	Personality and Autonomy .....	116
5.3.1	Personality as Precondition of Autonomy.....	117
5.3.2	Autonomy as Precondition of Personality?.....	125
5.4	The Principle of Respect for Autonomy .....	131
5.4.1	Beauchamp's Theory of Autonomy of Agency.....	132
5.4.2	The Tacit Preconditions for Autonomy of Agency .....	139
5.4.3	The Principle of Respect for Autonomy.....	145
<b>6</b>	<b>Dying Autonomously</b> .....	151
6.1	Targets and Preliminary Clarifications.....	151
6.2	The Value of Life, the Personality and the 'Right' to Die .....	155
6.2.1	The Argument.....	155
6.2.2	The Value of Life According to John Harris .....	156
6.2.3	Voluntary Euthanasia and Personal Autonomy .....	159
6.3	The Categorical Objection.....	162

6.3.1	The Objection .....	162
6.3.2	Subtle Distinctions?.....	163
6.3.3	Sanctity of Life and Plural Society: First Conclusion.....	169
6.4	Gradualist Objections .....	170
6.4.1	Four Objections .....	172
6.4.2	The Social Context of Voluntary Euthanasia: Second Conclusion .....	178
6.5	Is There a Duty to Kill? .....	179
6.5.1	A Duty to Kill?.....	179
6.5.2	Who Has Veto Rights?.....	180
6.5.3	Dying Autonomously: A Closing Conclusion .....	181
<b>7</b>	<b>Extended Autonomy</b> .....	<b>183</b>
7.1	Targets.....	183
7.2	Personal Identity, Autonomy and the Passage of Time.....	184
7.2.1	The Perspectivity of the Terminology .....	185
7.2.2	Types of Case .....	186
7.3	Limits to Extended Autonomy? .....	190
7.3.1	The Problem of Implementation.....	191
7.3.2	The “Do no harm!” Objection.....	193
7.3.3	The Assumption of Non-identity .....	194
7.4	Conflicts Within Personal Autonomy .....	196
7.4.1	Ulysses Contracts .....	196
7.4.2	Paternalism or Self-Paternalism? .....	198
7.5	The Ethical Relevance of Biographical Identity .....	199
7.5.1	Self-Commitment and Coherence .....	199
7.5.2	Restoration Versus Installation .....	201
<b>8</b>	<b>Medical Paternalism</b> .....	<b>203</b>
8.1	Targets of Proof.....	203
8.2	A Definition of Medical Paternalism .....	204
8.2.1	The Definition .....	205
8.2.2	Elucidation of the Individual Conditions .....	206
8.3	Strong and Weak Paternalism? Beauchamp’s Criticism of Feinberg .....	212
8.3.1	Feinberg’s Distinction .....	212
8.3.2	Beauchamp’s Criticism .....	213
8.3.3	Three Objections .....	216
8.3.4	A Possible Basis for Beauchamp’s Argumentation.....	218
8.4	Personality as Reason for Justification .....	220
8.4.1	Personality-Based Justifications of Paternalistic Action .....	221
8.4.2	Anticipated Consent .....	224
8.5	From Paternalism to Cooperation – A Brief Conclusion.....	229
<b>9</b>	<b>The Interlacing of Persistence and Personality</b> .....	<b>233</b>
	<b>Bibliography</b> .....	<b>237</b>



# Chapter 1

## Introduction

If one were to look for *the* specific property of human beings, the following definition would be a promising candidate: human beings differ from other living creatures through their endeavor to lead a *personal* life. The aspiration to lead one's own life and go one's own way, to give one's actions a "personal" touch or develop a "personal" style, are two of the numerous ways in which the fundamental aim of human life, to have one's own personality and develop one's own character, is articulated. The central value attributed to personal life in our culture is moreover expressed in diverse terms, which either themselves represent widely accepted values, or are lined up as ethical claims, because they are the conditions on which a personal life can be led: self-fulfillment and originality are examples of the former; freedom, autonomy or integrity of the latter. Not only as far as articulation, legitimation and defense of individuality – one of the characteristic features of modernity – are concerned, but also in such contexts as those in which the exceptional moral status of human beings in comparison to other *Lebensformen*<sup>1</sup> is to be specified or justified, this usually occurs with recourse to the human personality:

Whatever else we are, we are persons; and it seems likely that this fact will prove fundamental to the justification of the strong moral status that most of us want for ourselves and those we care about (Warren 1997, p. 90).

In view of this eminent relevance of personhood for the human *Lebensform*, it is not surprising that the concept of person is of central consideration in both theoretical and practical philosophy. In theoretical philosophy it is above all (the) different philosophical conceptions of the identity of persons over time that play a substantial role, whereas the discussion in practical philosophy is mainly concerned with an analysis of the properties and capacities on the basis of which an entity becomes a person. In addition, there are numerous aspects of the concept of person within the context of biomedical ethics. Thus, the different philosophical conceptions of the

---

<sup>1</sup>"Lebensform", used here in the Wittgensteinian sense, means our human way of being in the world (including first and second nature).

identity of persons over time are readdressed in the dispute over the criteria of the beginning of life and of death. And as far as questions of human dignity, the protection (sanctity) of life or the prohibition of instrumentalization are concerned, recourse to the concept of person is regularly taken in order to justify the exceptional ethical status of human life.

In this study I will show that the human capacity to lead a personal life can play a central role in biomedical ethics.<sup>2</sup> As explained in the preface my core thesis is that a complex conception of the *identity of human persons over time* provides biomedical ethics with a principle that factually guides our everyday ethical intuitions and should also guide adequate biomedical ethics.<sup>3</sup> By *principle* I understand – going along with the conceptions developed by T.L. Beauchamp and J.F. Childress – a principle of medium scope and medium profundity (cf. Beauchamp and Childress 1994). “Medium scope” means, on the one hand, that such a principle does not belong to the highest principles of a specific ethical theory, but can, more exactly, be developed on the basis of such varied background ethics as a kind of consensus principle. On the other hand, such a principle is open in content and requires specification through more specific rules according to a given context. Medium profundity is pertinent to these principles because each of them is valid *prima facie*, but can be restricted by other principles or further ethically relevant aspects.<sup>4</sup>

My conviction that a principle of personal identity can be rendered productive in a plausible way for adequate biomedical ethics is, however, accompanied by a skeptical thesis. With this, I relate in one respect to Birnbacher’s critical assessment regarding the use of the designation of person in the current debates in biomedical ethics (cf. Birnbacher 1997). On the other hand, I am convinced that, on the basis of a differentiated theoretical conception, one can and should adhere to a suitable and meaningful usage of the designation of person in form of a conception of personhood (as explained in Sect. 1.1 and developed throughout this study).

In contrast to most philosophers involved in the discussion on personal identity, I believe that the philosophical issue of the “identity of the person over time” is not aimed at a uniform phenomenon and is therefore not well-defined. For this reason, before I can justify my core thesis and give an overview of my line of argumentation, more kinds of usage of the concept of person – four different ones aimed at the questions of the identity of person and three diverging epistemological-methodological approaches – must be differentiated.

---

<sup>2</sup>“Biomedical ethics” does not refer to a special set of ethics with its own principles, but rather, to a specific topic area. Since the focus of my considerations is directed towards the “Principle of personal identity”, the following is only concerned with a section of that topic area.

<sup>3</sup>My argumentation is therefore committed to an internal conception that is specified within an ethical practice.

<sup>4</sup>Cf. There to in more detail Quante and Vieth (2002) and Vieth and Quante (2010).

## 1.1 The Concept of Person: Kinds of Usage

The concept of person is used in three different ways.<sup>5</sup> Firstly, in some theories one finds purely *descriptive* usage, whereby a person is understood as an entity that is capable of certain achievements or exhibits certain features (so-called “person-making characteristics”). In attributing the status of being a person, the evaluative participant perspective is in fact indispensable. However, within the frame of descriptive usage the issue as to whether the status of being a person is tied to an exceptional normative status remains open. But this does not preclude that, along with the properties and capacities accompanying the status of being a person, there are such that have particular weight in ethical terms. However, refined arguments are necessary to justify this relevance within the context of the descriptive usage of the concept of person; this justificatory function does not pertain to the concept of person itself. Secondly, the concept of person is used purely *normatively* in some theories, to denote the specific ethical status of an entity. On the basis of which properties and capacities an entity is attributed this specific status is not directly answered by the concept of person used in this way. In its normative usage, the definition indicates that one needs an independent, possibly complex criterion on the grounds of which an exceptional moral status can be justified. The purely normatively used concept of person indicates her status but does not justify it. Thirdly, there are often uses that can be defined as a combination of the first two, whereby, along with the concept of person, a certain set of properties and capacities descriptively, and simultaneously an exceptional moral status is ascribed to an entity.

Even if these three ways of using the concept of person can be differentiated, yet *similarities* between the three usages can be detected: (a) In all three usages the concept of person indicates a certain status of an entity: its *personhood*. The concept of person is therefore a predicative identification, even if it seems to be something else in statements such as “this person is insufferable”. But even here, “person” only serves to establish the reference due to its predicative meaning. (b) In all three usages, the status of being a person does not count as an unanalyzable property of an entity, but refers to the conditions of personhood. (c) Finally, it is these properties and capacities that make an entity a person, that justify the specific ethical status of a person.

The *difference* between these three usages in respect of the relation of properties and capacities on the one hand and specific moral status on the other can be described as follows: In the purely descriptive usage, the properties and capacities, on the

---

<sup>5</sup>In a further usage, “person” in the sense of “human individual” is used, whereby “human” is meant purely biologically rather than in a possibly evaluative or normative sense. In this usage, the designation “person” neither refers to certain properties or capabilities of persons, nor is reference being made to the ethical aspirations tied to the status of being a person. This usage of the person designation is masked out in the following, as it is obviously not tied to the denotation of this term and can be replaced by the expression “human individual”. This procedure, which I favor in this study, avoids the danger of an unnoticed or unaccounted for usage of the evaluative and normative dimensions of the term person; cf. Quante (2012) for a detailed analysis of this.

basis of which an entity is a person, are picked out without addressing the topic of their ethical relevance. In the purely normative usage, the specific ethical status of persons is picked out without establishing which of the person-specific properties or capacities justify this status. In contrast, in the combined usage, the exceptional ethical status of persons is actually justified by those properties and capacities that constitute personhood (or by a person-specific subset thereof). In this case, one imputes an ethical theory that assigns a specific ethical weight to these properties and capacities.

## 1.2 Personal Identity: Four Different Questions

One of the reasons some debates about personal identity continue to run side by side oblivious to one another is that behind this label (at least) four different questions are hidden (cf. Quante 2007a). This is because, for one thing, they are – rightly or wrongly – made use of when dealing with these different issues concerning the concept of identity, and for another, there is a factual relation between the different issues. The four questions, which shall be differentiated, are the following:

- Conditions of Personhood
- Unity (at a point in time)
- Persistence (over a period of time)
- Personality

(four questions)

*Conditions of Personhood* Behind this question lies the problem as to which properties and capacities the classification of an entity as a person is based. What should be on the list of so-called “person-making characteristics”? Since the concept of personhood will not be discussed during the course of this study, but will only be handled alongside the issues related to the question of the identity of a person over time, a few short remarks will have to suffice here.<sup>6</sup> Personhood is a complex property that is attributed to an individual with specific properties and capacities at her disposal. This includes being the subject of mental episodes, being able to relate to herself in the first person form, a (rudimentary) consciousness of time and (at least rudimentary) knowledge of her own existence over time. In addition, a person must possess logical and instrumental rationality, be capable of communication (in the broadest sense) and be able to recognize other individuals as persons. Some of the

---

<sup>6</sup>I don’t have to enter into the discussion which person-making characteristics have to or should be accepted when analyzing the meaning of “personhood” since the systematically decisive move in my argument is to analyze the persistence of a human person without recourse to personhood at all on the one hand. On the other hand I suggest an evaluative conception of personality which is compatible both with nearly all conceptions of personhood suggested in the philosophical literature and, even more important, with our daily understanding of personhood.

properties and capacities necessary for personhood are gradualizable, for example, one can be more or less rational. The status of being a person is generally conferred when properties (and capacities) are present to an adequate degree and habitual enough in an individual (i.e. they are practiced by her). If, for example, a person is not able to make decisions rationally enough in only one specific field, this does not lead to her status of being a person to be disallowed in general. Personhood is also a so-called “threshold concept”, i.e. all individuals who exhibit the required properties and capacities to an adequate degree are persons in the same sense. It is not, for instance, the case that a more rational individual is accorded personhood to a higher or more intensive degree than an individual whose capacity for logical or instrumental rationality is less pronounced. The status of being a person is also recognized when an individual is at some time disposed to practice these capacities, and has practiced these capacities at prior times in her existence. Thus a sleeping person or a patient who is currently unconscious due to general anesthesia is still a person even though she is for the time being unable to actualize the required capacities. In contrast, the potential to acquire properties and capacities sufficient for personhood does not suffice to make an actual person. For example, a healthy and normally developed 1-year-old human infant is a potential person in the sense that it is part of the normal development for members of her species to develop properties and capacities sufficient for personhood during the course of her life. However, this infant is not a person, since neither the properties and capacities nor the required dispositions pertain to her.<sup>7</sup> The list of properties and capacities necessary for personhood is probably just as unlikely to be completely fixed once and for all as are the combinations and degrees of the properties and capacities that are sufficient for the status of a person to be recognized. Some properties and capacities might depend on a certain cultural context, at least as far as their significance within the overall list is concerned, even if the above mentioned properties and capacities can in the main be counted inarguably as minimal requirements. The degree to which one must be able to attribute these properties and capacities to an individual in order to regard him as a person is likewise not invariable, but defined within a cultural setting. For example, it depends on the complexity of a society what degree of logical and instrumental rationality is considered necessary. According to this conception, “personhood” is a thick concept containing a variable list of criteria. It includes both descriptive and evaluative aspects. The latter play a twofold role: for one thing, because the concrete contents of personhood are not independent of societal ideals and behavioral norms, and for another, personhood contains essential evaluative aspects because persons are constituted through evaluative relations to themselves as well as recognitive relationships which are manifested in the individual personality of a person.

---

<sup>7</sup>For the consolidation of personhood in the species-specific properties and a (partial) justification of potentiality arguments established thereon, cf. Chap. 3. For even if the potential of becoming a person and the status of being a person are differentiated with the above implementations, this does not mean that the former is irrelevant from the ethical aspect.

*Unity* Although the issue of personal identity is usually understood as the question of identity over time, sometimes questions are (also) dealt with under this heading that apply to the unity of the person at a point in time (cf. Bayne 2010). These questions are firstly, whether the person is to be considered a psycho-physical unity, or as an entity ‘composed’ of body and soul. Secondly, there is the question of which conditions must be fulfilled so that it is possible to speak of exactly *one* person at a (particular) point in time. Thirdly, this problem area also includes the question of how the unity of self-consciousness is to be analyzed from the philosophical point of view.<sup>8</sup> In this inquiry I will limit myself to the question of the identity of the person over time (see Chap. 2).

*Persistence* If one asks how the identity of the person over time is to be understood philosophically, or which conditions must be fulfilled so that one can say of a person at a point in time that she is identical with a person at a different point in time, then one inquires into the requirements for the identity of the person over time, or the persistence conditions of persons.

*Personality* Above all in pedagogical or personality-psychological and socio-psychological literature one finds a usage of “identity” in connection with persons which has left its mark in philosophy, e.g. in existentialism, the theories of the Frankfurt school, or in the works of Charles Taylor. For instance, if, when speaking of national or cultural identity with regards to persons, someone wants to facilitate identity finding, bemoan loss of identity, or argue identity crises, he is not thinking of numerical identity. The identity of person is in fact taken to be the evaluative self-understanding and self-image of this individual, his way of comporting himself towards social demands so that what he is and wants to be is manifested therein. In the following, the concept of personality will be reserved for those phenomena that characterize the social life of persons.<sup>9</sup> Thus, this concept represents the respective embodiment of being a person in the biography of the respective person, for her evaluative attitudes, her emotional and habitual make-up, her convictions, wants, life plans, self-image of her capacities, her strengths and weaknesses. In other words: personality indicates who or what a person is and wants to be.

*The Connection Between the Four Questions* Although in the literature the concept of identity comes into use in the treatment of all the problem areas differentiated just now, in the following it shall be used exclusively in the logical sense (numerical identity). It is probably clear why one can forego the concept of identity within the frame of the question of the requirements of personality. Moreover, to avoid terminological confusion, it makes sense to forego the evaluative identity concept in the debate on the phenomena outlined by the concept of personality and instead to

---

<sup>8</sup>For dualists who regard the body-correlated subject as the person, the third question is immediately posed. But even then, when one comprehends the person as a psycho-physical unit, this problem arises if the unity of self-consciousness is regarded at a point in time as a necessary condition for awarding personality to an entity.

<sup>9</sup>Cf. the contributions in Friedrich and Zichy (2014) for an overview concerning the concept and conceptions of personality.

speak of the self-image, the self-understanding or in general of the evaluative relationship to oneself. But what about the relation between identity and unity, or identity and persistence? Both cases can be understood as special treatment cases of numerical identity on space-time (or at least temporally) expanded entities. The question of the unity of a person then becomes the question as to the material requirements which must be fulfilled so that one can speak at a point in time of *one* person. The question of the persistence or identity of the person over time can then be formulated correspondingly as a question regarding the material requirements which must be fulfilled so that one can say of *one* person that she exists at two different points in time. In the philosophical dispute of the latter issue stimulated by John Locke, two questions took center stage: Can there even be conditions to be fulfilled over time for the identity of a person, which (themselves) can be formulated without recourse to this fact? The dispute over this first issue is being held by followers of the simple view, who answer this question in the negative, and representatives of the complex strategy, who give a positive answer. If one affirms the first question, a subsequent question arises: What are these fulfillment conditions? The various versions of the complex view argue over the correct answer to this question (see Chap. 2).

### 1.3 Divergent Epistemological-Methodological Approaches

A further aspect, in respect of which the various theoretical approaches are to be differentiated, concerns the underlying epistemological-methodological approach (cf. Quante 2014a). It is expedient to differentiate between three perspectives.<sup>10</sup>

The *observer perspective* is to be understood as the stance of the disengaged, purely theoretically interested observer, whose epistemic achievements are not constitutive for the observed phenomenon. In this approach, only causal and functional designations of observable phenomena are analyzed. The treatment of the problem of personal identity within this purely descriptive, causal-explanatory stance does not differ from the treatment of the unity and persistence issues of other observable objects (e.g. ships, landscapes or trees).

In contrast to the observer perspective, the analysis of personal identity within the *participant perspective* contains essential elements which are due to the concerted knowledge of a shared *Lebensform*. The evaluative self-image of persons (personality) and the rationality of the mental become determinable from here and are partly constituted by the participant's presupposition of meaning. In every process of understanding, the assumption of a shared *Lebensform* is operant and

---

<sup>10</sup>The parlance of "perspectives" and the notion that these refer to epistemological-methodological approaches should leave open which ontological conclusions are to be drawn for the respective issues. An analysis of personal identity based on the Cartesian perspective is, however, mainly distinguishable in that ontological conclusions are drawn from the epistemic characteristics of self-consciousness.

inescapable. The parlance about shared *Lebensform* should be understood as including as minimal requirements such basal facts as embodiment, sentience, neediness and basal rationality and also concern for one's own wellbeing. At the other end of the spectrum one finds, as maximal requirements, aspects arising from a highly specialized *Lebensform* within a culture. The claim that the participant perspective is essentially evaluative also requires explanation and justification. On the one hand, this assumption does not preclude that descriptive aspects are also central to this perspective. Only the non-eliminability of evaluative aspects is claimed. This thesis can be justified in two ways: on the one hand, from the participant perspective, one works with rationality assumptions and expectations. But the concept of rationality is itself normative: following rules of rationality does not mean being subject to regularity, but rather, that the normative claim based on behavioral expectations must be fulfilled. On the other hand, the presuppositions of meaning are also oriented on evaluations in that e.g. certain events or conditions are rated as positive, but others as negative. Even if the interpreter tries not to take his own evaluations as the benchmark for the interpretations, he must nevertheless firstly presume to know what it means to evaluate events and states of affairs, and secondly, try to trace the evaluations from the viewpoint of the interpreted individual. Without a general knowledge of this evaluative dimension of the processes, the sequence of events in question cannot be understood as a sequence of actions.<sup>11</sup>

The *Cartesian perspective* arrives at its approach of the issue of personal identity through the special epistemic situation of self-consciousness. Persons have first-person knowledge about their own temporality and a privileged status in respect of their propositional alignment. The Cartesian perspective has, on the one hand, the purely theoretical stance in common with the observer perspective, but in contrast, the Cartesian perspective claims that personal identity can only be sufficiently grasped with recourse to the first-person perspective and the accessible facts thereof. The central role of the "inner perspective" of the person links the Cartesian perspective with the participant perspective. However, in contrast to the Cartesian perspective, this is understood as a moment of social and evaluative self-understanding in the latter. From the Cartesian perspective, self-consciousness – and herein it is related to the observer perspective – is conceived as a purely descriptively ascertainable phenomenon. On this view, the phenomenon of self-consciousness is to be grasped in a way that is oriented on perception. In other words, according to the conception upon which the Cartesian perspective is based, self-consciousness is about a "theoretical" phenomenon in the sense that it is taken to be a special form of knowledge. In contrast, self-consciousness, though also accepted as an essential element in the participant perspective, is at the same time a perspective based on the

---

<sup>11</sup> The differentiation between the observer and participant perspective takes preference here over the difference between first-person and third-person perspectives frequently used in the literature, because the latter does not encompass the differences relevant to my considerations. The Cartesian perspective demonstrates that the first-person perspective does not automatically include the evaluative aspects of the observer perspective, so that first-person and third-person perspectives do not coincide.



assumption that self-consciousness cannot be embraced free of its social and evaluative aspects: even if in this perspective the “theoretical aspects” of self-consciousness are well recognized, these are held to be inseparable from the “practical” aspects.

## 1.4 The Structure of the Study

The argumentation that will be unfolded in the following chapters rests on a strict separation of the issues of personal identity in the sense of persistence on the one hand and personality on the other. This procedure is the expression and consequence of my skeptical thesis, according to which the philosophical issue of the identity of the person over time is not aimed at a uniform phenomenon and is therefore not well-defined. The central question of this study: whether “personal identity” is possibly a central justificatory resource for biomedical ethics, can only be answered successfully if the issue of persistence is distinguished from that of personal identity in the sense of personality, and a separate solution applied to each problem area.

In the first section of the *second* chapter, a suggestion for philosophical analysis of personal identity, which is committed to the Cartesian perspective, will be rejected. According to this simple view, the identity of a person over time is a fact that is not further reducible and which reveals itself essentially in the first-person perspective. In contrast, at the second stage, with the biological approach, a complex analysis committed to the observer perspective will be proposed. Unlike other complex analyses, this biological approach waives the concept of person and draws only on the purely biologically understood concept of human being.

Notwithstanding this the analysis of human persistence proposed in the second chapter has indirect ethical relevance. The discussion about questions of the beginning of life (*third* chapter) and the death of the human individual (*fourth* chapter) will develop the biological approach and will highlight the clarifying function of the differentiation between persistence and personality for the partly convoluted bio-ethical debates on beginning of life and the death of human beings. Alongside this critical function, a further indirect ethical relevance of the here proposed analysis of human persistence is the fact that it provides the basis for controlled usage of potentiality arguments.

The following chapters are dedicated to the concept of personality and the thesis that personal identity is, in this sense, a central principle of biomedical ethics, which factually guides our ethical intuitions and should also be used explicitly as a justificatory resource in adequate biomedical ethics. So as to reach this target, in the first step of the *fifth* chapter, the conception of personal identity as personality proposed here will be unfurled. Since the principle of respect for autonomy is indisputably of central significance in current biomedical ethics, in the second step I will show that an adequate concept of autonomy cannot get by without recourse to the biographical dimension of personality.

In the subsequent chapters, on the basis of the discussion of the ethical questions on the ethical permissibility of voluntary euthanasia (*sixth* chapter), the binding

character of advance directives (*seventh* chapter) and the substantiation of medical paternalism (*eighth* chapter) will, for one thing, further define the content of the concept of personality proposed here. For another, the analysis of these questions shows that ideas about the specific form in which humans live their lives as persons factually guide – and indeed should guide – our ethical intuitions as regards the permissibility of voluntary euthanasia and the binding character of advance directives. Over and beyond this, the explicit integration of the concept of personality in our notion of autonomy proves to be helpful in identifying certain cases of medical paternalism as ethically justified interventions, if they ultimately serve to conserve or restore the personality of the patient. However, this justificatory function of the principle of personality for biomedical ethics presupposes that no absolute or categorical requirement is tied to the status of being a person. Only then, when the concept of person is freed of such excessive demands, can the justificatory function it has qua “personality” be disclosed for biomedical ethics. Only then will it be able to continue to play a central role in an ethical theory adequate for a secular and pluralistic society (cf. Quante 2014b; Willems 2016).

The strategy employed in this study has its price from the ethical point of view, in that personhood can no longer be detached from all ethical considerations as an absolute or categorical principle. From the theoretical standpoint, the strict differentiation between personal identity in the sense of persistence on the one hand and the personality on the other leads to a further problem: they are both indisputably interwoven in our everyday perspective; and clearly there are manifold relations between persistence and personality which must be analyzed philosophically. The questions raised here cannot, however, be dealt with in this study, and are only sketched in the form of a forecast in the *ninth* chapter.

## Chapter 2

# Human Persistence

Both the various ways of applying the concept of person and various epistemological-methodological approaches can be found in the context of the issue which has been discussed in an intense philosophical debate ever since Locke added a chapter about the identity of persons in the second edition of his *Essay Concerning Human Understanding*: the question concerning the identity of persons over time.<sup>1</sup> Locke's proposal provoked critical reactions associated with the names Leibniz, Butler and Reid. In the second half of the twentieth century, decisively initiated by the works of Wiggins and Williams, a widespread discussion of this issue developed within analytical philosophy. Initially, the dispute was carried out between supporters of a psychological criterion developed following Locke's memory criterion and supporters of a body criterion which was quickly further developed into a brain criterion. But soon philosophers such as Chisholm and Swinburne were participating in the debate, taking the whole discussion to be misguided because it was based on the premise that there could be an informative empirical criterion for personal identity through time. Thus a discussion thread was resuscitated that formulated anew the protests of Leibniz, Butler and Reid against Locke's proposal. In the course of this the various thought experiments to be found in the literature are used to show that every informative empirical criterion for personal identity through time is bound to lead to unacceptable consequences. Hence, two fundamentally different notions meet head-on; and even within both camps the theories exhibit in part serious differences.<sup>2</sup> However, the different proposals can be broken down into simple and

---

<sup>1</sup> Within the context of the consideration of the simple view, the parlance of the identity of a person over time should leave open the possibility that this concerns a special phenomenon which does not permit analysis according to the general pattern of the persistence of concrete, space-time existing entities.

<sup>2</sup> For the development of this discussion and the central thought experiments central to the discussion cf. Quante 1999a; for standard definitions of the various identity criteria cf. Noonan 1991, Chap. 1.

complex theories. The characteristic feature of the simple theories is the following set of assumptions:

The identity of the person over time is not reducible to empirically observable relations. Such empirical criteria have only epistemological functions. Synchronic and unity relations over time that can be observed exclusively from the first-person perspective are constitutive for the identity of the person through time. Thus the identity of the person is a simple (unanalyzable) fact and essentially tied to the first-person perspective.

(Characteristics of simple views)

In contrast, complex theories can be characterized by the following assumptions:

The identity of persons through time is analyzable and reducible in the sense that they are constituted through empirically observable continuity relations.<sup>3</sup> The identity of a person over time is a complex application of persistence, i.e. the identity of space-time expanded entities through time, and not essentially tied to the first-person perspective.

(Properties of complex views)

The classification of one of the types of theory as simple should not be understood as though there were no further justifications in these theories on the identity of the person but should in fact emphasize that personal identity over time is, according to this type of theory, an irreducible, simple fact: Personal identity can be explicated, but cannot be analyzed without recourse to this fact. The classification of the other type of theory as complex instead of using the definition ‘reductionist’ as is customary in the literature is expedient, because the latter could give the impression that a complex view could necessarily challenge or ‘explain away’ personal identity over time. This is not the case: What is contested is merely the existence of a not further analyzable special fact, as assumed in the simple view.<sup>4</sup>

Complex answers would seem to be *prima facie* counterintuitive, since they contravene intuitions that arise out of first-person experience.<sup>5</sup> In contrast, simple positions are convincing at first glance because they are based on first-person access to the identity of the person over time. In the debate, supporters of the complex view feel obliged to reveal and safeguard their approach through elaborate argumentation, so as to combat the semblance of implausibility. However, defenders of the simple view usually react by attempting to prove that the complex view has counterintuitive consequences. To this end they call on basic intuitions stemming from first-person experience, without actually expounding or explaining them further. The simple view appears to be adequately justified merely by proof of the

---

<sup>3</sup>The skeptical possibility that there is no continuity below observability is put aside in this study (cf. Nozick 1981, p. 35 on this).

<sup>4</sup>One option to be differentiated from the simple and complex view consists in arguing an elimination thesis. Thus, e.g. Unger concludes in his early work on the vagueness of the person designation that there is neither person nor personality (Unger 1979a,b).

<sup>5</sup>For this reason, within the complex camp between ‘reductionist’ approaches, which analyze the intuitive understanding of personal identity over time philosophically, and ‘revisionary’ approaches can be differentiated (cf. Nida-Rümelin 2006). The latter assume that a suitable theory of personal identity over time must, as opposed to the intuitive everyday image, be revisionary if it is to be philosophically satisfying.

contra-intuitiveness of the complex view. From this allocation of the burdens of proof and the usage of such fundamental intuitions it ensures that the basic assumptions of the simple view are seldom deployed. To remedy this deficit, the line of argumentation will be reversed in the following and it will begin with a discussion of the basic assumptions of the simple view (Sect. 2.1). This will show that these theories are not capable of analyzing personal identity through time satisfactorily. Following that, I will outline the biological approach as the complex theory I propose (Sect. 2.2).

## 2.1 Easy Solutions to a Complex Problem?

Taking up a formulation of Butler's, Chisholm differentiated between a strict and philosophical application of the identity concept for persons and a loose and commonplace use for other space-time expanded entities as regards identity through time.<sup>6</sup> This differentiation can be called the basic idea of the simple view.

While one could speak of artifacts such as ships (Chisholm's example) or natural objects such as trees (Butler's example) as having identity through time in a loose way, and not demand that all parts of the ship or tree must be the same at two points in time, this would be a different matter in the case of persons. Besides, as Butler and Reid emphasize, any attempt to philosophically prove the identity of a person over time by specifying empirical criteria is destined to fail, because all these criteria are bound to be epistemologically weaker than the immediate evidence of the fact itself that can be found in self-consciousness. These objections, directed historically against Locke's theory, are today being directed in the same way by the representatives of simple views against complex theories.

If one understands the difference focused on by the simple view in an epistemological way, as Butler and Reid also did, then the evidence of one's own identity respectively as given in self-consciousness becomes decisive. This first-person knowledge is connected with a certainty that is in principle inaccessible through any complex analysis based on empirically observable relations. Thus, the simple view joins with the Cartesian perspective and the assumption that a basis of indisputable or perfect knowledge is revealed in self-consciousness. In this way, both the main criticism of the simple view of complex analyses and the arguments made use of therein become comprehensible.

---

<sup>6</sup>On this cf. Leibniz (1958, p. 85 ff.), Butler (1836, pp. 251–257), Reid (1983, pp. 212–218) and the reuptake in Chisholm (1969, 1970a, 1971a).

### 2.1.1 *The Charge of Inadequacy*

Both supporters of the simple view and supporters of the complex view concede equally that theories based on empirically observable relations cannot eliminate the logical possibility<sup>7</sup> of the ontological indeterminacy of a person over time that has been illustrated frequently in various thought experiments. The ontologically undetermined cases are those in which, according to the complex theory, there is no fact that is independent of conventions or human evaluations as far as the identity of a person over time is concerned. For one thing, the graduality of empirical relations leads to the eventuality of borderline cases: which physical or psychological changes are still compatible with the continuing existence of a person, and which are not, cannot be determined unequivocally. For another thing, the indeterminacy also occurs through the possibility of so-called fission scenarios,<sup>8</sup> i.e. such cases in which at  $t_1$  there are two candidates, B and C, who, in terms of the criteria demanded by the respective complex theory, have the same relation to a person A at  $t_0$ .

Complex theories cannot logically exclude these possibilities, and must recognize that, as regards the issues of graduality, fission and conflicting persistence criteria, the question of the identity of a person over time is, in the end, a matter of convention in borderline cases. This answer is deemed inadequate by the simple view for two reasons. For one thing, the evidence in self-consciousness, which is constitutive for the identity of a person through time, is not a matter of convention. For another, it would be unacceptable that in this way the basis of our ethical practice of attribution and evaluation of actions as expressed in praise and blame, reward and punishment, would not depend on facts but rest on conventional decisions. What is more, it would be inappropriate in a gradually waxing and waning continuum to want to pin such a grave difference from the ethical viewpoint on a conventionally fixed position.

In their criticism of the fact that complex views permit indeterminate cases while having only inadequate solutions to offer, the simple views rely on singularities of the first-person perspective that are manifested in personal memories and the anticipation of future own mental states. Together with the synchronic unity, i.e. unity of self-consciousness at the present point in time given for the present self-consciousness, this unity relation lasting across a certain time span, which is accessible in the first-person perspective, constitute the simple fact of personal identity.<sup>9</sup>

---

<sup>7</sup>The philosophical usefulness of thought experiments shall be conceded here for argument's sake.

<sup>8</sup>I am here ignoring the other case of the fusion of two persons discussed in the literature, as it does not raise any additional problems for the arguments developed in the following.

<sup>9</sup>This criticism presumes 'Only-X-And-Y-Principle' which implies that the answer to the question as to whether X is identical to Y may only depend on such factors that apply exclusively to X and Y. The strategy of excluding cases of duplication by definition is called the 'Ad hoc answer' by Nida-Rümelin (2006). Within the frame of the consideration of the biological approach it will, however, be shown that this answer is supported or motivated by other theoretical assumptions and is therefore not ad hoc.

### 2.1.2 *With the ‘Weapons’ of the Simple View*

In the first-person perspective, a person possesses identity through time through personal memories (past) and anticipations (future).

*Memories* When it comes to memories in the context of personal identity, this refers almost exclusively to personal memories, i.e. memories of having done or experienced something at an earlier point in time. For the simple view, memories are decisive in two ways: For one thing, they demonstrate in a first-person way the fact that a person existing now must already have existed at an earlier point in time. For another, this proves that the criterion suggested by Locke, whereby memories are constitutive for the identity of a person through time, is circular. If I remember flying over the Alps yesterday, or going to my hotel, then, according to the simple view, two factors are indubitably certain: first the fact that it is *my* memory and secondly, the fact that it *was* me who experienced the flight turbulence or tried everything to make myself understood at the reception. The content of my memory does not consist merely of the fact that I am conscious of it now. It also includes ‘how-it-was-to-make-this-experience’ (to-have-performed-this-action) that is present to me in first-person form. On the basis of this analysis, the memory criterion for personal identity is accused of distorting the facts. Speaking about a mental episode as *my* memory presumes that *I* am *identical* with the person who had that experience or carried out that action.

*Anticipation* The inadequacy of the complex view is also proven to the supporters of the simple view in that it is incompatible with the ability of a person to anticipate future mental states as her own. If, after one kind of chaotic series of events or another, the situation arises that it is not ontologically defined whether A at  $t_0$  is identical to B at  $t_1$ , according to the complex view, then A should not be in a position to anticipate the mental state of B as his own. But precisely this seems to be possible. According to the simple view, in anticipation, a person pre-empts her own future mental state. This first-person fact does not permit either gradualization or epistemologically undecidable or ontologically undecided cases.

In its argumentation, which is essentially bound to the first-person perspective, the simple view draws on two principles, which can be called the ‘principle of criterion-free self-reference in self-consciousness’ and the ‘principle of the primacy of self-attribution’.

*The Principle of Criterion-Free Self-Reference in Self-Consciousness* the first-person reference in self-consciousness is taken to be characterized by the fact that its self-reference does not need identity criteria or a prior identification. Somebody referring to himself as ‘I’ has not, in a first step, identified an entity with himself by applying identity criteria, in order to then label them as himself. The much avowed

immunity to false reference is taken to be due to precisely this fact, namely that self-reference in self-consciousness has nothing to do with identification.<sup>10</sup>

*The Principle of the Primacy of Self-Attribution* This principle says that the first-person self-attribution being manifested in ‘my mental state’ is more fundamental than the use of identity criteria of personal identity. A mental state is not identified by me in my self-consciousness first as a state belonging to me and then called ‘my own’ by me. On the contrary, one of the irreducible factors of self-consciousness is that the ‘mineness’ (Heidegger’s *Jemeinigkeit*) of actual mental states is immediately obvious. Sure enough, a person can be mistaken as regards the type identity of a self-conscious state and, for instance, interpret a feeling of jealousy as warranted indignation (or indeed wishful thinking as a well-founded argument), but she cannot be mistaken as regards the question of whether this actual self-conscious state is her own or not. An antecedent application of identity criteria is not only unnecessary for answering this question, it is impossible.<sup>11</sup>

From the viewpoint of the supporters of the simple view, memories, anticipations and the two principles just mentioned indicate both that the identity of the person over time is essentially of a first-person nature; and that it is neither gradually or ontologically indetermined, nor constituted through extrinsic factors or conventions of one kind or another. Therefore any attempt to justify this simple fact via empirically observable relations is doomed to failure. However, the question in the following will be whether this analysis of the simple view is really adequate.

### 2.1.3 *The Problem*

According to the analysis of the semantics of ‘I’ suggested by Kaplan, the actual situation of usage determines the semantic assignment of this indexical expression where it is being used and not merely mentioned.<sup>12</sup> The speaker cannot refer to anything other than himself in the actual situation of usage, either in modal or such intentional contexts in which a speaker attributes a propositional stance to another speaker. Expressed in a technical slogan: ‘I’ always has the greatest scope. (Castañeda 1982, p. 65) In philosophical explanations of self-consciousness, the bondage of self to the present, i.e. the nunc centricity of self-consciousness, has been hinted at many times.<sup>13</sup> The unique epistemic relations in self-consciousness,

---

<sup>10</sup>This feature of first-person attitudes is also accepted by many representatives of complex theories (cf. e.g. Shoemaker 1963, 1996 or Perry 1979, 1983) and will not be questioned in the following. Such an analysis has, among others, the advantage that it can do justice to the special epistemic conditions in self-consciousness without having to champion the implausible thesis that ‘I’ is not a referential expression (viz. Anscombe 1975).

<sup>11</sup>To date, the most comprehensive and clearest analysis of intuitions and principles based on the simple view was provided by Nida-Rümelin (1997, 2006).

<sup>12</sup>The following considerations ensue from the analysis by Kaplan (1989).

<sup>13</sup>Cf. e.g. Foster (1979) and the theory developed in continuance from German Idealism by Rohs (1996, 1998, Chaps. 2 and 3).



that are frequently emphasized regarding both criteria-free self-reference and evidence of ‘my’, apply only to the self-attribution of current first-person mental states such as ‘I am in pain now’ or ‘I believe that p’. Who thinks that he believes that p, knows that he believes that p. And who thinks that he is now in pain knows that he is now thinking that.<sup>14</sup> This can be called the ‘I-here-now’ structure of self-consciousness.

If these views about semantics and the nature of consciousness are taken as a basis, then both the analysis of personal memory and that of anticipation suggested by the simple view run into a dilemma. According to the findings just illustrated, a *memory* is an actual first-person mental state in which a person remembers having done or experienced something at an earlier point in time. But if in the sentence ‘I remember eating an ice cream yesterday’ the ‘I’, firstly, always has the widest scope, and if, secondly, the ‘I-here-now’ structure of self-consciousness is taken into consideration, then the sentence must be analyzed as follows: I (here-now) utter that I (here-now) remember that I (here-now) ate an ice cream yesterday. Thus, there is no direct reference to an ‘I’ of yesterday, but rather, I attribute to myself here and now the property of having eaten an ice cream yesterday. From within the first-person perspective there can be no guarantee that *this* self-attribution pertains. The semantic finding of the direct reference of ‘I’ is indeed consistent with the simple view’s ontological assumption that there is an identical entity at the two points in time, but it does not provide any additional support for this thesis. Taking the truth of this self-attribution as guaranteed would amount to stretching the area in which the unique epistemic relations of self-consciousness apply in exactly one case beyond the area of actual first-person episodes. Since this precisely concerns the controversial case of identity of a person over time, such a procedure (cf. Bermúdez 2012 as an example) would only be ad hoc.<sup>15</sup> Moreover, this procedure is open to two more objections. *Firstly*, an analysis of memory restricted to the first-person perspective does not get along with the fact that memories have to be caused in a suitable way. As every déjà vu experience shows, some perceptions ‘feel’ like memories. So as to differentiate these phenomena from real memories, one must posit a suitable cause as a further condition. However, this cannot be grasped in the first-person perspective. If the proponent of the simple view makes a case for how it was to have had this experience or carried out that action was being experienced in the memory thereof, this raises the *second* objection: type and token are being confused. Because mental episodes are concrete, dateable entities, the actual memory of how it was to have eaten an ice cream yesterday cannot be a repetition of this past

---

<sup>14</sup>This is consistent with the person in question neither believing p nor really being in pain, even if it admittedly always involves a certain amount of effort to construct plausible examples for that kind of situation.

<sup>15</sup>Evans (1991, p. 213 ff., cf. especially note 19) also reaches this conclusion. Hamilton’s criticism (1995, p. 343 f.) of these considerations misses its target because, firstly, it interprets ‘remember’ as a successful verb, and secondly, on this basis concludes that there is therefore a logical guarantee for identity over time. Evans does not have to argue that in the case of a correct usage of ‘remember’ identity over time is present (as Hamilton assumes). The moot point is what comprises this identity.

event (independently of whether ‘I’ refers to actual or past self).<sup>16</sup> The knowledge of what it is like to carry out such actions or to be the subject of such experiences is absorbed into the content of the current memory. The past, concrete (token) experience is not present in the current situation.

Therefore in the case of remembering there is no indisputable first-person reference to the respective ‘former self’ and her past self-conscious mental episodes. This impression can only arise if the special features of ‘I’ and the nunc-centric constitution of self-consciousness are disregarded and a type-token-confusion has occurred. This illusion is facilitated through the fact that in the real world the additional conditions (according to the complex view) necessary to qualify a current mental episode as a memory are normally given. These truth conditions for memories, which can be defined in a complex view, cannot be analyzed in the simple view.

In the case of *anticipation*, the simple view is ultimately no better off either. For one thing, the representative of the simple view cannot analyze anticipations such that a current, present self is referring to herself or himself at a future point in time. Such imputed synchronism of a present and a future self cannot be asserted consistently within a framework of the modal time definitions, essential for self-consciousness, of ‘past’, ‘present’ and ‘future’, and the ‘temporal becoming’ essential to these definitions (cf. Quante 2017a). Thus, the ‘weapon’ of anticipation can only serve as an illustration of the intuition that the ‘mineness’ (Heidegger’s *Jemeinigkeit*) of the self-consciousness is in each case indubitably evident and does not apply any identity criteria. This is where the principle of the primacy of self-attribution comes into play. Bearing in mind the results we have arrived at so far, anticipation of a moment in the future must be understood as a current mental episode in which a self attributes to itself the property of being the subject of a certain mental episode in the future, in a first-person way. It is therefore necessary to differentiate two principles here: the principle of the primacy of *actual* self-attribution and the principle of the primacy of *diachronic* self-attribution. In view of the above considerations, the latter cannot be sustained: for one reason, because there can be no presence of a future self, and for another, because it is based on a type-token-confusion.<sup>17</sup> The anticipation of a future mental episode is a different individual from the mental episode which will be experienced by a self in the future (independently of whether I will be this self or not).

The plausibility of the principle of the primacy of self-attribution stems solely from the epistemic features of the present self-consciousness and cannot be extended

---

<sup>16</sup>Two alternatives are conceivable: For one thing, one can comprehend mental events as abstract entities that are not space-time datable individual things. For another, one can try to conceive them as universals that are numerically identically instantiated in various space-time places. In his dispute with Davidson’s event conception, Chisholm seems to have in mind this (cf. Davidson 1982 and Chisholm 1970b, 1971b, 1985). Maybe a systematic motif of Chisholm can be accounted for in the above presented context, since he supports the simple view.

<sup>17</sup>The problem of the principle of the primacy of actual self-attribution will not be further expounded in the following. A consideration of this principle leads to the issue of whether the synchronic unity of self can be analyzed purely internally from the first-person perspective, or only with recourse to external factors.

to include diachronic conditions. Thus, the attempt by the simple view to underpin philosophically the *prima facie* existing counter-intuitivity of the fact that complex views must permit ontologically indetermined cases of personal identity over time, ultimately fails. Complex views of personal identity over time can allow both the principle of criterion-free self-reference and the principle of present self-attribution, because these concessions only pertain to the synchronic unity of the person.<sup>18</sup>

The simple view reacts to this result with two objections: Its defenders raise the charge that the restriction of the principles of criterion-free self-attribution and the primacy of self-attribution to the synchronic unity of person is a disallowed abstraction. And they claim that the data of first-person phenomenology show that in self-consciousness we do indeed experience a temporally extended self. Both objections have a case. The restriction of the present unity of person in self-consciousness to one point in time is an abstraction, since in reality this experience is always temporally extended. In phenomena such as listening to a melody or reading a philosophical text, the unity of self is present in a first-person way over a temporal interval. Moreover, the self that is missed by Hume and therewith the synchronic unity of the I in the awareness of being able to experience two different mental conditions simultaneously, is likewise ascertainable in the first person (see e.g. Chisholm 1994, p. 198, Foster 1979, p. 172 ff., Swinburne 1986, p. 155). To be sure, it is notoriously controversial which ontological conclusions should be drawn from these phenomenal experience data: does one have to assume as the basis for this (both) simultaneous and successive unity in self-consciousness, or can one ‘construct’ this self philosophically out of momentary atomic self-conscious episodes? However, a proponent of the complex view can make a concession to the simple view at this point as regards the question of the identity of a person over time, by taking the special first-person form of reality of this identity over time as given. For in the end nothing is achieved thereby, as far as the question of the identity of the person over time is concerned.

The problem posed for the simple view ensues from the fact that the identity of a human person over time always includes phases without self-conscious first-person episodes. For instance, it quite obviously belongs to our understanding of personal identity over time that after waking up from a coma, no new person begins to exist, but rather, the person who already existed before the coma has regained consciousness (above all when she continues to be in command of her memories and faculties). This difficulty in satisfactorily integrating any interruptions in the

---

<sup>18</sup>In view of the question of personal identity over time that is of interest here, this concession, is – as we shall see shortly – *possible*, but it is not necessary. For one thing, it should be remembered that ontological conclusions, in the sense that a person is a *res cogitans* set apart from her body, cannot inevitably be drawn from the admission of epistemic particularities. On the other hand, the admission of the epistemic particularities in self-consciousness is not tied to the admission that the Cartesian perspective is suitable for the analysis of these phenomena. Furthermore, in respect of personality disorders, there is good reason to doubt the exclusivity of the first-person perspective in determining the synchronic unity of a person (and with it the principle of the primacy of actual self-attribution); cf. Gunnarsson (2010), Clarke (2013) and the papers in Hughes et al. (2006) for a detailed discussion.

current self-conscious stream of consciousness during the existence of a person with the instruments of the simple view dooms the whole project to failure. To be sure, the decisive assumption for the simple view, that there is a strictly identical I before and after the break in the stream of consciousness, is not logically precluded. However, this possibility cannot, as will now be seen, be justified plausibly with the instruments of the simple view, without betraying the assumption that the first-person perspective is essential to the identity of a person over time.

### 2.1.4 *Easy Alternatives?*

Most supporters of the simple view have seen the problem that arises from the interruption in the current stream of first-person experience and suggested solutions which can be reconstructed as diverse strategies.

The *first strategy* consists in extending the principle of the primacy of self-attribution, contrary to the above criticism, to such cases which go beyond a current unbroken stream of self-conscious events (cf. Bermúdez 2012). Thus, Madell (1981, p. 135) presumes that the special feature of self-attribution, which declares a mental episode to be ‘mine’, is infallible and unanalyzable and also applies to diachronic cases such as that of memory or anticipation, in which interruptions in the stream of self-consciousness are bridged. Lund (1994, p. 191) also follows this extension of the principle of the primacy of self-attribution, albeit limited to the case of memories, and pronounces diachronic self-attribution to be sheer infallible. Although, considering the consequences entailed in other solution strategies, there are good reasons for this view, bearing the above formulated objections in mind, it is not convincing.

The *second strategy* was adopted by Leibniz in his *New Essays on Human Understanding* (Book II, Chap. 27). He postulated ‘imperceptible perceptions’ which constitute an unbroken stream analog to those of self-consciousness which however, remains below the person’s first-person experience. This conception being an integral part of Leibnizian metaphysics, does not, however, suffice to constitute the identity of the person which, according to Leibniz, consists of discernible self-consciousness. Since he, in contrast to other supporters of the simple view, saw that his foundation of a strict identity that is not reducible to empirical relations is insufficient for the identity of a person across time, he feels compelled to fall back on the goodness of God. It is doubtful whether this alliance of ‘metaphysics and morals’ can convince systematically.

The *third strategy* consists in conceiving the relation of person and self-consciousness according to the model of thing and property or substance and expression. Phases of interruption in the stream of self-conscious mental episodes do not pose danger for the identity, because self-consciousness is interpreted as the activity of a basal soul substance that outlasts these ‘gaps’ as an inactive substance. Swinburne (1986, Chap. 8–10), who proposed this conception, ultimately ties the strict identity of the person not to self-consciousness as the expressions of the soul,

but to the substance which instigates these actions. Apart from the fact that within the mental all problems that arise for concrete objects with the thing-property conception recur, and irrespective of the burden of proof of such a theory, two objections can be formulated. The first attests that the philosophical model of thing and property is inadequate for the analysis of self-consciousness (cf. Quante 2017a, b). This extensively discussed – above all in German Idealism – and well-founded thesis is also taken up by Madell (1981, w.134), when he accuses such a conception as that of Swinburne of ultimately still wanting to treat the person as an ‘object’. The second objection is that in this way the identity of the person is no longer tied to first-person experience and that therewith an indispensable feature of the simple view is being relinquished.<sup>19</sup>

The *fourth strategy* uses transcendental-philosophical arguments. An attempt is made to demarcate the strict identity of a person over time as a condition of the possibility of actions (Rohs 1996, Chap. 10). In first instance, the following general objection can be raised against the transcendental justificatory strategy: It cannot be proven that the condition in question is necessary for the presumed fact that has been taken for granted; it is mostly just a sufficient condition. Secondly, for our context the special objection can be articulated that the epistemic particularities conceded within a currently interrupted stream of self-conscious episodes are sufficient for actions (cf. Quante 1997a).<sup>20</sup> What is more, the transcendental-philosophical strategy also has to concede that the strict identity of a person over time cannot invariably be based on the first-person data accessible, since transcendental conditions are not normally a constituent part of the first-person perspective.

The *fifth strategy*, which was taken up by Foster (1979), bridges the gaps in first-person experience by linking up with laws of nature that specify the persistence conditions of the respective entity with which the person correlates or through which she is constituted. In the case of human persons, these are those nomological

---

<sup>19</sup>Although Chisholm (1986, p. 73 ff.) also takes up this strategy, he leaves open whether the indivisible substrate of personal identity should be conceived as monad or material sub particle. Clearly, he is primarily concerned with the indivisibility of this substrate, which should make the strict identity possible, rather than with a substance-dualistic answer to the problem.

<sup>20</sup>In Rohs (1996, Chap. 10 and 1997, p. 236 ff.), two further transcendental-philosophical arguments are to be found. First, the strict identity should be a necessary proviso for our practice of ascribing and evaluating actions. It is, however, debatable whether such a reconstruction really is suitable for our ethical practice. And second, the strict identity of a ‘standing and staying self’ should be a necessary proviso for experiences (in the Kantian sense) and communication (Rohs 1988a). The claim of self-conscious achievements drawn on diachronic and intersubjective invariance does not just stem from a very strong conception of experience and communication, which one does not have to share. For the question of the identity of a person that interests us here, its consequences are much too strong, since the recourse to a transcendental I raises the question of how the empirical many and the transcendental I relate to each other (cf. Cassam, 1997). Analog to Chisholm’s conception of mental events as universals/abstract entities, we find here, root of the attractiveness to take self-consciousness as a universal. Such an answer is patently insufficient for an analysis of personal identity; cf. Nagel’s conception of the objective self (1986, Chap. 4) and the considerations in this regard in Zuboff (1978, 1990) or Sprigge (1988).

conditions which constitute human persistence. The linking up of the person's identity of over time with the respective nomological regularities can be explained better with the aid of the relation of supervenience which is generally understood as a non-reductive dependency relation between two independent areas of entities. The chief motivation of this strategy is to avoid having to postulate constituting facts for the identity of a person which are neither empirically nor from the first-person perspective ascertainable. Here, supporters of the simple view run into into a dilemma: Lund (1994, p. 183 f.) rejects the supervenience theory because this way personal identity will ultimately be bound to empirical relations, while conceding that a person's identity is in principle a non-discernible fact. Inversely, Foster ties the conditions of personal identity over time to empirically observable relations which are also made use of in the framework of the complex view. With this the simple view is relinquished as an adequate answer to the analysis of the identity of human persons over time and it is accepted that persons' identity over time is constituted through other relations. Even if a person does not explicitly refer to these criteria in the first-person perspective, the persistence of the human person nevertheless consists in those (empirical) relations whose existence is required in the condition of supervenience. The theoretical framework of the simple view has thus been abandoned.<sup>21</sup> The identity of persons over time is tied to the identity conditions for the entity with which she 'correlates' or through which she is constituted, while the validity of the theory is restricted to the area of nomological possibilities.

To sum up, we can say that the prospects for the simple view are poor, given that the identity of a person overlaps with phases of interruption in the current stream of self-conscious mental episodes.: The different ways in which proponents of the simple view handle the problem of 'gaps in self-consciousness', and the respectively articulated criticism of the alternative solution strategies argue for the assumption that the complex view is the more promising way of analyzing the identity of the person over time. Thus – as the discussion in this section suggests – the complex answer to the question of the identity of a person over time should fulfill the following *conditions of adequacy*:

1. It should as far as possible be compatible with the intuitions that are called up by the simple view (above all, the independence of the existence of personal identity from social norms or linguistic conventions).
2. It should capture the stability and regularity of the existence of these relations, in order to put across why the fulfillment of the truth conditions are normally assumed implicitly in the daily practice of diachronic self-attributions of first-person mental episodes, so that these truth conditions are not themselves made a subject of discussion.<sup>22</sup>

---

<sup>21</sup> Foster (1979) emphasizes that this analysis applies to human persons and is just as compatible with the assumption of a non-bodily existing self as with the continuing existence of a human person as a purely spiritual entity after death. However, the biological approach suggested in the next section also accepts the logical possibility of this.

<sup>22</sup> Our everyday conception of the identity of persons over time thus has a 'default-and-challenge' structure; in contentious cases we are obliged to have recourse to the facts based on our first-person experience. It is these causal enabling conditions which constitute human persistence and can be compassed through the biological approach.

3. It should be able to explain the phenomenon of conflicting intuitions regarding thought experiments and the impression given of the undecidability of the question in these contexts.
4. It should tie up with plausible assumptions as regards the general problem of identity over time, and not have to devise postulates that are only valid for the specific case.
5. It should also be able to capture the intuitions relevant to (biomedical) ethics in respect of the beginning and end of existence.
6. It should be able to explain the confusion in the face of real ‘puzzle cases’ (irreversible coma, brain transplants, etc.).
7. It should be able to explain (as fully as possible) the intuitive equation of human being and person.
8. It should be free of the weaknesses of the simple view.

(conditions of adequacy)

## 2.2 Human Persistence: A Complex View

The basic idea of the complex view is that the identity of persons over time is constituted through empirically observable relations, so that the truth of statements about the identity of persons over time depends on whether these relations hold. According to the complex view, such an analysis is possible without having to refer to the fact of the identity of a person over time as unique, and it is not essentially tied to the first-person perspective.

Certain changes are incorporated in the transition to the complex view as opposed to the approach of the simple view, most notably a change in the methodological-epistemological approach: The complex view replaces the Cartesian with the observer perspective.<sup>23</sup> The connection to the observer perspective requires the integration of a causal element, since this perspective is marked out by causal or functional explanatory access. Such a causal element can be found in most variants of complex theories and is also implied by the fifth problem-solving strategy of the simple view. If one shares the widely accepted assumption that causal relations are connected with laws, it is reasonable to assume that there is a dependency on causal relations. This element of the complex theory, whose main features will be outlined further in due course, explains the intuitions that the relations constitutive for the identity of the person are empirically observable, stable facts that are not constituted through conventions or norms. The connection with causal laws limits the validity

---

<sup>23</sup>This perspective rather than the participant one proves suitable to define the existence of the identity of a person over time as a purely descriptively compassable one, independent of norms, assessments and linguistic conventions. In contrast, in the participant perspective not only does the first-person perspective play an important role, but it is also signalized by the constitutive function of evaluative elements such as expectation of meaning, presupposition of rationality and norms (see Chap. 5).

of the theory to situations possible under causal laws; this limitation of the scope of validity explains to a large extent why both our basal intuitions and our central concepts get confused in the context of such thought experiments as leave the area of what is possible under causal laws.

*Persistence* Since the identity of a person over time is, according to the complex view, a case of persistence that does not differ in principle from all others, the cumbersome reference to the identity of a person over time will be replaced by the term ‘persistence’ in this section. In general, two models of analyzing the persistence of space-time existing entities can be differentiated: ‘endurance’ and ‘perdurantism’ (Lewis 1986; p. 202 ff., Loux 1998, Chap. 6). The central thesis of *endurance* states that concrete individual things exist wholly at every moment of their existence and wander, as it were, as a whole through time. In contrast, space-time extended entities should, according to the basic assumption of *perdurantism*, be understood as four-dimensional entities that consist of temporal parts. At every moment only the respective present time slice exists, not the whole object. The suggested variant of a complex theory that follows is committed to the ontology of temporal parts.

### 2.2.1 *Causality: First Group of Presuppositions*

Since causality plays a central role in the complex theory suggested below, let me enumerate briefly the premises on which my argumentation is based.<sup>24</sup> As regards causality, the central conception for the observer perspective, three levels should be discerned: the level of singular causal explanations, the level of causal relations and the level of causal laws. In *causal explanations* an event is named as the cause of another event, whereby reference is made to both the causing and caused events by way of qualification. Causal explanations exhibit an evaluative dimension in two ways. On the one hand, picking out an event from a complex network of causally necessary conditions as *the* cause is based on pragmatic reasons relative to interests and on our prior knowledge. On the other hand, the explanatory power of such statements is based partly on the way the events in question are described. This verbalizes the reference to both our background theoretical assumptions and our interest-led relevance criteria. The explanatory power of these statements is based on the two events referred to actually being linked in the alleged correlation of cause and effect. Such *causal relations* exist independently of explanatory interests and evaluations. They can be understood purely descriptively and independently of

---

<sup>24</sup>The comments following are not a justification of these theses. As regards causality, what applies to philosophical theses in general, applies in particular here: none of them is undisputable. The outline sketched in following does not, in my opinion, stake a claim on anything that does not enjoy wide acceptance and is moreover neutral towards some debatable issues (e.g. event ontology, reduction or particular forms of causation such as mental or agent causation).



background theoretical assumptions or interest-led qualifications.<sup>25</sup> A purely space-time succession of events does not suffice for a causal relation to exist between them. But rather, the causal relation in question is required to be an instantiation of a *causal law*. Therefore, the events standing in causal relation must be available for description as being types of events linked by a causal law. Causal laws also have to be facts that are to be found independently of evaluations and interests. It is consistent with the consolidation of the explanatory power of causal explanations and the constitution of causal relations that the causal laws in question contain *ceteris paribus* conditions. This means that not every instantiation of a type of event referred to as cause in the law also automatically entails an instantiation of the type of event referred to in the law as effect: Causal laws can have exceptions, whereby the *ceteris paribus* conditions point out that these exceptions can be explained causally by factors not taken into account in the law. Hereby, it remains open whether these further factors can be embraced in the same theoretical framework in which the law is formulated, or whether maybe a reallocation to a different theory (e.g. from biology or chemistry to physics) is required.

This conception of causality shows two features that are relevant to the following considerations: For one thing, it facilitates a differentiation of causal explanations and causal relations so as to retain the evaluative aspects of our explanation practice without evaluatively ‘infecting’ the ontological level of persistence itself. With this, it is possible to comply with the intuition that the existence and non-existence of persons cannot be a matter of valuations and conventions. For another thing, the liberal conception of causal laws avoids the implausible corollary that only in physics can there be genuine causal explanations. The above model acknowledges that biology, for instance, possesses independent explanatory relevance. Since this conception is to apply to causality only as far as it occurs in the framework of the natural sciences oriented on the observer perspective, it remains neutral with regard to aspects of causality not embraced by this perspective (e.g. agent causality, mental causation) as regards both their existence and the nature of their law.

### 2.2.2 *Sortal Dependence: Second Group of Presuppositions*

*Preliminary Remark* A sortal predicate F, if applied to an object X, tells us what kind of entity this object is (standardly stated as F(X) as “X is a cherry”). If X is a persisting entity which can undergo changes, we have to distinguish two cases: In case A the change is such that F no longer applies to X (“X is no bachelor anymore.”) without X ceasing to exist. In case B the loss of X being F necessarily results in X ceasing to exist (“X is no human organism anymore.”)

Sortal predicates allowing for A-cases are called phase sortals, since they can apply to X at certain periods of X’s persistence and not at other periods of X’s persistence. Sortal predicates allowing for B-cases are called constitutive sortals, since

---

<sup>25</sup>Regarding the debate about an extensional (Davidson 1982) or intensional conception of events (Kim 1993, Chap. 3, Rheinwald 1994) these considerations remain neutral, at least for those theories in which the identity of events does not depend on evaluations.

they have to apply to all periods of *X*'s persistence (a loss would amount to *X* ceasing to exist).

If one asks where this difference between constitutive and phase sortals stems from a plausible answer (taken as a premise in this study) is that constitutive sortals deliver criteria for persistence (in case *X* is a persisting object) or identity (in case *X* is an abstract object).<sup>26</sup>

1. *Sortal dependence*: If one asks whether an entity *a* at one point in time is identical with entity *b* at another, or if one wants to know what counts as one entity, then neither the definition of numerical identity nor the principle of indiscernibility of the identical can be of any help.<sup>27</sup> The former already assumes that *a* and *b* are individuals, while the latter provides no answer to the question as to what comprises the persistence of an entity. If questions of identification and individuation are to be answered, then the field of the logics of identity must be abandoned. This is often carried out – as it will be in the following – through recourse to sortal concepts. My basic assumption here is that a statement '*a* is the same as *b*' depends on sortal concepts. However, two rival approaches must be distinguished here: the thesis of sortal relativity and the thesis of sortal dependence.

The conception of *sortal relative* identity, which goes back chiefly to Geach (1980), can be characterized by three assumptions (Rapp 1995, p. 158 ff. and 388 ff.):

- (a) The statement '*a* is the same as *b*' is incomplete and must be supplemented by sortal concept *F* if it is to be meaningful.
- (b) '*a* is the same as *b*' cannot be resolved into '*a* is *F* & *b* is *F* & *a* is the same as *b*'.
- (c) It is possible that *a* and *b* are the same relative to a sortal *F* and not the same relative to a sortal *G*.

In contrast, the conception of *sortal dependent* identity, which goes back chiefly to Wiggins (1980), contains the following assumptions<sup>28</sup>:

- (d) The statement '*a* is the same as *b*' refers to a sortal concept *F*, on which our practice of individuation and identification is implicitly or explicitly based. The statement is not meaningless, but elliptical.
- (e) Not all predicates that apply to *a* and *b* denote sortals that are equally fundamental to our practice of individuation and identification. One can distinguish

---

<sup>26</sup>I leave open here whether these criteria are delivered via the meaning of *F* or via its (hidden) reference to persistence relations. Furthermore, constitutive sortals are taken to deliver criteria to count tokens of *X*, i.e. to distinguish exemplars of a species.

<sup>27</sup>The following portrayal owes a lot to the study by Rapp (1995). But in contrast to Rapp's considerations, which are oriented on linguistic pragmatism, I take a realist conception as a basis, at least as far as human beings are concerned.

<sup>28</sup>This characterization corresponds to the conception based on the following and does not entirely align with either the conception of Wiggins or that of Rapp.

between constitutive and non-constitutive sortals: 'a is the same as b' refers to a constitutive sortal F, on which the truth of the statement depends. A constitutive sortal is not only one which is assigned to the entity in question at every moment of its existence, but rather, at the same time one which, firstly, provides the necessary persistence conditions and secondly, whose acquisition or loss means the beginning or end of the existence of the entity in question.

- (f) It is not possible to have two constitutive sortals F and G that apply equally to a and b and yield conflicting conditions of individuation or identification.

It is important to distinguish clearly between these two conceptions, because the pertinent criticism of the implausible consequences of the thesis of sortal relativity is usually looked upon as also being adequate to rebut the conception of sortal dependence I make use of. But as (e.) and (f.) show, the latter conception downright excludes sortal relativity. Admittedly, it thereby takes up burdens of proof which will be specified shortly.

2. *The necessity of origin*: A further premise that will be employed in the following states that the concrete event of its genesis necessarily belongs to the identity of a persisting entity. This means that a can only be the same F as b when a and b are generated by the same event (Forbes 1985, Chap. 6 and Kripke 1981). With this, a purely quantitative criterion of the specific essence of an individual is presumed, which will mainly become relevant in the context of the beginning of life (cf. Chap. 3).
3. *Natural kind terms*: Independently of one another, Kripke (1981) and Putnam (1979) proposed for terms that denote natural kinds the thesis that their meaning should be defined externally. This means that e.g. the concept 'tiger' depends in its semantic content on how tigers are really conceived. So, what is meant by 'tiger' depends on what we discover regarding this species. If, for example, we discover during the course of the progress of our knowledge, that dolphins are not fishes, but mammals living in water, then we have not had a different concept of dolphins in the past, or yet systematically mistaken the reference to dolphins, but rather, in using the sortal concept 'dolphin', we have been referring the whole time quasi indexically to real dolphins.<sup>29</sup>
4. *Natural kinds*: Whereas Putnam and Kripke only suggest an externalist theory for natural kind terms, without answering the question of what natural kinds are, the theory of persistence portrayed in the following enters into additional burdens of proof by drawing on a realist position of natural kinds, at least for highly developed animals. Even if there is no consensus in biology, or rather in the philosophy of biology as to how species should be precisely individuated and which precise ontological status they possess, nevertheless, a far-reaching realism can be detected as regards higher developed animals.<sup>30</sup> In the following it is assumed as a premise that in biology, specific laws are discovered which describe

<sup>29</sup>The externalist analysis of concepts for natural kinds does not imply the stronger thesis that the reference alone establishes the meaning.

<sup>30</sup>Since the following is concerned only with the persistence of the human individual, this suffices. A general theory of biological species is not necessary for my purposes.

the organization, functions and normal development of biologically normal members of this species.<sup>31</sup> These laws are causal laws which can be devised from the observer perspective in as far as they refer to processes of evolving, growing, aging and dying. This concerns causal laws specific to species, which express the conditions of fulfillment for the persistence of members of this species. These regularities, which are referred to via specific sortal concepts used in biology,<sup>32</sup> are the nomological correlations that are not dependent on evaluations and conventions and which are necessary as a basis of a complex theory of persistence. The realism about natural kinds is justified additionally, since this way it becomes understandable why specific sortal concepts are more fundamental than others in our practice of identification and individuation. This is because in one case detectable regularities guide this practice implicitly and not in another. Furthermore, it can in this way be appreciated why there can be no conflicting identity criteria. Since causal laws are not in competition with one other, at most the case can occur that there are general and specific laws, e.g. those for mammals and those for human beings.<sup>33</sup> In this case, the laws pertinent to the species provide the persistence criteria respectively.<sup>34</sup> The thesis that our practice of individuation and identification presupposes e.g. the validity of current regularities for human beings, moreover explains why the diverse thought experiments which have been constructed in the discussion about the identity of the person, evoke conflicting intuitions: these scenarios transcend the area of application of our persistence criteria. In addition, this thesis makes it understandable why there is a tendency in the context of theoretical considerations to use the terms 'human being' and 'person' coextensively: it is the former term which guides our identity statements when we talk about human persons. In so doing, the concept of the human being is used in a purely biological sense (i.e. biology as a science). The general parlance about persistence has to be rendered more precise by indicating the respective sortal concept from which the identity criteria in question may be won. For this reason, the following is concerned with the persistence of the human individual; in my conception, there is no persistence of the person.<sup>35</sup>

---

<sup>31</sup> In isolated cases any occurring aberrations or deviations can concur with these laws due to the *ceteris paribus* clauses.

<sup>32</sup> This recourse to biology is not meant as a contention that the everyday use of this sortal concept and everyday statements on the persistence of members of this species function basically in a different way. For one thing, the regularities included in this way can in part also be grasped socially, and for another thing, there is resort to the factum of linguistic division of labor (Putnam 1979). But there are also areas (e.g. in embryology) which are for the most part inaccessible to the social approach.

<sup>33</sup> This formulation presumes that causal laws are facts. If, instead, causal laws are understood as propositions, then the thesis reads that two true causal laws cannot contradict one another.

<sup>34</sup> In contrast to Rapp's suggested connection to the pragmatics of our identity statements my suggestion enters into a higher metaphysical hypothec. In my view this is justified above all by the fact that it allows to explain the non-conventionality of the persistence of members of certain biological species. But it has to be noted here that therewith the analysis of persistence suggested in the following is restricted to higher species and is bound to actual biological laws.

<sup>35</sup> The thesis that the persistence conditions for human embryos must be coined on the basis of the concept of the human being is also defended by Ayers (1993, Chap. 22–25), Olson (1997),

### 2.2.3 *The Biological Approach*

The precondition to analyze persistence via empirically observable relations of causal space-time continuity on which causal laws are based, leads as regards the human individual to the theory that here the concept of the human being – understood purely biologically – rather than the concept of person must be taken up. A *demanding* concept of person, which can only be applied to an entity when it possesses certain properties and capacities to an adequate degree (Dennett 1981, Chap. 14), does not possess the attributes required for an analysis of persistence. Obviously, evaluations and norms play a role in the characterization of an entity as a person, so that the independence required is lacking. Because of this evaluative character of personhood it can also not be expected that the concept of person denotes a natural kind through which the special causal laws required by the analysis of persistence may be gained. For one thing, this follows from the possibility that quite heterogeneous kinds of entities can be persons, so that possibly even conflicting causal laws may be relevant for the amount of entities defined as persons, or, in the case of artificial persons, no pertinent laws exist. For another, the evaluative character of the concept of person leads to the causal links ascertainable on the observer perspective showing an under-determinateness as regards the personhood of an entity, since the evaluative aspects cannot be captured. Furthermore, a demanding concept of person can only apply as a phase sortal, i.e. a sortal concept which can apply to an entity at certain periods of the existence over time and not apply at others: It is obviously not the case that an interim comatose human being, who no longer fulfills the conditions of the demanding concept of person, stops existing, or that a new entity starts to exist after reconvalescence. The demanding concept of person can therefore not be understood as a constitutive sortal that determines the conditions for the persistence of entities of this kind. For, to recall, a constitutive sortal records those properties and capacities of an entity, whose acquisition and loss accompany the beginning and end of the existence of this entity.

If the demanding concept of person located in the participant perspective and including our perception of personhood and personality is unsuitable for an analysis of persistence, then maybe an undemanding concept of person could still be defined which – free of evaluations and norms – is used exclusively in the observer perspective (recall the distinctions introduced above in 1.3). If one, for example, understands by person merely an entity to which both psychological and physical properties can be attributed without going further into detail about the exact nature of these psychological properties and capacities, the concept of person is, however, no longer being used in the accepted sense.<sup>36</sup> If one is looking for the persistence of the *person*, then such an undemanding concept of person is prima facie an advantage, since

---

Snowdon (1990, 1991) and Wiggins (1976, 1980, Chap. 6). It is important to note that the above thesis does not imply, first, that only human beings can be persons, and second, contrary to Wiggins, that it is intended as an analysis of the identity of *persons* over time.

<sup>36</sup>This point has been made by Harry G. Frankfurt against Peter F. Strawson's conception of the person developed in "Individuals" (Strawson 1959, Chap. 3); cf. Frankfurt (1988, Chap. 1).

it is the most likely to facilitate the fulfillment of the conditions required of a complex theory. So the main issue is whether it is the psychological properties, the physical properties, or both classes together which determine the persistence of a person. But in this discussion it is conspicuous that the different theories swing between a demanding and an undemanding concept of person. This shows up above all in the properties and capacities that are taken into account in the analysis and, in the end, which conditions are acknowledged as being sufficient for persistence. The following dilemma ensues from abiding by the concept of person within the context of the analysis of persistence: the more sophisticated the underlying concept of person is, the more problematic it becomes, on the one hand, to build up a persistence analysis oriented solely on the observer perspective that does not underdetermine the phenomenon. On the other hand, an analysis oriented solely on the observer perspective is all the more successful, the less demanding the underlying concept of person is. However, it is not clear what an entity thus analyzed in respect of persistence conditions, has to do with personhood.

The way out of the dilemma described in the following entails dispensing with the concept of person for the analysis of persistence and instead calls on that sortal concept which denotes the natural kind of the respective entity. In the case of human individuals this is the biological concept of 'human being'.<sup>37</sup> With this, the persistence conditions for human individuals are determined by the biological regularities relevant to members of the human species.<sup>38</sup> So it is not the body of a person in the sense of an aggregate of material, which enters into the conditions of persistence, but the causal continuity of an *organized* body whose material parts can be replaced without endangering the persistence of the organism. Even the brain, which plays a prominent role in the context of the debate about the identity of the person over time, is not merely a collection of specific atoms or cells, but a functionally specific organ that is individuated not by its constitutive material parts but through its organized state, which refers to the human organism as a whole.

The central psychical functions of the human being relevant in most of the complex theories can be integrated into the biological approach: either directly, when they can be captured from within the observer perspective, or indirectly via their biological bases of realization. However, in contrast to many other complex theories, in the biological approach psychical functions are not necessary conditions for human persistence. Objections can be raised against such approaches that support a biological approach. According to these alternative complex theories, psychological properties and capacities are not sufficient, but nevertheless necessary for the persistence of a human person. Even if it is possible, within the framework of such theo-

---

<sup>37</sup> I have explained the reasons why the other variants of the complex view do not represent a satisfactory alternative to the biological approach in Quante (2001a).

<sup>38</sup> For the portrayal of the general strategy it can remain open here whether 'human being' denotes a natural kind, or whether the laws that can be formulated within the framework of the basis of biology are relevant for other animals (e.g. apes), too. There is no need for a definition of the persistence of the human being as opposed to all other forms of life. Whether 'human being' defines a natural kind, or how the natural kind that includes the human being is to be precisely defined, is a question to be decided empirically.

ries, to bridge interruptions in the stream of psychological events by dispositions, two grave problems still remain. If one takes the case of a human being lying in an irreversible coma from a specific point in time and, due to the destruction of the brain areas needed for psychological states and episodes, leading only a vegetative life, then, according to the complex theory, the individual in this life period cannot be the same individual that existed before the onset of the coma. This case can be seen as a special one at the end of life, on which one cannot build a theory, or because of which one should not give up an otherwise plausible theory. But the second difficulty is present in every human individual: According to such a theory, the phase in the development of a human life in which the basis for psychological states and events have not yet been developed, do not belong to the individual's later existence. Both consequences are extremely counterintuitive. If, in contrast, one accepts the assumption that the persistence conditions in question stem from the biological concept of the human being, then both the early stages in the development of the human embryo and the irreversible phases of merely vegetative life can be understood as phases in the existence of a persisting human individual.<sup>39</sup>

In contrast to other complex theories which use the concept of person, the biological approach also has the advantage of not being dependent on a naturalistic conception of the propositional attitudes that are central for personhood and personality (memories, intentions, etc.). On the contrary, in the framework of this conception it suffices to make available the necessary basis for these mental episodes within the framework of a description of biological functions oriented on the observer perspective. Since this is not aimed at reconstructing a demanding concept of person via biological concepts, it is sufficient if the biological enabling conditions which are necessary so that human individuals can be persons, are captured. Personhood and personality cannot themselves be understood within the framework of a naturalistic conception, but belong to the participant perspective. Since, in compliance with the overall conception of the biological approach suggested here, no persistence conditions can be determined at this level, this area can be completely omitted in questions of human persistence.<sup>40</sup> Inversely, the factual constitution of the biological organism plays a central role as a basis for realization of personhood and personality. One needs only to think of how the growing up, aging, being ill or dying specific to humans characterizes our concepts of personhood and personality.

The starting point of the biological approach is the human organism as a self-integrating unit. As long as one does not additionally subscribe to the naturalistic thesis that the only phenomena and entities that really exist are those which can be

---

<sup>39</sup>The *prima facie* counterintuitive consequence that the biological approach involves is that a corpse is a different entity from the human being it once was (cf. Chap. 4; Rosenberg arrives at the same conclusion, 1983, p. 27 f.).

<sup>40</sup>The exclusive connection to the observer perspective does not allow analyzing the organized body as 'my body', i.e. the body a person attributes to herself. This would admit the first-person perspective back into the analysis, so that Hamilton's criticism could bite (1995, p. 346).

fully grasped within the framework of the observer perspective, no reductive or even eliminative consequences arise from the complex analysis suggested here.<sup>41</sup>

### 2.2.4 Conclusion

All in all, the biological approach suggested here fulfills the adequacy conditions formulated at the end of the first section: The bond with the biological regularities ensures the independence of persistence from social norms or linguistic conventions, explains the stability and regularity of the existence of these relations and, with this, also why, in the daily practice of diachronic self-attributions of first-person mental episodes, the truth conditions for persistence can normally be assumed as given and why these truth conditions do not themselves have to be discussed. At the same time, it becomes understandable why we intuitively tend to equate ‘human being’ and ‘person’ in issues of personal identity. Furthermore, the bond with factually valid laws explains in general why the thought experiments dealt with in the literature evoke conflicting intuitions and the impression of undecidability: the concepts with which we answer questions of persistence are utterly unsuitable for dealing with these cases, and the concept of person itself, which is drawn on in these contexts, does not facilitate any clear cut answers.<sup>42</sup> Furthermore, the omission of the suggested fixation on psychological states and episodes from the concept of person also makes it possible to embrace everyday intuitions as regards the beginning of life, death or coma states. All in all, the biological approach can be formulated completely in the observer perspective and can manage without special postulates which would have to be set up exclusively for the analysis of the identity of persons over time. The burdens of proof of the biological approach are considerable, though fewer than those of the simple view or the alternative complex theories. Above all, a realist conception of natural kinds is needed, which is backed by biol-

---

<sup>41</sup> Cf. Quante (2000a); the declination of a naturalistic conception of propositional attitudes thus is not committed to taking (and sticking to) the Cartesian perspective. On the contrary, the simple views can be accused of masking the evaluative and ‘hermeneutic’ aspect of personhood and personality just as much as the naturalistic complex theories do, by adapting the first-person perspective to the observer perspective. In the face of this, it is suggested here that personhood and personality be assigned to the participant perspective that contains the first-person states as an integral part. Connected with this is the thesis that self-consciousness also always contains evaluative and volitional aspects (Tugendhat 1979). Chisholm (1970a, p. 36 f.) agrees in one sense, in that he makes a case for a ‘loose’ rather than a strict application of the concept of identity for personhood and personality. With this, expressed in the terms of my approach, he sums up the difference between an analysis oriented on the observer perspective and one on the participant perspective. But what comprises personal identity when personhood and personality are excluded, remains hazy.

<sup>42</sup> The extrinsic conditions criticized by the simple view (such as the exclusion of doubling) are not an ad hoc solution, but result from the relevant biological laws containing *ceteris-paribus* conditions. Neither the manipulation nor the presence of constraints that hinder the normal development of a human organism are directly embraced by this approach.



ogy or, rather, the philosophy of biology, at least in regard to human beings. In an adequate formulation of the biological approach, use will also be made of categories which go beyond the framework of biology itself: first and foremost, the concepts 'organism' and 'integration' or, rather, the integrative achievement (cf. Chaps. 3 and 4). At the same time, the scope of the biological approach is limited. This not only has the already described welcome consequence of banning undecidable thought experiments to the realm of 'science fantasy' for good reason, but also the effect that the persistence conditions are limited to the normal determining factors of the specific relevant laws for each natural kind. So, within the framework of this theory, nothing can be said about artificial persons and about the personal identity of God or angels. Even in regards to human beings, it emerges that technical interventions such as e.g. the transplantation of brain tissue (or parts of the brain) as well as the artificial replacement of specific organic functions of the brain, possibly exceed the normal application area of this theory. Whether this is the case and, if applicable, what consequences are to be drawn from such results, must be revealed by the discussion of the issues in question in the following two chapters.

## Chapter 3

# The Beginning of Life

The question of when a human life begins is usually posed in ethical contexts.<sup>1</sup> Whoever wants to know if – and when – abortion is ethically permissible links the question of the beginning of life with genuine ethical problems: the question of the ethical status of the human individual in her first stages of life is one way to frame the issue. The same applies to the disputes concerning the admissibility or inadmissibility of research on human embryos and preimplantation diagnostics. Whoever enquires about the point in time at which an individual human life starts to exist does this mostly with a moral intention. Furthermore, the question of the beginning of life is frequently expressed in the first-person perspective, as exemplified by the title of Ford's book *When did I begin?*.

Both this first-person perspective and above all the seemingly almost unavoidable link<sup>2</sup> to the ethical issue of a 'right to live' or the 'moral status' of the beginning of human life are, however, not conducive to an appropriate answer to the question of when the persistence of a human being begins, but lead in various ways to implausible results. Thus, in this chapter – in alignment with and in continuation of the results of the second chapter – an answer to the question of the beginning of a human organism's life is to be arrived at, which on the one hand is free of ethical considerations and on the other is bound to the observer perspective. Since thereby the following argumentation differs from a large proportion of the available literature, let us first formulate the targets and issues to be taken up in this chapter.

---

<sup>1</sup> This question is understood here as being a question of when a specific human organism begins to exist, i.e. live.

<sup>2</sup> A printing error which slipped into the following review of Ford's book (1991) provides evidence of how difficult it evidently is to not make ethical contexts the subject of discussion. In Flaman (1991, p. 39) the original subtitle of Ford's book, which is "Conception of the human individual in history, philosophy and science", reads as the "contraception" of the human individual.

## 3.1 Targets

### 3.1.1 *Preconditions*

This attempt to answer the question of the beginning of life of a human individual bases on the hitherto developed main features of the biological approach, whereby two questions are to be omitted. One is the question as to how the living is generally distinguished from the non-living. The other is the question of how the human being can be defined as a biological kind. To this end, in the following both the human individual's state of living and the affiliation to a kind are assumed as given.<sup>3</sup> In addition, the arguments in this chapter – and in the following one – rely on two concepts which will not be analyzed in detail. These are the concepts 'organism' and 'integration.' Although I will enter into an elucidation of both concepts within the framework of the elucidation of my proposal for a definition of the beginning of life of a human individual, I cannot present an exhaustive analysis thereof in this study.

### 3.1.2 *Targets*

The main concern of this chapter is to answer the question of the beginning of life of a specific human individual (Sect. 3.2). Even if a general definition of the beginning of an organism's life is assumed, the criterion for the beginning of life proposed here is limited to human organisms.<sup>4</sup> This means that no general theory will be developed for the conditions of persistence, but rather that it will be restricted to human individuals.<sup>5</sup>

Above all, this is not about developing a general ethical theory for dealing with the beginning of human life. In accordance with the basic conviction of the biological approach, the persistence of a human individual can only be determined with recourse to the concept of the human organism and this can only be done from the observer perspective. Both together lead to the ethical scope of this biological approach being limited, as the discussion of the bond between persistence, potentiality and personhood shows (Sect. 3.3). But even if no comprehensive ethical theory for dealing with the beginning of human life can be deduced from the answer to the question of the beginning of life of the human individual given by the biological approach, this answer is nevertheless relevant for some biomedical issues such as e.g. the difference between abortion and contraception, or the evaluation of human genetic interventions (Sect. 3.4).

---

<sup>3</sup>For this reason, the issue of the affiliation to species of transgenic creatures plays just as minor a role as the question of the ethical admissibility or inadmissibility of the genetically modified link between human and non-human genes; cf. Quante (2001b).

<sup>4</sup>The distinction between definition and criterion is analog to that of the definition and criterion of death and borrowed from that discussion context (cf. the next chapter on this).

<sup>5</sup>The attempt to develop a general theory of 'biological individuality' runs into sheer insurmountable obstacles in view of the variety of forms of life (on this cf. Wilson 1999).

## 3.2 The Beginning of Life: Human Persistence (I)

In this section, as a first step, an answer to the question of the beginning of human life, which arises from the conception of human persistence unfolded here, will be proposed and justified (Sect. 3.2.1). Following on that, in the second step this answer will be defended against the main objections that can be found in the literature, whereby the different aspects of the biological approach to human persistence will be employed (Sect. 3.2.2). Finally, attention will be directed to the hard problem to be confronted by the answer suggested here – the graduality of the achievement of integration of the human organism (Sect. 3.2.3).

### 3.2.1 *When Did This Human Organism Begin to Exist?*

If one asks the question of the beginning of human life in this way, then three important decisions have already been made. Firstly, it is being asked – post festum – when an already existing entity began to exist. The question is thus being asked at a point in time at which the individual in question already exists, so that the reference to ‘this’ is clearly fixed. The following is neither concerned with possible entities nor with answering the question of how one can say of a certain situation that it contains the potential to produce a new human organism (see also Sect. 3.4 in this chapter). Despite this retrospective perspective, the initial question will for two reasons not be asked in the first-person form: for one thing, the first-person perspective has proved unsuitable for answering questions of persistence. And for another, the first-person form of the question implies the wrong conclusion that ‘my’ existence can only have begun at the point in time at which I possessed self-consciousness. The third decision that was made with the above wording of the question consists in not using the equivocal concepts of person or ‘human being’, but rather, in speaking explicitly of the human organism. Thus the purely descriptive observer perspective directed at causal-functional explanations is taken up rather than the comprehending participant perspective. More precisely, the reference to the concept of organism signalizes that the following will be concerned with the interplay of general metaphysical considerations and biological facts.

I understand an organism to be the self-integrating *process of life* rather than the body and its constitution (similarly van Inwagen 1990, p. 145 ff. and Wilson 1999, p. 101 ff.).<sup>6</sup> Life is therefore not the property of a space-time entity, but itself a process extended in space-time. The appropriate material body with its constitution

---

<sup>6</sup>With this event or process conception of life and the organism, the biological approach does in fact depart from the concept of common sense; DeGrazia (2005) has developed a conception very similar to the one suggested in this study, in which he doesn’t commit himself to such process conception of life and organisms. My aberration is, however, justified in that such a conception is not only more compatible with scientific facts, but also leads to ethically worthier results with regard to the determination of death (cf. Sect. 4.5.2. in Chap. 4 on the latter).

is thereby the most important enabling causal condition for the occurrence of this life process. Inversely, one can say that this life process is realized in the actualization of the properties and capacities of such a material body. Beyond this decisive changing of points between the organism or life as processes versus body and property, the talk of a biological individual shows ambiguities which will also have to be discussed briefly.

In his comprehensive analysis of the concept of biological individuality, Wilson points out that both the biological and the philosophical literature reveal many blind spots regarding the issue of individuating living creatures, due to a one-sided choice of examples: ‘Horses and Oak Trees’ (Wilson 1999, p.5). Therewith, the majority of *Lebensformen* are totally disregarded so that the resulting conception is too specific and unsuited to a general theory. According to Wilson’s analysis, a biologist can mean six different things with the thesis that a living entity is an individual (Wilson 1999, p. 60):

1. It is a **particular**. A biological entity is a particular just in case it is neither a universal nor a class.
2. It is a **historical entity**. A biological entity is a historical individual if it is composed of spatiotemporally continuous parts.
3. It is a **functional individual**. A biological entity is a functional individual if the parts which compose it are causally integrated into a functional unit.
4. It is a **genetic individual**. A biological entity is a genetic individual if its parts all share a common genotype.
5. It is a **developmental individual**. A biological entity is a developmental individual if it is the product of a developmental process.
6. It is a **unit of evolution**. A biological entity is a unit of evolution if it functions as an important unit in an evolutionary process.

(Various meanings of ‘biological individual’ according to Wilson)

In contrast to Wilson, who tries to give a comprehensive answer to the problem of biological individuality by distinguishing between different kinds of individuals (cf. *ibid.* Chap. 5), the following concerns only human organisms and the biological laws pertinent to them. So the special features of other *Lebensformen* do not have to be taken into consideration here. However, it is significant that Wilson points out a general ambiguity that accompanies the use of the concept of organism in biology. He registers dithering between the functional and developmental sense of the individual, whereby the condition of having to develop out of a single cell is frequently added (*ibid.* p. 62). Even if the two conditions: development and emergence from a single cell are applicable in the case of the development of a human being, the concept of organism is in the following limited to the functional unit that exhibits the attribute of ‘causal integration’ (*ibid.*).<sup>7</sup>

---

<sup>7</sup>Wilson names the following as paradigmatic attributes of higher organisms, which are mostly and wrongly imputed to be broadly binding: spatial and temporal continuity, spatial and temporal boundedness composed of heterogeneous causally related parts, development from a single cell to a multicellular body, subject to impaired function if some of its parts are removed or damaged, ability to reproduce sexually, and genetic homogeneity (Wilson 1999, p. 9 – an even more comprehensive list can be found on p. 56).

The concept of integration, which I use to characterize organic life processes, is also not unproblematic. By the making of a whole I understand a unity of parts and leave open here in which form this achievement of integration is to be analyzed more closely – as a causal process, in cybernetic or system-theoretical terms, or teleonomically. In our context, it is only important that we can speak of an achievement of integration when the parts of a whole are oriented to one another and the subfunctions react to one another in such a way that one can describe the whole (the organism) adequately<sup>8</sup> as self-regulating and self-maintaining. The concept of integration is indispensable, as it is an integral part of our concept of an organism. What is decisive is the *holistic* insight that the organism itself and not some of its parts or even ontological entities independent of it achieves the achievement of integration.<sup>9</sup> This notion is, however, certainly compatible with the failure of a certain organ being enough for the organism to no longer accomplish the achievement of integration, so that the failure of this critical organ can be identified with the death of the organism. Even when the organism as a whole is the ‘subject’ of the integrative achievement, it can only attain this when the necessary components function adequately (or have been substituted). Furthermore, it cannot be ruled out that in different phases of the existence of an organism (or ‘components’) this achievement of integration has causal support by different organs. Therefore, according to the definition proposed here, the organism begins to exist as soon as it accomplishes the achievement of integration itself.

### 3.2.1.1 Definition

In order to answer the question of the beginning of life of a human organism, I will begin with the following suggestion for a general definition:

Beginning of life: The start of the existence of a specific organism.  
(Definition of the beginning of life)

This definition is general because it is not yet limited to certain kinds of living creatures, but applies equally to all organismic life processes. The limitation to a specific kind follows only on the level of the criterion (see Sect. 3.2.1.3). Despite this generality, the above definition is not ‘empty’, as the following explanations demonstrate.

---

<sup>8</sup>The stronger condition, that one can only describe this process with loss of information if one waives the concept of self-regulation, I also consider plausible as regards organisms. But since I cannot defend it here, above I only draw on the weaker condition. However, my above assumptions imply that the concept of integration can be analyzed without recourse to the participant perspective.

<sup>9</sup>The integration should therefore be understood not just as a merely causal bond, but as the organism’s own causal achievement (cf. Mishler and Brandon 1998, p. 302 f. and Wilson 1999, p. 52 ff. on this).

### 3.2.1.2 Explanations

- (a) “*Specific*” organism: This component is necessary because the reference must also be fixed; there must always be an exact definition for that entity whose beginning of life is being sought. Since life emerges from life and single living entities fuse into a new organism, coming into existence through the dissolution of a subsequently no longer existing organism, or start to live as offshoots, it must always be clear which particular organism is under discussion.
- (b) Specific “*organism*”: This definition limits the concept of the beginning of life to the organism, so that life beyond the organism (e.g. single cells) or combinations of organisms are excluded.

Alongside these two clauses that are contained explicitly in the above proposal for a definition, there are two further implicit provisos which are central to the following considerations:

- (c) *The observer perspective – first implicit proviso*: Since, according to the biological approach, persistence can be understood as a causal-functional relation defined by the causal laws valid for the natural kind of the respective individual, the purely descriptive observer perspective is pertinent to the question of the beginning of life (cf. Chap. 1). This means that the concept of personhood or personality is not viable here, since the latter can only be applied in the comprehending participant perspective (cf. Chap. 5). Therefore the question of persistence and with it the question of the beginning of life and of death must be kept free of ethical considerations. According to the biological approach, these issues are to be decided by descriptive arguments consisting in a bond between metaphysical considerations and biological facts. These descriptive arguments, though relevant to ethical problems, are not themselves ethically impregnated.
- (d) *The causal element – second implicit proviso*: Since persistence can be analyzed as a relation supported by causal laws, the biological approach contains a causal element whose more detailed theoretical definition is significant in the following. The causal laws drawn on for the persistence of human organisms are of a biological nature and contain *ceteris-paribus*-conditions, i.e. they allow for exceptions which can be explained by these conditions. These additional parameters can be found on the biological level as well as coming from other fields (possibly physics or chemistry, but also from the social sciences). Furthermore, such conditions that belong intrinsically to the organism in question (e.g. the normal development of DNA or the genetic copying process), or are extrinsic in relation to the organism in question (e.g. normal processes in the organism of an expectant woman, or environmental conditions) can also fall under the *ceteris-paribus*-conditions.<sup>10</sup>

---

<sup>10</sup>Due to this fact, the biological approach does not fulfill the so-called “Only-X-and-Y-Principle” either; only such facts may be relevant for the identity of X and Y over time as, according to this principle, belong intrinsically to X (resp. Y) (cf. Quante 1995a, p. 41 ff., also Garrett 1990; Noonan 1991, p. 16 f.). The intuitions based on the “Only-X-and-Y-Principle” are either due to the Cartesian intuitions of the first-person perspective, which have already been rejected (cf. Chap. 2), or they rest on the confusion of (numerical) identity and persistence (cf. Chap. 1).

This structure governed by natural laws explains why our practice of the specification of identity is based on normal cases of development and pointers to anomalous and special developments or technical interventions do not hold water as an objection to the criterion for the beginning of life proposed below. Moreover, the involvement of causal laws will permit a precise meaning of potentiality to be won (see Sect. 3.3).<sup>11</sup>

### 3.2.1.3 The Criterion of the Beginning of Life

The criterion I propose, which I will explain and defend against objections in the following, is:

The beginning of life of a human organism is the initiation of the activity of the individual genome of this organism, which normally happens at the four to eight cell stage (between the second and fourth day after the fertilization of the egg-cell) and is the beginning of the self-regulation of this individual life process.

(Criterion for the beginning of life)

This criterion is a combination of general metaphysical assumptions, specific implications of the biological approach and general biological facts.<sup>12</sup> After the penetration of the sperm into the egg-cell (impregnation), about 24 h pass before the fusion of the haploid genomes of the gametes is completed and the zygote has formed. Then the blastomeric stage begins in which first in two, then four and subsequently in reduplicating steps ever more numerous cell unions are formed, which on the one hand show a genetic identity and on the other are enclosed in the zona pellucida (cf. Bodden-Heidrich et al. 1997, p. 66 ff.). The genome that has been formed by the fusion of the gametes is only activated from the four cell stage on and then takes over the regulation of the further development process (Bodden-Heidrich et al. 1997, p. 70; Ford 1991, p. 113).<sup>13</sup> It is this initiation of self-regulation which, according to the above criterion, is interpreted as the standard case of the beginning of an individual human life process, because from this moment on an active integration through the organism itself follows, while the previous stage in development is

<sup>11</sup> The bond between causality and persistence is shown in general form by Brennan (1988, p. 26 ff. and 99 ff.) and elaborated as regards biological entities by Wilson (1999, p. 52 ff.). Implicit or explicit references to a causal component of persistence in the context of the discussion of our problem of the beginning of human life can also be found in Williams (1990, p. 169) and Ashley (1976, p. 127 – the latter even speaks of “definitive laws”).

<sup>12</sup> In the discussion it is more or less undisputed that in terms of scientific facts there is no dissent (either regarding the knowledge or with respect to the areas in which the facts are not completely clarified, or hardly clarified at all). In fact, the dissent only arises on the level of the philosophical interpretation of these facts, either as regards their metaphysical or ethical significance. In respect of the biological facts, the following remarks rest on the views of Bodden-Heidrich et al. (1997), Fisher (1991), Ford (1991) and Morowitz and Trefil (1992).

<sup>13</sup> Whereas Ford (1991, p. 118) places the moment of activation between the two and four cell stage, Bodden-Heidrich et al. (1997, p. 70) locate this “switch-on” in the phase between the four and eight cell stage.



still steered by the ‘maternal’ RNA and further components of the oocyte (Bodden-Heidrich et al. 1997, p. 70).

Because this criterion is both complex and not a matter of course, a few further explanations shall be given, since in this way certain strategies for objections can be repudiated as unfounded (Sect. 3.2.1.4). Following that, various objections to this criterion will be discussed, for, even if the above named point in time has not, to my knowledge, been suggested explicitly in the literature, there are nevertheless numerous research reports which favor either an earlier or a later point in time. The various arguments for and against these different positions must therefore be vetted for their validity as opposed to the criterion proposed by the biological approach (Sect. 3.2.2).

### 3.2.1.4 Explanations

- (a) “*Human*” organism: The above criterion is explicitly limited to human organisms. Thus questions of the beginning of life of other biological entities and also e.g. artificial life are not discussed.
- (b) “*Activity*”’: Since, according to the biological approach, organisms are individual life processes, whose integration into a life process can be credited as their own achievement, the mere presence of the capacity for self-regulation is not sufficient. On the contrary, an activity must be detected which can be understood plausibly as the exercise of this capacity for self-regulation.<sup>14</sup> In accordance with the above criterion, this is accomplished with the ‘switching-on’ of the individual genome.
- (c) “*Normally*”’: The above criterion allows for exceptions. Factually, human organisms whose existence began at a different point in time as that named in the criterion do exist. These are multiple monozygotic embryonic formations, whose existence only starts after the fission of the original life process. For these multiples it also applies that their respective existence begins as soon as their individual genome is activated and exercises self-regulation.<sup>15</sup> The above criterion does not apply in such cases as must be embraced by the *ceteris-paribus*-conditions of the relevant biological laws.<sup>16</sup>

---

<sup>14</sup> So the objection by Tonti-Filippini (1989, p. 41 – all the following quotes come from this source) against this criterion (this activation “occurs from within the zygote itself”) comes to nothing; the spatial location is not enough to bring forward self-integration. The zygote therefore only has the potential (“contains its own programming”) to initiate an integrated life process, but this actually begins only at the moment of the “switching-on of the embryonic genome”. In his criticism, Tonti-Filippini misjudges the objective thrust of the discussion about the switching-on of the genome when he replies that “genetic uniqueness or individuality is neither sufficient nor necessary to establish ontological individuality”. According to the biological approach, at least, the argument about the switching-on of the genome is about self-regulation as being the execution of a capacity and not merely a potential or an unused capacity.

<sup>15</sup> In the next section I will go into more detail about the various aspects of the problem of fission.

<sup>16</sup> In the case of multiple embryo formation, Williams (1990, p. 175) also falls back on a disclaimer. However, this consists in explicitly assimilating the no-fission-condition in his criterion for the beginning of life, whereas this case is covered by the underlying structure of the laws and thus the ad-hoc objection does not hold water.

(d) The “*individual genome of this organism*”: This proviso comprises two aspects which are essential to the biological approach and the following argumentation.

(i) For one thing, the individual genome is oriented to a specific organism. This means that an organism is not individuated via the genetic identity. Genetic identity cannot suffice as a general criterion for the identity of a biological organism, as the case of the monozygotic multiples shows. And it cannot be essential, because for one thing copying errors in the causal history of an organism, or – through technical interventions such as chimeric formation or organ transplants – hybrids can occur.<sup>17</sup> The condition of strict genetic identity can be replaced by the condition of causal continuity in connection with the limitation to essential genetic identity – at least for variants not caused by technical interventions. The problem of genetically identical multiple formations is solved by insisting on space-time and causal cohesion in addition to genetic identity for the existence of an organism. According to this criterion, e.g. monozygotic twins are defined as two genetically largely identical, but numerically different organisms.<sup>18</sup>

(ii) For another thing, the above definition contains a fixed reference (“to *this organism*”) because of the post festum perspective. This establishes whether this is a normal case of the beginning of life, or whether it is one of the possible exceptions, and guarantees that this individual life process has lasted up to the moment of the fixing of the reference. In this way the objection is circumvented that many of these individual life processes are terminated at a very early stage. For one thing, such an objection is not really cogent against a criterion for the beginning of life, since the impression of implausibility only occurs when wider, e.g. ethical considerations are enlisted. And for another, it is plainly necessary to differentiate between three questions:

1. How great is the probability that a fully developed human being will develop out of this incipient human organism?
2. Does this human organism have the potential to develop into a fully developed human being?
3. When did this human organism that is now at this precise stage of development, begin to live?

(Three questions)

---

<sup>17</sup>It should be remembered that because of the recourse to biological laws, the biological approach has no validity in such cases in which technical interventions have been executed.

<sup>18</sup>With this, the objections raised by e.g. Dawson (1993, p. 44 ff.) to the “genetic argument” are not cogent here, since neither is the issue of affiliation to a species clarified, nor is the genetic identity drawn on as the only criterion for biological individuality.

It is the third question that the above criterion should answer. Whether this answer is plausible must ensue from the following discussion of possible objections and alternatives.

### 3.2.2 *Objections*

Two groups of objections can be raised against the criterion of the beginning of life proposed and explained just now. The first group consists in objections that are intended to show that the envisaged point in time is set too early (Sect. 3.2.2.1). In contrast, the objections in the second group try to prove that the human organism begins to exist at an even earlier time (Sect. 3.2.2.2). The following discussion of these objections has the aim of defending the proposed criterion. In addition, the biological approach will be developed contextually and justified as a whole.

#### 3.2.2.1 *I'll Come Later*

Most of the objections that can be found in the literature are raised against the “Zygotic Principle” (Williams 1990, p. 169), according to which a human organism begins to exist at the instant of the fusion of the haploid genomes, i.e. as a zygote. Since the counterproposals connected with these objections settle the time for later than the criterion proposed here, they are also instructive for my own approach. Furthermore, it is revealing to see for what reasons the diverse objections do not hold water.<sup>19</sup>

- (a) *Fission*: One of the main objections to criteria of the beginning of life which establish the beginning before the fourteenth day after impregnation consists in the reference to multiple embryo formation.<sup>20</sup> There is usually no further differentiation between two variants of this objection. In the first form, the objection submits that even the *possibility* of fission at the point in time  $t_1$  suffices to say that a human organism cannot be an individual at an earlier point in time  $t_0$ . In the second form, the objection submits that if there is a *factual occurrence* of fission at  $t_1$ , then the existence of the individual emerging from it cannot have begun before this point in time.

---

<sup>19</sup>In the literature on personal identity, the familiar thought experiments about fission and fusion have their realist counterparts in this context. For this reason the various solution strategies and thought models known from this discussion also reappear (cf. Quante 1999a for a detailed portrayal).

<sup>20</sup>Cf. on this e.g. Anscombe (1984, p. 111), Bole (1989, p. 650), Engelhardt (1977, p. 228), Ford (1991, p. 100, 111 f., 132 ff. or 171), Grobstein (1988, p. 60), Kuhse and Singer (1993, p. 66), Warnock (1990, p. 229) or Wilson (1999, p. 78).

- (i) There are various defensive strategies one could come up with against the objection to *possible* fission. The right one is that a non-actualized possibility does not – with hindsight – represent an objection. If a human organism has not factually divided in the time period from  $t_0$  to  $t_1$ , then there is no reason why it does not persist during this time period even if the possibility of this fission might have been present (see below Excursus 3.2).<sup>21</sup> Another reaction propounds that a later launching would be implausible because of the continuity of the development process, since from the moment when the multiple embryo formation can be ruled out there can be no “qualitative jump” (Bodden-Heidrich et al. 1997, p. 90) comparable to the nascence of the zygote. Such a later launched demarcation must therefore remain “to a certain extent arbitrary” (Moraczewski 1983, p. 304).
- (ii) If the objection to the criterion being defended here is raised with reference to factual multiple embryo formations, then it must be allowed. For these exceptions it is accepted that multiple embryo formations only emerge at the moment of fission. In such a case, in my view, three entities exist in twin formation. The first (A) exists from the moment the new genome is activated. The other two (B and C) begin to exist when, after the fission of A, each respective genome has activated. But since multiple embryo formations are the exceptions in human beings, they fall under the *ceteris-paribus*-conditions and can be dealt with by the biological approach (see Excursus 3.2).

*Excursus 3.1* Starting with the treatment of the fission problem in the literature on personal identity, there are further conceivable reactions. Most notably, these are resorted to by theologians working with the idea of ensoulment. If one imputes the indivisibility of the human soul, then one has to find either the moment from which time onward fission of the human organism with which the respective soul is bound can be ruled out (see, for instance, Ford).<sup>22</sup> Or one must interpret factual fissions in such a way that they are compatible with the acceptance that an indivisible soul can survive this event. Tonti-Filippini (1989, p. 43 and p. 50 fn. 3) indicates that one can also interpret fission in line with the model of a plant offshoot as a kind of copying process by which the original organism A continues to exist while a new organism B,

---

<sup>21</sup>This answer is given explicitly by Tonti-Filippini (1989, p. 42) and also implicitly by Williams (1990, p. 175) who includes factual non-fission in his variant of the “Zygotic Principle”. In contrast, Forbes (1985, p. 135) argues – and Stone (1987, p. 818 f.) follows him in this – with the indication that even the possibility of fission is irreconcilable with Leibniz’ law of identity. But this argument confuses identity and persistence and inadmissibly transfers the modal characteristics of the former to the latter.

<sup>22</sup>Without these extra theological premises there is no reason why twin formations should not be interpreted in accordance with the above suggestion that a human organism ceases to exist as two new ones begin to exist. The idea of ensoulment therefore seems primarily to be what makes Ford search for a point in time at which no further twin formation can occur (also see Bole 1989, p. 652 fn. 2). But Ford inexplicably omits to distinguish between possible and factual fission. Tonti-Filippini, on the other hand, rightly rejects the objection of possible fission and seeks a solution strategy for factual fissions, in order to defend the “Zygotic Principle”.

that is likewise ensouled, comes into existence (see also Ashley 1976, p. 127).<sup>23</sup> Moraczewski (1983, p. 306 f.) interprets the process of twin formation thus, that a cell which does not belong to the proper human organism (to the “inner cell mass”) in any case, separates due to a defect and forms a new organism.<sup>24</sup> Billings (1989, p. 122 and 124) subscribes to the thesis that it is already established before the fertilization whether fission will take place, so that in cases of this kind it can be assumed that there is a kind of multiple occupancy. Even when, at the moment of fertilization, there appears to be only one individual, nevertheless the two organisms A and B, which will be parted after the fission event, are already there.<sup>25</sup> In all three ways, the concept of an indivisible soul could be compatible with factual multiple embryo formations, so that in principle not even the factual fission contradicts the “Zygotic Principle”. But as I am not taking up the problematic concept of ensoulment and instead speaking of the human organism, and because there are no empirical reasons for construing one of the twins as the original, I accept factual fission as the end of the existence of the original organism and the beginning of two new human organisms.

*Excursus 3.2* In the context of dealing with the fission problem, there is occasionally differentiation between the case that the occurrence of the fission is genetically determined and the case that it is brought about by external circumstances. Contrary to what Billings (1989) implies, to my knowledge it has not yet been clarified which influences cause a multiple embryo formation. The assumption that it is purely genetic determination is implausible, however. In the literature, the view is often taken that in the first case of genetic determination both the “zygotic principle” and the criterion of the beginning of life being defended here must clearly be unfounded (cf. Anscombe 1984, p. 112 and Quinn 1984, p. 143 fn. 5). An entity with genetic determination that divides at a later point in time is not an “individual substance” (Anscombe 1984, p. 112), but a “special kind of proto-organism” (Quinn 1984, p. 143 fn. 5). Even those who hold on to the “zygotic principle” accept this indirectly, since they endeavor to interpret human multiple embryo formations as instantiations of the second case. So Ashley (1976, p. 127), for example, speaks of coincidental influences that trigger the process. In contrast, both cases are to be treated alike in the framework of the biological approach.<sup>26</sup> For one thing, the intrinsic potential of a split at a future point in time does not gainsay that it is the potential

---

<sup>23</sup>The most likely objection, that we quite possibly have no clue as to who is the copy or rather the offshoot and who the persistent original, is blocked by Tonti-Filippini (1989, p. 43) with the pointer that no ontological consequences can be drawn from an epistemic conundrum. Even if this reposte is accepted, it does divulge the general snag that the concept of ensoulment cannot be dealt with empirically (which does not apply to the biological approach in this form).

<sup>24</sup>With this, Moraczewski’s answer to the fission problem contains implicit recognition of the objection of internal-external differentiation – cf. objection (d).

<sup>25</sup>It is not clear to me which “scientific evidence” Billings (1989, p. 122) is referring to. What is more, he must implausibly look upon the potential for fission as adequate for the factual existence of several organisms. But it is interesting that with his conception he adopts a strategy which is known in the literature on twin formation as the thesis of multiple occupancy.

<sup>26</sup>For this reason, the lack of empirical knowledge regarding the causal factors that lead to multiple embryo formation has no effect on the biological approach.

of *one* organism at the present time (otherwise normal amoebae could not be organisms). And for another, the disparate treatment of the two cases rests on the intuition that the persistence of an entity must depend on conditions that are intrinsic to this entity – in fact the reverse inference holds: if a factor is extrinsic it does not endanger persistence. This “Only-X-and-Y-Principle” is, however, usually rejected by the biological approach, for which reason both intrinsic and extrinsic conditions can count as *ceteris-paribus*-conditions (see also the discussion on objection (d.) below). This intuition about intrinsicness can only be made plausible in the framework of an approach that either confuses persistence and (numerical) identity, or ties up with the first-person perspective.

(b) *Fusion*. Alongside multiple embryo formation and above all with reference to animal experiments and future technical options, there is reference to the possibility of the fusion of several cells into a new organism.<sup>27</sup> But even here, no further differentiation is made between the objection to *possible* fusion and the case of *actual* fusion. The principal possibility of a recombination of totipotent cells from various blastomeres to a new organism cannot yet be considered as scientifically verified. However, in the following it will be conceded for the sake of argument. With reference to possible fusions, this potentiality does not itself suffice to deny an existing entity the status of an organism. Counterfactual progressions are not generally sufficient to annul the above criterion of the beginning of life in the case of a factual development that has taken its normal course. If one imputes that a currently unavailable technique has been used, the following applies: if one e.g. distributes to new organisms the eight cells of an eight cell blastomere which regain their totipotency in isolation, the original organism ceases to exist and where applicable new ones come into existence. This is compliant with the above criterion since it applies to normal cases and such technical interventions lie outside its limits. Occasionally, developments of natural fusion are described in the literature. With this a beginning twin formation occurs which, however, then regresses and launches into normal development. Even these cases are not an objection to the criterion of the beginning of life being defended here. For one thing it is not clear whether they can occur after the point in time suggested here. If this should be the case, then either they can be interpreted as anomalous courses by which one organism and then – after fission – several and then – after the fusion – a new organism begins to exist. Or – and this is the alternative that is favored here – the beginning and factually reversible fission is not counted as executed, but interprets the process as fluctuation in the degree of integration of a persisting organism (see Sect. 3.2.3 of this chapter).

*Excursus 3.3* A further special case that only half fits into the context of the objection to the fusion is so-called Siamese twins. This case does not fit exactly, because we normally assume that this concerns two organisms that are not completely isolated from one another or share certain parts of the body. In the context of the

<sup>27</sup>For examples of this objection cf. Bole (1989, p. 650), Ford (1991, p. 139 f.), Grobstein (1988, p. 60) or Kuhse and Singer (1993, p. 67).

beginning of human life this special case plays a role when one counts the placenta as part of the beginning of human life (see e.g. Flaman 1991, p. 43 f. or Fisher 1991, p. 207). So twins share an organ.<sup>28</sup> Even if this special case probably stretches our conceptual apparatus to the limits of its usefulness, one can say from the viewpoint of the biological approach that this poses no real problem. If one accepts that the jointly used body part really belongs to both organisms, then two things should be remembered. For one thing, it is possible that two events occur to a body part simultaneously, even at the same space-time location.<sup>29</sup> For another, it is also conceivable that an event is a constituent of two integrated life processes.<sup>30</sup> Although I prefer the first option to the second one, I find both answers more convincing than Flaman's strategy (1991, p. 44). He maintains that the parts in question might de facto only belong to one organism, but that we cannot recognize or decide this.

(c) *Totipotency*. A further major objection<sup>31</sup> against the “Zygotic Principle” consists in the allusion to the totipotency of the cells of a blastomere which is only lost after the eight-cell stage.<sup>32</sup> If this objection is not understood as an indication of the possibility of fission or fusion, both of which have in fact already been quashed, then it signifies that the totipotency of the cells as such speaks against viewing the blastomere as a self-integrating organism. Against the objection being understood in this way it must be said that the alleged totipotency is on its part only a potential. As long as the cells are part of the blastomere, they show only multifunctionality and no totipotency, otherwise fission would always have to occur. Therefore this is already an integrative achievement steered by the “maternal” RNA and other components of the oocyte which however, in respect of the integration of the individual cells, is reversible when individual potentially totipotent cells are detached from the cluster or stimulated by further factors to form multiples (also cf. Bodden-Heidrich 1997, p. 71 and Flaman 1991, p. 41).<sup>33</sup>

<sup>28</sup> Fisher and Flaman include the placenta with the embryo, so as to meet the objection of lacking internal-external differentiation without taking into account that they then have to meet the objection to fusion in the form of Siamese twins.

<sup>29</sup> Events occupying the same space-time location then have to be individuated via diverse causal relations. This is possible even when, ultimately, events are normally individuated via space-time locations. The spatiotemporal coincidence is then a special case which can be resolved with the help of the causal relations. In general, the identity of events – pace Davidson (1982, Chap. 8) – cannot be defined via causal relations, as these relations must themselves be individuated as concrete occurrences via the events accomplishing them; on this, cf. Quine (1992).

<sup>30</sup> This is conceivable because the special features of the first-person perspective, which lead to the Cartesian intuition that a mental event can essentially only belong to exactly that subject whose experience it portrays, do not figure within the framework of the biological approach

<sup>31</sup> This can be found e.g. in Bole (1989, p. 649 f.), Ford (1991, p. 137 f.); Kuhse (1990) or Kuhse and Singer (1993, p. 67 f.).

<sup>32</sup> Whether this objection even applies to the criterion suggested here depends on whether the totipotency gets lost after the “switch on” of the genome or beforehand. But since the objection is generally invalid, the lack of empirical knowledge has no effects on the biological approach.

<sup>33</sup> With this, nothing is gained in the ethical question of whether one may do research with the totipotent cells released out of the blastomere or what moral status they have.

- (d) *Internal-external differentiation*. A further objection that is championed above all by Ford (1991, p. 146 ff. and 156 ff., cf. also Warnock 1990, p. 229), rests on the observation that the demarcation between the human organism and its environment is not clearly detectable in the early developmental stages. Against this it must first be argued that demarcation or incorporation do not pose a problem for the identity of an organism as long as it can be interpreted as the result of her own integrative achievement, as is the case here (cf. also Tonti-Filippini 1989, p. 45). Secondly, it can be registered that the human organism segregates itself as a self-asserting unit from her environment, above all the organism of the expectant woman (cf. Bodden-Heidrich 1997, p. 78). Thirdly, our concept of the organism tolerates a degree of vagueness since the integrative achievement and thereby the degree of integration of the individual components can vary.
- (e) *Hydatidiform mole*. The hint at the possibility that only a placenta and not a human organism emerges from a zygote is used as an objection to a zygote already being a human organism (Bedate and Cefalo 1989, p. 644; cf. also Ford 1991, pp. 82 and 157). However, this – as any other indication of possible aberrations – is not a valid objection (cf. also Suarez 1990). When normal development has occurred, then in accordance with the criterion being defended here, this human organism began to exist when the genome was activated. If there is aberration such as the hydatidiform mole, the life of the human organism ends due to a genetic defect at an early stage.<sup>34</sup>
- (f) *Parthenogenesis/cloning*. With the possibility of parthenogenesis which, however, is not normally successful in humans (cf. Morowitz and Trefil 1992, p. 52), and cloning a human individual, a further objection is formulated which does not apply to the criterion being defended here. Since there is no demand for a human organism to emerge from the activation of a genome which in turn has arisen through the fusion of sperm and egg, the objections draw a blank. With reference to the question of the beginning of life, the problem posed for clones is analog to that for normal zygotes (cf. Ford 1991, p. 149 f.).<sup>35</sup> Neither would pathogenesis pose any new problems regarding the beginning of life if it could proceed successfully via technical aids (cf. Quinn 1984, p. 144). In both cases the moment must be found at which the active self-regulation begins.
- (g) *Anatomical criterion*. Against the fixing of the beginning of life suggested here it is also argued that e.g. a zygote or a blastomere has nothing much in common with a normal human organism. Therefore, one should take an anatomical criterion to the effect that the human organism only begins when it has developed specifically human characteristics, i.e. in about the 24th week of development (cf. Becker 1981, p. 32 and Morowitz and Trefil 1992, p. 9, 16, f., 62, 74 or 81).

---

<sup>34</sup>The discussion of the empirical findings by Bedate & Cefalo being held between Suarez (1990) and Bole (1989) suffers from their shared assumption that the differentiation between purely intrinsic and extrinsic factors is relevant to the persistence issue. Since this condition does not play a role according to the biological approach to human persistence, I will not go into it further.

<sup>35</sup>Tonti-Filippini (1989, p. 42) aptly emphasizes that no problem ensues from the existence of copies for the master copy of the clone.



Against this one would argue not only that it would be difficult to define accurately an approach to our normal commonsensical idea of a human organism. From the biological standpoint, a protozoa can also be an organism (Bodden-Heidrich et al. 1997, p. 35 f.), and the recourse to commonsensical evaluative criteria of the standard examples of fully developed organisms is implausible. But, above all, both Becker and Morowitz & Trefil speak of stages in the development process of *an* entity – and it is hard to see who this entity is supposed to be if it is not the human organism that goes through her various stages of development here (cf. Sect. 3.2.3 of this chapter). Ultimately, both these approaches are affected by an ethical problem, because they also want to settle the moral status together with the beginning of life. So in their book, Morowitz & Trefil are looking for the point in time from which the development of the human organism differs from all other biological species (they call this the “humaness”), because they assume that the members of the human species have special moral status. If one frees oneself of the relation between the beginning of existence and moral status, there is no longer a plausible reason for the anatomical criterion.<sup>36</sup>

- (h) *Brain life*. With his brain life approach, Lockwood has suggested an alternative criterion for the beginning of life, which shall be discussed briefly here, as it contains implicit criticism not only of the criterion being defended here, but what is more, challenges the entire strategy of the biological approach. In contrast to the supporters of the anatomical criterion, Lockwood assumes that the question of the beginning of life can be answered in an ethically neutral way (Lockwood 1990, p. 245). However, with that he is searching for the beginning of the “human being”, not the human organism. For him they are different entities (Lockwood 1985, p. 11 and p. 19; Lockwood 1988a, p. 200 and Lockwood 1990, p. 246).<sup>37</sup> His answer is that the human individual, in contrast to the human organism, begins to exist as soon as the brain evolves as the material substrate of psychological capacities that are vital for personhood. Hereby there is, parallel to brain death (cf. Chap. 4 on this) a symmetrical criterion for the beginning of life: brain life (Lockwood 1985, p. 20). Lockwood’s conception presents a comprehensive alternative to the biological approach and can only be invalidated by the comprehensive evolvement of the latter. Hence, at this point only the main differences shall be brought to mind. In general, complex analyses of human persistence that work with the concept of person prove to be problematical (cf. Chap. 2). This also applies to modified approaches that only draw on a body criterion restricted to the brain. Since the concept of person does not define a natural kind, neither does it deliver causal laws that regulate

---

<sup>36</sup>Since Becker and Morowitz & Trefil interpret the moment at which “humaness” is attained not only as a morally decisive threshold, but also at the same time as the criterion for the beginning of life, their approach resembles the moral strategy that is adopted in the debate about a suitable criterion of death e.g. by Hoffman (cf. on this, Sect. 4.1 of the next chapter).

<sup>37</sup>Hence, Lockwood’s argumentation corresponds to the one in the context of the debate on the criterion of death named “ontological strategy” (cf. Sect. 4.4.2.1 of the next chapter on this).

the persistence of the entity in question. So Lockwood, when he is to determine the persistence of the brain, is implicitly referred back to the human organism, since the nascent process of a specific brain can only be grasped at this level. Above all, every theory that treats the human organism and the human person (the “human being”) as two separate entities runs into the problem of having to define the relation of these two entities to one another. If, on the other hand, the concept of personhood is treated as a phase sortal, this difficulty does not arise. And at the end of the day the normal human organism has the potential to develop into an organism that yields her integrative achievement with the help of the brain (see also Ford 1991, p. 81 f.). Without the preconditions of specific assumptions of certain variants of complex theories of personal identity, this development step does not indicate a break that must or can be interpreted as the evolvment of a new entity. Even the symmetrical concept of brain death and brain life alone is not an argument, though maybe theoretically attractive.<sup>38</sup> Lockwood’s brain criterion is thus, while not indebted to the ethical aspects of the concept of person, nevertheless the result of the use of the concept of person in the context of the persistence issue, since it is searching for the “seat” of the capacities central to personhood. So this approach is, similar to the ethical or ontological approach in the context of the debate about the definition of death, a theoretical alternative, which can be evaluated only through a comprehensive comparison of its merits.

### 3.2.2.2 I Was Here Already

Above all, it is the defenders of the “Zygotic Principle” who have introduced counter-arguments against the various suggestions that base on the objections discussed in the last section and which could in part be used to defend the criterion proposed here. My proposal dates the beginning of existence too late from the viewpoint of the defenders of the “Zygotic Principle”, so in this section a discussion of the reasons why it is implausible to bring the moment of the beginning of life even further forward will follow. For this, three alternative conceptions will be confronted with the approach being defended here.

- (a) The *zygote conception*. The suggestion that the human organism begins to exist as a zygote is rejected by the biological approach because, while the new individual genome of this particular organism does exist with the fusion of the haploid genome, it does not take up self-regulation. Since the organism is understood here as a self-integrating life process and life itself as an activity and not understood as a property or merely the capacity for exercising it, the mere existence of the new genome does not suffice.<sup>39</sup>

---

<sup>38</sup>When, in the next chapter, the criterion of whole brain death is defended, in no way is the brain life criterion being construed from it (cf. also the discussion of the fourth objection in Sect. 4.5.1 of the next chapter).

<sup>39</sup>That the “switch on” of the genome and the final loss of the up to that moment still possible totipotentiality of the blastomere cells presumably occur together, also speaks for the criterion for

- (b) The *Ootide conception*. The suggestion that the beginning of life should be dated even before the union of the haploid genomes (Bodden-Heidrich et al. 1997, p. 67) is justified in that with the completion of the second fission, at the pronucleus stage, the genetic uniqueness of the newly developing human being has already been fixed. Against this suggestion two things can be said. For one, the criterion of being fixed is unclear. If one e.g. assumes strict causal determinism, then in a certain sense everything has already been “fixed” at the beginning of the existence of the universe. It seems to me imprudent to burden the criterion of the beginning of life with metaphysical ballast of having to introduce a plausible qualification. But above all, this argument is based on the false inference from the direct imminence of an event to the occurrence of the event itself: An imminent union is not a union that has already occurred, even if the result that would ensue from that union is fixed beforehand. The new genome does not yet exist even if the whole situation is preset so that only a certain genome can emerge. This becomes clear in that e.g. a technical intervention during the process can result in the fusion of the haploid genomes not taking place. Here it is not the case that an already existing human genome is destroyed, but it is prevented from beginning to exist. Broadly speaking: from the fact that it is already fixed at  $t_0$  how an imminently developing X will be constituted it does not follow that X already exists at  $t_0$ .<sup>40</sup> For this reason, “it is of no consequence for uniqueness whether the haploid genomes contained in the male and female pronuclei are initially separate” (Bodden-Heidrich et al. 1997, p. 67), but to the contrary for the question of the beginning of life.
- (c) The *gametes conception*. A further suggestion to bring forward the beginning of existence has been submitted by John Harris: “An individual’s life begins when one of the gametes from which that individual will develop is first formed” (Harris 1998, p. 77). Harris defends this at first glance more than surprising suggestion in three steps. *Firstly*, he confesses that his “Gametic Principle” (ibid. p. 79) is exposed to the objection that egg and sperm cell are obviously two

---

the beginning of life proposed here. This is not the case because thereby the objection of totipotentiality is dispensed with as being – as shown – invalid, but because there is a case for the new quality of self-regulation resp. integration.

<sup>40</sup>As regards the bringing forward of the beginning of existence, in the text by Bodden-Heidrich et al. there is a certain tension that indicates that the authors are uncertain or at variance here. Later on there is discourse to the effect that the “self-regulation of the embryo probably occurs at the pronucleus stage, but at the latest in the zygote (Bodden-Heidrich et al. 1997, p. 78 f.). Here, the authors hark back to the zygote conception and reduce the question of the beginning of existence to one merely of definition (ibid. p. 77). In all, the ethical and political interest in bringing even the oocyte under the protection of the German embryo protection law is consistently noticeable in the contribution by Bodden-Heidrich et al. Regarding the repeated allusion to the continuity of development that would make a later onset criterion for the beginning of life implausible, let it just be said at this point that Morowitz & Trefil (1992, p. 43 f.) use precisely this argument to justify as late a moment as possible. One can see clearly in both cases how effective existing ethical suppositions are, independent of the ontological problems, and how they lead to implausible conceptions.

different entities that are neither causally nor spatiotemporally connected.<sup>41</sup> This objection is countered by Harris with the hint that the “Zygotic Principle” is no better, whereby he refers to the possibility of twin formation and fusion (ibid. p. 80). This defensive strategy is unconvincing for two reasons: for one thing, one cannot simply deduce that one has to estimate a criterion for the beginning of life that dates the beginning of existence later, as e.g. Ford does. For another, we have seen that the objections Harris draws on here are not even sound (see also Stone 1987, p. 816 f. fn. 3). With his suggestion, Harris draws *secondly* on a specific principle of potentiality:

Something has the potential to become a zygote, and whatever has the potential to become the zygote has the potential the zygote has (Harris 1999, p. 298).

In a general form, this principle of potentiality then reads:

A has the potential for Z if, when a certain number of things do and do not happen to A (or to A plus N), then A or A plus N will become Z (ibid. p. 299).

Against the again obvious objection (cf. Finnis 1996b, p. 50) that with this he does not distinguish between the potential of an organism to develop and the potential of a set of entities and limiting conditions to build an organism, Harris simply insists that he takes for granted just this potentiality criterion and no other. Whilst I would like to postpone the detailed explanation of the problem of potentiality for the moment (cf. Sect. 3.3 of this chapter), it is already apparent here that Harris’s position for one thing shows difficulty in restraining the candidates for “A plus N” plausibly so as not to allow the principle of potentiality to run idle. For another, it becomes clear that one can accept the potentiality implied by Harris without already being committed to the gametes conception. If one makes a distinction between the two kinds of potentiality, then one can accept the potential of a constellation of entities and constraints to lead to the existence of a new organism without having to count the beginning of the existence of these entities themselves as already being the beginning of the existence of that organism. And, since our concept of a human organism is irreconcilable with the causal and spatiotemporally separate form of existence of egg and sperm, everything speaks here for a differentiation between the kinds of potentiality rather than a softening of the concept of the organism.<sup>42</sup> The reason Harris holds on to his principle of potentiality is tied up with his *third* argument for the “Gametic Principle”. Persons are distinguished by their ability to

<sup>41</sup>This obvious objection is raised by e.g. Ford (1991, p. 78) and Anscombe (1984, p. 113).

<sup>42</sup>Stone (1987, p. 816) also raises the objection against the gametes conception, that it is incompatible with our concept of identity, that both sperm and egg are identical with the later human organism without being identical with one another. Further, he points out the problem that it is possible for a specific sperm to fertilize a different egg so that this sperm must then be identical with two different organisms. Whereas I view the first of Stone’s arguments as plausible, the second one bases on mixing the modal properties of the identity relation with the criteria for persistence and is therefore not taken up by the biological approach. But Harris himself, similarly to Hare, undermines the difference between factual and future possibly existing entities and thereby also mixes persistence and modality in a problematic way (cf. the fourth section of this chapter on this).

assume an evaluative stance to their own existence (Harris 1985, Chap. 1 and Harris 1999). What is more, their pre-personal identity is closely associated with their genome: “Pre-personal identity is the relation between individuals and their DNA” (Harris 1998, p. 85). It is unarguable that a manipulation of the gametes at  $t_0$ , from which an actual (at  $t_2$ ) existing person once (at  $t_1$ ) emerged, could have had effects on this person, so that one could have injured or benefited this person by e.g. damaging or repairing the gametes.<sup>43</sup> So, at this point, one can agree with Harris that the genetic constitution enters into the personality, which Harris (1998, p. 85) refers to as “a person’s life story”:

The history of Tom and Dick is the history of their relationship to their DNA in the form in which it makes them the persons they are and will become (ibid. p. 83).

With this, he has all he needs to justify his criterion of the beginning of life:

The history of A goes back to each of the gametes from which A derived and can, as we have seen, be crucially affected by what happens to those gametes (ibid. p. 82).

At this point it becomes clear that Harris’s deliberations are not based on ontological, but ultimately on moral considerations. For one thing, his refusal to make a difference between the potential of a set of entities and conditions to produce an organism and the development potential of this organism itself, in as far as it has emerged factually, can only be rendered plausible when one inquires exclusively about the moral relevance of this potential. Harris always discusses potentialities in the context of the question of moral status, as the following passage bears out:

The potentiality of something, or some things, has moral importance on the assumption that actualizing a particular potential is what matters. We would not worry about what precisely it is that has the potential to be a person, or an adult human being, if persons or adult human beings did not matter. We are only interested in the potentiality argument because we are interested in the potential to become a particular, and particularly valuable, sort of thing (Harris 1999, p. 299 f.).

With this remark Harris blocks Finnis’ objection. His argument can be reconstructed thus: If the factual existence of a person X at  $t_2$  is taken for granted, one can damage the interests of this person existing at  $t_2$  just as well counterfactually by destroying the potential of the zygote (at  $t_1$ ) as by damaging the potential of the gametes (at  $t_0$ ). Since, factually, this sperm and this egg have led to the development of this human individual, both potentials must be treated alike from the moral viewpoint. And since potentialities are only of interest in terms of their moral value, there is no reason to differentiate between the two kinds of potentialities. This is the first step of the argument. The second rests on pushing the identity of a human person as far back as there are causally defining factors that were factually necessary for the development of this person. For this person is interested in these factors in

---

<sup>43</sup> Counterfactual since the factual existence of the person in her actual constitution is assumed in the description. Furthermore, Harris must assume that e.g. the damage to the gametes is such that the „cross-world-identity” of the factually existing person is conserved with that which existed in the counterfactual scenario. This does not seem to me to be a trivial condition (cf. Sect. 3.4 of this chapter).

the retrospective perspective of her biography. And, thus the third step, because of the interweaving of the identity of the person and her genome, she takes a particular interest in the fate of that genome from which her own DNA has evolved.

From the viewpoint of the biological approach, Harris's argument bases on the false strategy of making ontological demarcations dependent on questions of moral status and evaluation respectively. Firstly, it is thus doubtful whether one should really only allow the moral value to count as a differentiating criterion for kinds of potentiality. As will be shown shortly, it is in fact contrariwise the case that only the waiving of running together of moral evaluation and ontological problems allows a plausible conception of the potentiality of human organisms to be developed (cf. Sect. 3.3 of this chapter). Secondly, Harris uses the concept of personhood implicitly in order to answer the ontological question of the beginning of a human life. However, this leads to implausible results in the framework of both a moral and an ontological strategy. The not yet united egg and sperm are not an organism but two separate entities. Finally, Harris's conception has, thirdly, the internal difficulty of crediting the individual DNA of a human being with a special function for the identity of the personality, because otherwise the gametes could not be marked out above all other causally necessary conditions for the existence of a specific organism. But in this demarcation he does not start from the human organism (the "pre-person", in his terminology) but from the factually existing person. The gap formed here could only be closed by Harris by his postulating the genetic determination of personhood. Without this strong and surely by Harris unintended thesis, only the genetic individual is individuated, but not the human organism, let alone the person. The concept of person and with it the ethical considerations conveyed through it are then masked.

If one takes these findings seriously, it is apt to choose the biological approach so that the criterion of the moral relevance of potentials can be left out just like that of personality. Freed of these elements, Harris's proposal must, however, still seem implausible since he can only justify it by indicating that the "Zygotic Principle" is also a problematic thesis as regards the beginning of the existence of a human organism, as the "Zygotic Principle" and the "Gametic Principle" do not represent exhaustive alternatives. Anyway, the non-unified egg and sperm are more clearly not a human organism than the zygote, even if the latter stage in the development process should likewise not yet be recognized as the beginning of the existence of a human organism according to the criterion being defended here. Harris's proposal is of interest and has been discussed in detail here not because it is a plausible suggestion for the beginning of the existence of a human organism, but rather, because his argument makes clear where the considerations concerning the beginning of human life can lead when they are not kept free of moral considerations and the concept of person.

### 3.2.3 *The Graduality of Integration: The Hard Problem*

The objections discussed up to now against the criterion for the beginning of life proposed here have not proved valid, and the alternative conceptions have not proved plausible. The main reason the objections could be rejected is that they have not tackled the hard problem which is that the integrative achievement of the human organism increases little by little at the beginning of its existence. Most of those who enquire about the beginning of the human organism accept integration and self-regulation as the decisive criterion.<sup>44</sup> Those who distinguish between the embryo as an “individual multicelled organism” (Grobstein 1988, p. 59) and the so-called pre-embryo point out that the latter does not exhibit the degree of integration and self-regulation that can be expected from an organism.<sup>45</sup>

It is unarguable that the development of a human organism can be described as a process of increasing integrative achievement and self-regulation (cf. Boddend-Heidrich et al. 1997). Due to the fact that active self-regulation only begins with the activation of the own genome of a human organism, the thesis has been defended here that this moment should be accepted as the beginning of the human organism’s existence. Before that, we have a set of entities and limiting conditions which, though it already segregates itself in its entirety from the organism of the pregnant woman, does not yet take on board her integrative achievement and thus cannot be counted as an organism (pace Boddend-Heidrich et al. 1997, p. 78 f.). After the “switch on” of the genome there is in my view no further event which can be counted as the moment of the beginning of a new human organism (except multiple embryo formations).<sup>46</sup> Because of this gradual increase in the integrative achievement, from the moment of activation of the individual genome of the new human organism, one should speak only of developmental stages of this organism.<sup>47</sup> The distinction between pre-embryo and embryo, which from the scientific viewpoint may well make sense in the differentiation of various stages in development, is misleading if it is associated with the ontological distinction between two numerically different entities.<sup>48</sup> However, it was precisely in the context of the ethical debate about the

---

<sup>44</sup>Cf. e.g. Fisher (1991, p. 226 fn. 87), Flaman (1991, p. 40 and 42 f.), Ford (1991, p. 146, 168 und 170) or Tonti-Filippini (1989, p. 41).

<sup>45</sup>Cf. also Grobstein (1988, p. 58 f.). With reference to the so-called pre-embryo, Warnock (1983, p. 238 f.) also speaks of a mere cell mass or cell collection (cf. Warnock 1983, p. 241 and 243 or Warnock 1990, p. 215 and 228), although she considers the question of the beginning of human existence to be unanswerable. With this she also draws ex negative on the criterion of integration and self-regulation.

<sup>46</sup>For this reason, an organism cannot survive phases of completely halted integrative achievement. Such an event is tantamount to the death of that organism.

<sup>47</sup>This gradual growth should not be confused with the thesis that an organism comes into existence gradually. As regards this metaphysical question, my above considerations remain neutral (cf. the discussion on this in Quinn 1984, p. 149 ff.).

<sup>48</sup>Strictly speaking, according to the criterion being defended here, the beginning of the existence of the human organism lies within the stage which is otherwise called the pre-embryo, so that the above remarks apply only to the period that comes temporally after the beginning of existence.

permissibility of experiments on embryos, as the study by Mulkay (1997) has shown, that the distinction between pre-embryo and embryo was not understood as the difference between two developmental stages of one entity,<sup>49</sup> but as the difference between numerically different entities, which has substantially influenced the course of the debate. This shows not only how important it is to define the question of the beginning of human existence precisely, but also that the difference between human life that is not yet an organism and human organisms in general is seen to be ethically significant (cf. the next two sections of this chapter). For this reason it is, however, consistent that human life (e.g. cells or tissue) is to be considered ethically relevant even before and independently of the existence of a human organism, the thesis is that the various developmental stages of a human organism have effects on its moral status and ethical claims.

So far, Mary Warnock's thesis according to which the question of the beginning of the existence of a human life is "either unanswerable or unintelligible" (Warnock 1983, p. 238) has been opposed for good reasons. With the exception of the thesis that factual existence makes an ethical difference as opposed to the possibility of coming into existence, the above remarks have been borne by the effort to keep ethical aspects out of the question of the beginning of life of the human organism. Before this strategy is applied with reference to the question of the potentiality argument, I would like to close this long section with a final elucidation. In his elaborate criticism of Ford's argumentation, Fisher (1991, p. 203 ff.) points out that scientific facts alone are not sufficient to answer the question of the correct criterion for the beginning of life. More exactly, it is always a case of interpreting and evaluating these facts correctly. Morowitz and Trefil (1992, p. 6) also speak out in this regard when they write that the question of "When does an individual life begin?" cannot be answered with scientific methods. One must understand these statements thus: that science, or more exactly: biology alone cannot answer this question. But rather, advanced metaphysical assumptions whose plausibility is proven only through philosophical analysis and argumentation are called on for such an answer. This concession must not, however, lead to the misapprehension that with this the biological approach does base on a moral foundation after all. The evaluations and interpretations that unavoidably play a role in the justification of a criterion for the beginning of life are not ethical or even normative, as would be the case when one operates with the concept of person or "human being" in opposition to "human organism". On the contrary, it concerns interpretations and evaluations of scientific facts in the light of general conceptual and metaphysical assumptions. And this kind of evaluation is not only compatible with the analysis of human persistence defended in this study, but actually required by it.

---

<sup>49</sup>The thesis that the distinction between pre-embryo, embryo, human creature and person concerns the developmental stages of an entity, can also be found in Billings (1989, p. 120), Carter (1982, p. 91), Grobstein (1988, p. 61), and above all, Quinn (1984, p. 143–149).



### 3.3 Persistence, Potentiality and Personhood

In hardly any other area of philosophy is the interference of metaphysical and ethical issues so pronounced as in the context of the debate on abortion, in vitro fertilization or experiments on human embryos. The reason for this is the triple use of the concept of person to signify members of the human species, to indicate an exceptional ethical status and to answer the question of the beginning and end of the existence of an individual. If one, along with Beauchamp (1999), distinguishes between a moral and a metaphysical concept of person, then hope rests on the latter, that by means of the declaration of a specific list of properties and capacities the question of the moral status of an entity can be answered, and with it the question of the permissibility of certain actions towards this entity.<sup>50</sup>

Allowing for the assumption that human organisms do not possess the properties and capacities necessary for personhood in the first phase of their existence, the discussion about the moral status of human life in this early phase speedily shifts to the question of whether the potential ability to develop into a person should be accepted as morally relevant or not. The argument being kindled here marks the point at which the considerations of the biological approach and the remarks in this chapter become relevant for the discussion. Such relevance must be surprising at first glance, since the biological approach for one thing simply puts aside the concept of person and for another, leaves out ethical issues through limitation to the observer perspective. Indeed, with the remarks in this section it is primarily the limitation of the biological approach for questions of practical philosophy that surfaces. But because some ineligible counterarguments are dispelled (Sect. 3.3.2) and the possibility of controlled use of the potentiality argument is sketched on the basis of the biological approach (Sect. 3.3.2.2), indirect ethical relevance does arise. However, to make this visible, the ethical premises stemming from the debate on the ethical status of beginning life and also on my own reflections must be addressed in a first step (Sect. 3.3.1).

---

<sup>50</sup>For a general overview of the application of the concept of person in the debate on abortion cf. Ach (1993), English (1975) or Macklin (1984). As Birnbacher (2001) describes, the attempt to justify an exceptional moral status via the concept of person leads to the concept of personhood via a list of properties and capabilities. Such a conception can be found, for example, in Harris (1985, 1999) and Tooley (1983, 1998). However, in an early study, the latter also used the concept of person normatively, and this usage must be strictly distinguished from the metaphysical usage. Whereas in the metaphysical usage the capabilities accompanying personhood are drawn on as a justification of the moral status, the concept of person in its prescriptive usage represents all entities that have a certain moral status without the implication of a justificatory relation (cf. Tooley 1990, p. 159, where Tooley speaks of a “pure moral concept”).

### 3.3.1 *The Bioethical ‘Guillotine’*

If one asks oneself why, above all in the context of the debate on abortion, there is such a bitter struggle over the concept of person and the permissibility of its application to human life in its earliest stages of development, then one soon discovers a link between the status of being a person and having a right to live.<sup>51</sup> This link – I call it the bioethical ‘guillotine’ – can be presented in the form of two syllogisms.<sup>52</sup>

(P 1) Only actual or potential persons have a right to live.<sup>53</sup>

(P 2) All human beings have a right to live at every moment of their lives.

---

(C 1) All human beings are potential or actual persons at every moment of their lives.

(First syllogism)

(P 3) Only actual persons have a right to live.

(P 4) Not all human beings are persons.

(P 5) All human beings are not actual persons at certain stages in their individual existence.

---

(C 2) There are human beings who at no moment of their existence have a right to live, and there are phases in the existence of every human being in which they have no right to live.

(Second syllogism)

On the one hand, these two conclusions differ from one another in respect of the evaluation of the potential ability to become a person, while on the other hand they also share two premises:

P\* There is a special claim: the ‘right to live’.

P\*\* This right to live, including the question of potentiality, is accorded only to persons.

(Shared premises)

---

<sup>51</sup> For a detailed critique and refutation of this widely accepted way of argument see Quante (2014b, Chapter IV).

<sup>52</sup> As this does not concern questions of animal ethics, the problem concerning the limitation of the concept of person to members of the biological species: human being will not be expounded further; cf. Birnbacher (2001). According to my own analysis there are, however, no conceptual or logical reasons for attributing personhood only to human beings. However, this extra complication can be ignored for the purpose of this study.

<sup>53</sup> An entity X is an actual person iff it exists actually and can be ascribed personhood and personality; X is a potential person iff X exists actually and has the capacity to develop into such a stage that personhood and personality can be ascribed to X then (this means that X actually isn’t a person but already an existing individual human organism).

I do not go along with either P\* or P\*\*. In the first place, the discourse about rights is ambiguous. In its strong meaning one understands the right to live as a categorical ban on carrying out homicidal actions, whereby that also applies to each individual. The right to live would be inalienable and the desire for active euthanasia therefore ethically inadmissible. Within the framework of philosophical ethics, such an understanding of rights seems to me not to be plausibly justifiable, since it has to make use of theological or such metaphysical premises as are not able to be made generally binding in a plural society (cf. Quante and Vieth (2002) on this, see also the discussion in Chap. 6). In its weak meaning, the discourse on a right to live signifies that it concerns a *prima facie* rightful claim which can only be overruled by other ethically relevant claims. Speaking of a right rather than a claim can be understood as an indication that a right to live could only be overruled when high-ranking ethical claims or goods oppose it. To leave out the connotation of the categorical and inalienable right to live that cannot be weighed up against other ethical claims, I will speak of ethical claims in the following and abstain from discussing rights.

If one imputes only a weaker concept of right, P\*\* becomes implausible, for now why other living creatures than persons should not *prima facie* have an ethical claim on their life not being ended cannot be made intelligible. If one presupposes a continuum of ethical claims, then it is not clear at first glance why a qualitative ethical leap should be necessary.<sup>54</sup> Reconcilable with this result is the assumption that special ethical claims as regards voluntary euthanasia or the autonomous disposition of one's own future can be construed as a consequence of the capacities pertaining to personhood, (cf. also Chaps. 6 and 7). In my view, personality and personhood are indeed both directly and indirectly – via the concept of autonomy (cf. Chap. 5) – relevant for questions of (biomedical) ethics. But the moral status of an entity is neither limited to a right to live – understood in the way I do, nor does the fact that an entity is entitled to ethical claims imply that this entity is a person.

If one rejects the two premises that are accepted by both sides of the debate then, for one thing, one dismisses the concept of person from the discussion on the ethical status of the beginning of human life and for another, avoids the situation of an all-or-nothing decision. In this way, the status of the potentiality of an entity to become a person contested between the two camps becomes tractable, because it is now no longer linked with an absolute difference in ethical status. The potential of an entity X to develop into a person can, in my view, be justified by an ethical claim by X when (and then: because) this potential is, firstly, realized in the complexity of X and, secondly, it is directed to the formation of capacities that attain positive ethical evaluation. The plausibility of this ethical thesis, which I cannot further justify here, depends not only on a decoupling of personhood and categorical rights, but also on

---

<sup>54</sup> Birnbacher (2001) and – with reference to Tooley (1983) – Leist (1990, p. 140 f.) also object to P\*\*. The above comments do not purport that a champion of the thesis that only persons have a special claim on life could not have further ethical or metaphysical arguments to offer. But I maintain that the attempts to establish a categorical difference between persons and non-persons are futile; cf. Quante (2014b, chapter I and IV).

the possibility of invalidating other objections which have been asserted against the potentiality argument.

### 3.3.2 *The Debate on the Potentiality Argument*

It is not the aim of this section to adjudicate in the debate over the ethical value of the potential of a human organism. But rather, to attempt to expose misunderstandings or respectively differing notions about what potentiality should be understood as, so as not only to make the potentiality argument manageable for the ethical problem in the first place, but also to reveal the “precise nature and source of the disagreement” (Buckle 1993, p. 107). As Buckle already remarked, this will certainly not be enough to settle differences of opinion, but it can help to divert attention to the really problematic aspects. For this, in a first step, the three main objections to the so-called potentiality argument will be analyzed (Sect. 3.3.2.1); in a second step, a description of how potentiality can be managed in the framework of the biological approach will be given (Sect. 3.3.2.2).

#### 3.3.2.1 Three Main Objections

The three objections, of which the first two surface almost stereotyped in the literature, are the logical-fallacy-argument, the reductio-argument and the argument of graduality. They will be introduced briefly and examined for their eligibility.

- (a) *The logical-fallacy-argument*: This argument accuses the supporters of the potentiality argument of a logical fallacy which is diagnosed as follows: From the potential of X (at  $t_0$ ) to become f (at  $t_1$ ) it is concluded that X (at  $t_0$ ) is already f. This fallacy is hidden because the argumentation proceeds not with the property or capacity f directly, but with the moral status which is justified by the possession of f. In other words, the supporters of the potentiality argument equate the moral status of an X with actual f and an X with the potential of becoming f. The argument against this is that “the fact that something will become x (...) is not a good reason for treating it now as if it had become x” (Harris 1983, p. 223).<sup>55</sup> Harris substantiates this thesis by pointing out that the fact that each one of us will be dead one day does not mean that we should be treated now as if we had already died.

The logical-fallacy-argument applies to applications of the potentiality argument in which the moral value that an X’s actual being F has is transferred to an X which actually isn’t F but has the potential to develop such. This transfer is taken to be justified either for purely logical or semantic reasons, or, due to the principle of

---

<sup>55</sup>The logical-fallacy-argument can also be found in Harris (1985, p. 11; 1998, p. 50 and 1999, p. 297) and in Engelhardt (1974, p. 224) or Warren (1998, p. 131).

universalization, because of the identical moral value of the potential ability to become *f* and of being *F*. In this case there really is a logical fallacy, because it is not borne in mind that there is no identity between the potential of becoming *f* and the actual being-*F*.

But there is a more plausible version of the potentiality argument available. The suggestion is that we should evaluate the potential to become *f* as being morally the same as the actual being *F* (cf. Steinbock 1992, p. 59). What one can reproach the supporter of the potentiality argument in this version with is not a logical fallacy but the lack of ethical justification for this assessment. It is certainly logically possible that two different things can have the same ethical value, and from the counterexample cited by Harris it merely follows that this is not so in every case. The logical-fallacy-argument is thus limited in its scope and in many cases comes down to a value disagreement as regards the question of what value should be attributed to the potential to become an *f*.

(b) *The reductio-argument*: This argument states that the equation of the moral status of an actual person with an entity, e.g. the zygote which has the potential to become a person, leads to implausible conclusions. For, since the gametes also – themselves have the potential to become a zygote and with it – qua transitivity – also the potential to become a person, then the gametes must also be granted the same moral status. But this is taken to be absurd and, if one also adds certain considerations of justice, would even lead to an obligation to procreate and possibly to contraception or sexual abstemiousness being morally wrong.<sup>56</sup>

What should one make of this argument? For a start, true to the motto: “One philosopher’s *modus tollens* is another’s *modus ponens*” one could, in effect, deduce from the link postulated by the *reductio-argument* that there is in fact an obligation to procreate or to treat gametes as actual persons.<sup>57</sup> If one concedes that this way out is unacceptable, the potentiality argument has nevertheless not yet been invalidated. On the contrary, one is now faced with the task of proposing a demarcating criterion which dismantles the link imputed by the *reductio-argument* (Leist 1990, p. 85). In order to justify such a demarcating criterion, various conceptions and kinds of potentiality must be differentiated and there must be a more precise specification of which potentiality is meant in the potentiality argument (cf. Sect. 3.3.2.2). Thus the *reductio-argument* is itself only of limited scope. It obliges the defender of the potentiality argument to solve the demarcation problem and justify his concept of potentiality. Should such a justification succeed, the implied *reductio ad absurdum* will no longer arise.

---

<sup>56</sup>This argument can be found among others in Harris (1983, p. 223; 1985, p. 11; 1998, p. 50 and 1999, p. 298) and in Sumner (1981, p. 104), Warnock (1990, p. 230) or Warren (1998, p. 131).

<sup>57</sup>As is generally known, this route is taken by Hare (1990) – for a discussion of his approach cf. Corradini (1994). One main reason Hare reaches this contrainuitive result is his refusal to distinguish between possible and factually existing human beings (cf. also Sect. 3.4 of this chapter).

(c) *The argument of graduality*: This argument is proposed almost casually by Mary Warnock (1990, p. 229 f.) and apparently identified by her with the reductio-argument. The objection to the potentiality argument maintains that the discourse on the conditions, due to which an X has the potential to become an f, cannot be specified. With this, a certain propinquity to the reductio-argument is present, for when a limitation in the attribution of potentials does not succeed through specification of the conditions which should be allowed, then the attribution of potentials becomes inflationary and ethically unacceptable consequences are the result. Just as the reductio-argument, the argument of graduality also confronts the defender of the potentiality principle with a demarcation problem. But Warnock submits a further consideration for which reason it is in my view expedient to count her objection as an independent argument. With regard to the conditions that must be present for a potentially fertile garden to be really fertile, she explains that the realization of this potential can demand more or less effort and input. The more conditions have to be fulfilled so that the imputed potential can actually be realized, the less valuable the existent potential will be estimated. So according to Warnock (ibid. p. 230) the value of a potential varies according to the nature of the constraints which are necessary for its realization. Warnock's objection brings with it not only the challenge of differentiating between different kinds of potentiality which differ as regards the respectively necessary conditions. She also quite rightly points out that potentialities permit gradual differences. Of course, taken on its own merit, this is not an objection to the thesis that potentialities should be accorded an ethical value. But if one bears in mind that, most notably in the context of the debate on abortion, categorical or absolute ethical protection – the so-called right to live – is linked with the allusion to the potential of X to become a person, then Warnock's objection gains weight. For now the relation between graduality and categorical protection is strained. The argument of graduality is borne out. Even if one establishes a threshold somewhere from whence all grades of potentiality are treated ethically in the same way, the assumption of graduality is compatible with both our ethical intuitions and our perception of the development of a human embryo. This study takes these findings into account, in that potentiality is allowed an ethical value while a categorical or absolute right to live is not linked thereto.<sup>58</sup>

The three objections to the potentiality principle do not actually lead to this argument being pronounced generally unfit. But they do show, firstly, that further ethical arguments are needed to justify the value of potentials. This task will not be pursued below, as the aim in this chapter is not that of developing general ethics for dealing with the beginning of human life. Instead, I will start from the *prima facie* assumption that certain kinds of potentiality have an ethical value. Secondly, the above

---

<sup>58</sup> Thus, my strategy is similar to the one suggested by Siep (2001), albeit with the difference that it is not the concept of person resp. personhood that is gradualized but the potential to become a person. For a discussion of Siep's suggestion cf. also Birnbacher (2001).

discussion has revealed that with potentials it will be difficult to justify a categorical or absolute right to live. This objection is allowed here. But above all, the objections impose on the defenders of the potentiality principle, thirdly, the obligation to solve the demarcation problem. How this can be done within the framework of the biological approach will now be elucidated.

### 3.3.2.2 Potentiality in the Framework of the Biological Approach

One of the main difficulties of extracting a precise meaning from the discourse on potentialities is that factual ontological and ethical problems have always already been interweaved in the literature.<sup>59</sup> This misleading linkage will be severed in the following.

- (a) *Kinds of potentiality.* In his description, Leist (1990, chap. IV) distinguishes eight different kinds of potentiality with the aid of which one has tried to reject the reductio-argument.<sup>60</sup> To be able to succeed here, it is first necessary to differentiate between the analysis of potentials as probabilities of future states and the analysis of potentials as a kind of causal powers of entities (Buckle 1993, p. 93 f.). The former is, as one can see from Harris's position, unsuitable for usefully undertaking demarcations since, in connection with a transitivity principle, potentials can be assigned arbitrarily. If, on the other hand, one understands potentialities as causal powers, then one must for one thing say to whom these powers are being assigned. And for another, one must distinguish between the various types of such potentials.<sup>61</sup>

In contrast to Harris or Engelhardt (1986, p. 111), who interpret potentiality as the probability that a certain development will take place, Buckle (1993), Stone (1987) and Ford (1991, p. 96 f.) take the potential of an entity as something from which causally determined developments are negotiated internally. Whereas e.g. the potential of a block of marble to become a statue can only be externally realized—through an artist's appropriate actions – the disposition of an organism to develop

---

<sup>59</sup>Harris (1999) e.g. chooses the potentiality concept which accommodates his ethical intuitions that are independent of potentiality. Ford (1991, p. 99) also claims that "the meaning of a potential person needs to be understood in the context of genetically human life and of the above moral concepts". In contrast, the question of the ethical value of the potentiality will be blended out in the following, since the controversy over the potentiality argument only becomes arguable when potentialities can be measured adequately, independently of ethical and other premises which are controversial in the debate.

<sup>60</sup>In addition, he distinguishes between potentiality and identity arguments, a differentiation that I maintain is mistaken, for reasons to be explained later on.

<sup>61</sup>In order to distinguish between the case of a sleeping person and a human embryo, one must understand a potential as the potential to develop a property or capacity in the future. A sleeping person, in contrast, has the capabilities linked with personhood, but is not factually making use of them. Capabilities will be understood in the following as being attributed to an individual at a point in time, whereas a potential consists in enabling a property or capacity to be developed some time in the future; cf. on this problem also Bole (1989, p. 651 fn. 10) and Engelhardt (1977, p. 24).

certain capacities is inherent.<sup>62</sup> Moreover, the potential of an organism is active, whereas the potential of a set of ingredients to become a salad is passive.<sup>63</sup> While, namely, a set of entities (e.g. a sperm in a syringe and an egg in a Petri dish) cannot do anything by themselves to instigate the generation of a new individual, once a human organism has been generated it has the active potential – though dependent on constraints – to initiate the development. While one has to ‘kick off’ the development of a set, it can only be stunted in an organism. The potential relevant to the potentiality argument must therefore be understood as being inherently active. This is attributed to organisms because they are capable of initiating development through self-regulation and self-integration.<sup>64</sup> And since the potentials inherent to an individual organism depend on the nature of its individual genome, the question of when an individual organism begins to exist is significant for the attribution of this kind of potentiality.<sup>65</sup> That is to say, while the active inherent potential to become an *f* presumes the existence of a persisting organism, this simply does not apply to the potential of a set of distinct entities to generate a new individual (Buckle 1993, p. 95). As opposed to Warnock, Lockwood (1988a, p. 202 and 1990, p. 243 f.) also rightly emphasizes that the question of the beginning of a human individual is relevant for biomedical ethics. Even if he thereby presumes a different criterion for the beginning of life,<sup>66</sup> his study nevertheless does contain the correct insight that an

---

<sup>62</sup> Inherence must not be confused with autarky. Many a conception of strong potentiality is implausible because it imputes that an inherent potential must be completely independent of favorable external conditions (cf. Stone 1987); cf. for this criticism also Wolbert (1998, p. 46). In his differentiation between active and passive potentiality, Tooley (1998, p. 122) avoids this error by pointing out that the potential need not be “fully active”.

<sup>63</sup> Cf. on the differentiation between active and passive potentiality also Ford (1991, p. 110) and Baumgartner et al. (1997, p. 230).

<sup>64</sup> The difference between the gametes that are not yet unified and the existing human individual is not decisive because no potentiality could be ascribed to a set of distinct entities. In this respect, the objection by Singer and Dawson (1993, p. 85) is reasonable. But for one thing, the decisive difference for the potentiality argument is that no active inherent potential can be ascribed to a set of distinct entities (cf. Steinbock 1992, p. 65). And for another, e.g. Harris’s consequences would only follow if one could identify the potential of a set of entities with that of the individual entities of this set. Whether this is possible, or whether in fact this is a composition fallacy, as I suspect, requires further discussion (cf. Elster 1981, p. 154 ff.). For the same reason, Leist’s (1990, p. 88 f.) objection to the attempt to solve the demarcation problem by means of the concept of the inherent active potential also fails.

<sup>65</sup> This is also the reason why the differentiation between potentiality and identity arguments undertaken by Leist (1990) is misleading, since a criterion for the beginning of life is exactly what is assumed for the attribution of potentiality referred to in the potentiality argument. Leist even notices this himself in his discussion of this kind of potentiality, when he understands by the recourse to strong potentiality by Stone (1987) “that in fact it means an identity argument on the basis of numerical or spatiotemporal identity” (Leist 1990, p. 89).

<sup>66</sup> The same applies to Engelhardt (1986, p. 110 ff.). Engelhardt’s thesis, that a fetus can only be ascribed the potential to engender something, but not the potential to develop, which is described as a curiosity by Baumgartner et al. (1997, p. 230), can be explained by Engelhardt’s criterion for the beginning of life, according to which a fetus/an embryo is not yet a somebody and therefore not identical with the later person.



ethically useful potentiality argument can only materialize using the active inherent potential of an organism (Lockwood 1988a, p. 208 and 1990, p. 243 f.).<sup>67</sup>

(b) *Conditions of adequacy*: The biological approach provides the theoretical framework within which the potentiality argument can be applied wisely. For one thing, a criterion for the beginning of life is proposed with which the question is answered as to when one can start to speak of a new individual human being. Further, the organism is understood as a self-integrating and self-regulating life process, so that in this way the establishment of the features “active” and “inherent” is assured. And finally, according to the persistence analysis of the biological approach, the development of the organism over time is understood as a causal process based on species-specific biological laws. This permits the character of potentialities to be justified as causal powers and not – as in Harris – just as probabilities. The biological laws also yield contextual limitations which allow relevant and irrelevant constraints to be differentiated. All in all, the demarcation problem in its different forms can thus be solved satisfactorily.<sup>68</sup> Even if the question of the ethical value of this potential has not been answered with this, the prerequisites have indeed been fulfilled which make it possible to conduct such an ethical discussion constructively in the first place.<sup>69</sup> The criticism of the so-called “identity arguments” misses the point if these arguments are not offered as an answer to the ethical question, but serve only as a prerequisite for the controlled application of the potentiality argument.<sup>70</sup>

---

<sup>67</sup>The reason he considers potentiality in connection with numerical identity to be morally relevant however, is justified by Lockwood in that the concept of person from whom the moral worth is derived is being applied. In contrast, the above considerations forego the concept of person and can therefore only be drawn on as prerequisites for rendering the potentiality argument at all manageable for ethical questions. On this point, one must agree with Wolbert (1998, p. 45 f.): The necessary specification of the problematic kind of potentiality does not suffice for this potential to be pronounced ethically relevant.

<sup>68</sup>Singer and Dawson (1993, p. 77 u. 87 f.) admit that although the potentiality argument is usable for normal or natural procedure, it is unsuitable in the context of technical reproduction (cf. also Steinbock 1992, p. 64). This assessment can be borne out. When potentiality is interlinked with the biological approach, then a contextually and methodologically controlled ascription of potentials is not based on “physical possibilities”, as Singer and Dawson (1993, p. 79) believe, but on specific biological regularities. Nevertheless, it applies that these can be overridden when there are technical interventions. In this way the biological approach can also explain why the potentiality argument loses its orienting function in the context of reproduction technologies. Hence the biological approach is bound to the thesis that one cannot ascribe the same potentiality to an embryo in vitro as to an embryo in vivo sans phrase.

<sup>69</sup>Despite the recourse to species-specific biological laws, the ascription of potentialities in the biological approach remains centered on the individual due to the involved criterion of the beginning of life. This means that the potentials specific to the members of this species cannot just be ascribed to a human organism with defective genome (perhaps in the sense of a second rank potential). But since, on the other hand, the affiliation to a species also plays a role, the biological approach at least allows a reconstruction of why we understand the lack of the normal potential as a loss – Wilkes refers to this in a nice turn of phrase as an “Aristotelian loss.”

<sup>70</sup>Thus the criticism by Leist (1990, p. 104 ff.) of the identity arguments is only justified when they are expected to deliver ethical answers at the same time. Leist, like many a defender of the identity argument, is buffalooed by the ambition to answer ontological and ethical questions at one and the same time. In contrast, the advantage of the biological approach consists in the clarification of the systematic bond that exists between the potentiality argument and the issue of the beginning of the existence of a human organism.

- (c) *Potential person?* With this, the potentiality argument is theoretically embedded in the framework of the biological approach insofar as it becomes manageable for ethical questions. The question remains at this point as to what the effects forgoing the concept of person can have on the analysis suggested just now. The immediate effect of the requirements of the biological approach is that neither the concept of person nor the concept of personhood or personality can be applied in its framework. So the potential of a human organism to achieve personhood in the course of its development cannot be encompassed directly. This is not because the social factors necessary for the development of personhood would not be ascertainable. Insofar as they are understood as causally relevant conditions, they count as factors which are subsumed in the *ceteris-paribus*-clauses. However, social constraints cannot be taken into account in the framework of the biological approach as regards the claims for approval that accompany recognitive relationships, as our approach is limited to the causal explanation and the observer perspective. Altogether, only the causal conditions enabling personhood are thus brought into focus. Because this is the case, the potential to develop into a person cannot be embraced solely by means of the biological approach, either. It is rather the case that from the observer perspective the target of being a person or developing personhood must be given. As regards the potential ability to develop personhood, the biological approach provides the answer to a question posed to it by the observer perspective.<sup>71</sup>

Alongside the methodological constraint, general discontent might well remain that in this way the real ethical debate has not even been initiated. This self-limitation is, however, deliberately allowed for by the biological approach, so as to provide the conditions for conducting a constructive debate. The two most important, even more far-reaching aspects for the ethical issue are, first, that this conception of potentiality permits gradualizations, because the potential is realized in properties and capacities that can increase in complexity during the course of an organism's development. This certainly meets our ethical intuitions as regards the problem of abortion half-way (cf. Steinbock 1992, p. 66 f.). And second, the analysis of potentiality points out that a distinction must be made between the factual and the merely possible existence of a human organism. As the following section will show, this has immediate ethical relevance e.g. on the evaluation of human genetic interventions and also plays a role in the context of the dispute about the relation between contraception and abortion.

---

<sup>71</sup>The situation would be different if one could naturalize personality or personhood e.g. on the basis of genetic determinism. But this possibility of a reduction of evaluative to purely natural concepts is challenged in the available study; cf. Quante (2000a), (2013c, Chap. VII) and Chap. 5 below.

### 3.4 The Ethical Relevance of the Beginning of Life

Although the point of the biological approach consists in answering questions of persistence without recourse to the concept of person and ethical intuitions, it does have indirect ethical relevance: For one thing it is possible therewith to retain the ethical asymmetry between abortion and contraception (Sect. 3.4.1). And for another, this approach permits ethically relevant differentiations to be drawn in the context of gene therapy and misleading discussions to be clarified (Sect. 3.4.2).

#### 3.4.1 *Contraception Vs. Abortion: Possible Vs. Potential Persons*

The discussion of the potentiality argument has revealed that meaningful discourse on potentiality, which is not affected by the reductio-argument, is possible when potentiality is ascribed to a persisting human organism. In this way, the implausible consequences of a potentiality concept can be avoided, as e.g. Harris contends. One of these counterintuitive consequences is e.g. the ethical symmetry between contraception and abortion which emerges from the intrinsic moral status of gametes and human embryos (Kuhse and Singer 1994). The thesis postulated here, that abortion and contraception are ethically not on a par, can be defended in opposition to the studies by Harris, Kuhse or Singer in that one differentiates between the various kinds of potentiality and only awards ethical relevance to those which can be understood as the development potential of a factually existing organism.

Hare has tabled objections against this strategy of justifying the asymmetry by means of the potentiality argument in various papers. Different from Harris, Kuhse or Singer, Hare deduces an obligation to procreate out of the symmetry. Although his line of attack is not directed at a reduction ad absurdum of the potentiality principle, Hare's arguments are interesting for the following on two counts: The discussion on Hare's position offers an opportunity to clear up a possible misunderstanding as regards the discourse on potential persons. Thus it can be shown why a supporter of the potentiality argument is not, even under the proviso of Hare's postulated golden rule, forced to construe symmetry between abortion and contraception, let alone have an obligation to procreate. In contrast to the dispute with Harris, the following is not about developing a concept of potentiality which does not lead to counterintuitive consequences, but about the question of whether the fact of the beginning of life possesses ethical relevance. In doing so, the premise that contraception is less questionable than abortion is accepted, whereby contraception is understood as "preventing a human individual from beginning to exist", whereas abortion is understood as "killing an already existing human organism through an external intervention".<sup>72</sup> So this is not about justifying the ethical asymmetry, but

---

<sup>72</sup>This is of course much too broad for a precise definition. But in our context it suffices to mark the decisive difference.

about proof that even when the golden rule is accepted the consequences drawn by Hare do not follow.

To see more clearly here, it is important to introduce a terminological distinction. If the following discourse is about *potential persons*, reference is being made to actual existing human organisms to whom the potential of becoming a person has been ascribed. In contrast, if I am speaking about *possible persons*, this concerns entities that might possibly exist at a future time, but do not yet factually exist.<sup>73</sup> Warnock (1983, p.238 f.) has put forth the thesis that the beginning of the existence of a human individual is irrelevant for the ethical question of how to deal appropriately with emerging human life. In opposition to this, Lockwood defended the position that the bond between the beginning of existence and the potential to become a person are ethically decisive. Though Lockwood thereby for one thing uses the concept of person and an interest-based ethical conception and for another, proposes a different criterion for the beginning of life from the one proposed here, his strategy is still in principle identical with the one I have unfurled in this chapter. So the objections Hare (1988) raises against Lockwood are – irrespective of the differences from Lockwood’s conception – relevant for the biological approach.

Lockwood’s initial thesis, in which potential and persistence are bound together, stated originally:

- (1) A potential for X generates an interest only where there is some individual for whom the development of the potential for X constitutes a benefit (Lockwood 1988a, p. 199).

Against this thesis, Hare (1988, p. 215) raises the objection that one must not speak of “constitutes a benefit”, but instead say “would constitute a benefit”, and points out an ambiguity in “there is” (ibid. p. 215 f.): In one sense, there is no reference to a specific point in time (“tenseless”), whilst in another, such a reference is included (“tensed”). Lockwood (1988b, p. 344) points out that he meant “would constitute” in his original formulation and also explains that he was speaking of an actual existing entity, a potential person, and not of a merely possible person. So (1) should be understood in the sense of the tensed reading, because now the moment of the utterance has been pointed out. Accordingly we get:

- (2) A potential for X generates an interest only where there is actual some individual for whom the development of the potential for X would constitute a benefit.

As Lockwood continues, (2) is now ambiguous as regards the area of quantifiers. Is the “there is actual” in the scope of the “would”, or vice versa? Is (2) meant in the sense of (2a) or (2b)?

- (2a) A potential for X generates an interest only where there is (tenselessly) an actual individual for whom the development of the potential for X would constitute a benefit.

---

<sup>73</sup>For an overview of this debate see the contributions in Roberts et al. (2009).

- (2b) A potential for X generates an interest only where the development of the potential for X would result in there being an (actual) individual who benefited from it.

Hare (1988) argued against Lockwood's original statement that the results of that are symmetry between the treatment of gametes and the human organism and consequently symmetry between abortion and contraception. The ethical significance of the beginning of life postulated by Lockwood could not be conjoined. Lockwood thereupon points out that this can follow only if one imputes reading (2b), whereas he, in contrast, intended (2a). This coincides with the findings arrived at in the previous section, since (2b) evidently takes for granted a conception of potential that is not bound to the factual existence of a persisting organism.

However, what Hare was thinking of as he pointed out the possibility of a tenseless reading of "there is" is in my view not only this difference between the two ways of understanding potentiality. On the contrary, with this he is steering towards his central thesis, according to which actual existence makes no difference in moral respects. That means that he is not really visualizing (2b) but (2) in the tenseless reading:

- (3) A potential for X generates an interest only where there is (tenselessly) some individual for whom the development of the potential for X would constitute a benefit.

In other words, Hare can see no reason to differentiate between potential and possible persons, because in his view both have to be treated equally as regards moral considerations. According to Hare, the real and the hypothetical existence are excluded from consideration in the universalization procedure as being an irrelevant difference (Hare 1995a, p. 359 and Hare 1998, p. 401 f.). So in (3) Hare is committed to the broad conception of potentiality (explicit in Hare 1995b, p. 312), but is not affected by the *reductio*-argument because he indeed accepts the symmetry between abortion and contraception in the sense of an obligation to procreate (Lockwood's reconstruction of Hare's objection is insofar not false, either). The reproach expressed by Lenzen (1995, p. 234) and Schöne-Seifert (1995, p. 213 f.), that Hare does not distinguish between potential and possible persons, spins in the void, as Hare himself points out (1995b, p. 314). He can see the difference, but considers it to be morally irrelevant.<sup>74</sup>

Lockwood (1988b, p. 349) rejects Hare's objection with moral arguments. In his opinion we have stronger moral duties to actual existing entities (potential persons) than to merely possible persons. In my view it is, however, rather pointless to want to encounter Hare's arguments in the field of morals, though his argument (Hare 1995b, p. 311), that no responsibility for future generations can be justified without

---

<sup>74</sup>Another possibility of getting through (3) without committing the theory to quantifying over possible human beings would be to presume strict determinism. Then the amount of all entities endowed with the potential of becoming a person could be quantified, since which entities exist at any particular time is predetermined. In contrast, Hare's own argumentation disallows this option (cf. Hare 1988, p. 222 ff.; Hare 1993, 1998, p. 402).

the recognition of (3), is plainly unsound.<sup>75</sup> His thesis, that mere existence is already useful (Hare 1995b, p. 311) and the denial thereof harmful (Hare 1995a, p. 356) is also just as problematic as his premise that the difference between factual and hypothetical existence is morally irrelevant.

In view of Hare's entire argumentation it is, in my view, decisive that one cannot refer to possible persons as spatiotemporally existing things by means of an existence quantifier. We simply do not have the necessary identity conditions at our disposal to pick out and individuate possible persons and ascribe interests or other properties to them in a controlled way.<sup>76</sup> The asymmetry between abortion and contraception is not thanks to an immediate moral argument (though there may be some), but due to the fact that a necessary condition for ushering in possible persons as individual bearers of ethical claims does not exist.<sup>77</sup> So the "world of difference between contraception and abortion" which Kuhse and Singer (1994, p. 145) speak of really does exist. However, it does not apply to the moral value but to the ontological status: It concerns the difference between our actual world and all to which philosophers refer as possible worlds. In other words, it is the difference between the area in which there are criteria for individuation and hence reliable arguments and the area in which this is not the case. Hare's arguments thus need no moral refutation, but already fail due to their untenable ontological preconditions.

### 3.4.2 *Gene Therapy, Persistence and Personality*

Starting from the "non-identity problem", which is analyzed by Parfit (1989, chap. 16), Zohar (1991a) proposed the thesis that gene therapy cannot really be understood as a form of therapy since a change in the genome results in a new human

---

<sup>75</sup>Such an obligation can be justified via contingent demands in case of their actualization. (cf. Schöne-Seifert 1995, p. 211), even if no actual obligation to install production can be justified therewith. For this one can speak about future generations via a universal quantifier. But with an existential quantifier, reference to possible persons should be avoided, as will be demonstrated shortly. However, it is precisely this difference that is blurred by Hare (1995b, p. 314 f.) when he transfers – in his example – from the group of future church attenders to "a possible person" (cf. also Birnbacher 1988 on the problem of responsibility for future generations).

<sup>76</sup>Cf. Quine (1979) on this; with reference to the objection of missing identifiability Schöne-Seifert (1995, p. 211 f.) is too generous, presumably because she wants to confront Hare on moral ground. Even in contexts of immediate procreative actions, neither the future action nor the possible person is individuated (the only impracticable alternative for Hare himself would be the presumption of strict determinism).

<sup>77</sup>Perhaps Hare overlooks this because in his deliberations on abortion and the golden rule he always assumes the factual existence of a person who evaluates retrospectively possible alternatives in the light of their actual interests (cf. Hare 1993, p. 173). Here, an individuation seems possible – at least prima facie; however, for this Hare has to understand names (or definite labels) as rigid designators, whereby he is bound to the strong reading of Kripke's conception: Rigid designators always designate the same individual, even in worlds in which the individual does not exist.

organism coming into existence. Since the persisting subject of the therapy thus goes missing, according to Zohar (1991a, p. 276) there can be no “person-regarding reasons” for such a medical measure, only impersonal ethical reasons at most, since, as Zohar retaliates to the objections by Kahn (1991a), a duty to prevent harm to an embryo can only be present when the persisting subject of this harm also exists (Zohar 1991b, p. 311). In contrast to Zohar’s arguments, Kahn (1991a) had proposed a differentiated theory of harm<sup>78</sup> and had also argued the thesis that the relationship between genetic and organismic identity certainly permits to some degree the change of the former, so that in opposition to the strict identity requirement of Zohar (1991a) the genetic identity is a “matter of degree” (Kahn 1991b, p. 316) as regards the persistence of a human organism.

Yet this debate is branded by Zohar and Kahn seesawing between the persistence of the human organism and the identity of person over time, whereby a few unclari- ties creep in.<sup>79</sup> This can also be observed in the dispute between Elliot (1993, 1997) and Persson (1995, 1997). Elliot (1993) follows up Zohar and Kahn’s deliberations and pleads explicitly for the thesis that Zohar’s objections do not apply as long as the persistence of the human organism is taken as a basis, as this allows a certain amount of changeability in the genome. Admittedly, Elliot also applies the concept of personal identity in the sense of the persistence of person (Elliot 1993, p. 36), whilst interpreting it elsewhere as personality (*ibid.* p. 30 fn. 9). The result of this unclarity is that subsequent to his essay a dispute develops between him and Persson as to how one should deal with the following problem: How should one deal with the case that due to a gene therapy intervention “numerically distinct persons could originate from one and the same conceptus” (Persson 1995, p. 17 f.). The further course of this debate (*cf.* Elliot 1997; Persson 1997) need not interest us at this point, as the analysis of personal identity excludes precisely this problem. In the case of a gene therapy intervention by which the persistence of a human organism is preserved, it can develop a different personality than is the case if an intervention remains undone. But that does not mean that the non-identity problem described by Parfit arises with reference to different personalities. Personalities are complex pat- terns of the properties and capacities of an organism and not independent entities with their own conditions for persistence.

Nevertheless, this debate raises questions that are also of interest to the biological approach. Accordingly, the beginning of existence of the human organism is tied to the actualization of an individual genome. And the question is now whether this might lead to similar consequences as those drawn by Zohar (1991a). The following concerns neither a comprehensive discussion on ethics nor a philosophical analysis of the concept of harm or the difference between person-regarding and impersonal moral reasons. On the contrary, it is about taking a closer look at the content of the

---

<sup>78</sup>I will not expand on this dimension of the debate in the following, as it does not affect the theses of the biological approach.

<sup>79</sup>Since both Zohar and Kahn differentiate between the person and the organism as separate enti- ties, from the perspective of my biological approach they blur the decisive difference between genetic interventions before and after the beginning of the existence of a human organism.

biological approach by relating it to this problem complex. This opens up the opportunity to expound a little on the premise of “necessity of origin” firstly (Sect. 3.4.2.1). Secondly, the role of the genome in the persistence concept of the biological approach can be explained and freed of possible misunderstandings (Sect. 3.4.2.2).<sup>80</sup> And thirdly, it grants the possibility of saying more about the relation between genetic identity and personality (Sect. 3.4.2.3).

### 3.4.2.1 Interventions Before or After the Beginning of Life: A Decisive Difference

The “non-identity-problem” Zohar associates with gene therapy takes two fundamentally different forms which can be expressed in two questions:

- (Q-1) Does a gene technical intervention lead to a different possible organism B coming into existence instead of a possible organism A, whereby at the time of the intervention neither A nor B exist?<sup>81</sup>
- (Q-2) Does a gene technical intervention on an already existing organism A lead to it ceasing to exist and possibly the development of a different organism B (or several organisms)?

(Two Questions)

While it is in principle useful to formulate the “non-identity-problem” with recourse to question Q-2, the first question that is being discussed in this section is misleading if it is asked in relation to the wellbeing of A or B. Kripke (1981, p. 127 ff.) formulated the thesis of the “necessity of origin” for human organisms as follows: that a human individual is of necessity individuated through the fusion of two specific gametes. Hence it is impossible for a human being to have developed out of a different gamete from those out of which she has factually emerged. In the second chapter of this study I understood the thesis of the necessity of origin as the requirement that an entity cannot originate through two different events. Thus, with the assumption that the identity conditions for events also contain the space-time locations, the identity of the space-time location is also required over and beyond the identity of the gametes named by Kripke. At first glance, Zohar’s thesis that every gene-technical intervention ends the persistence of an existing organism (our question Q-2) seems to apply likewise to the identity of the gametes. Consequently, e.g. Elliot (1993, p. 40 fn. 26) suggests that parallel to the moderate answer to Q-2, moderate identity conditions for gametes should also be permitted. It would then – at least assuming the conditions demanded by Kripke – be possible for one and the same human organism to begin to exist as long as the gametes have not been geneti-

<sup>80</sup> Even if the biological approach in general loses its validity for contexts of technical interventions, at least it can be partially shown to which assumptions he is obligated in this context.

<sup>81</sup> The reference to individual possible organisms could be avoided through a more technical formulation as regards possible worlds; it has only been chosen for the sake of simplicity, though it is strictly speaking incorrect.



cally changed in any way that would harm their own conditions of persistence. It is not clear to me what form such a moderate conception of persistence for gametes could take. But this way out by Elliot does not help the biological approach, at least, not when one makes the extremely plausible assumption that every gene-technical intervention on gametes leads to a change in the space-time course of events. The consequence of this is that a gene technical intervention (just as many other interventions and events) leads to a different organism coming into existence from the one that would have come into existence in a difference course of events.<sup>82</sup> So Q-1 is only a case of the non-identity problem in a specific sense. Such a gene-technical intervention results in a different possible world becoming actual – a world in which there are other entities than in the one which becomes actual when the intervention is not undertaken. For the ethical evaluation of this kind of intervention one cannot therefore usefully seek ethical reasons that presume the identities of the parties involved.<sup>83</sup> But the search for other ethical measurement scales nevertheless remains entirely consistent with this.

### 3.4.2.2 Genetic Identity: All or Nothing?

Regarding Q-2, Zohar (1991a) arrives at the result that every change in the genome of a human organism means the end of the persistence of that organism. Like Kahn (1991a) or Elliot (1993), I also consider this consequence – as the answer to Q-2 – to be highly implausible. But since the individual genome of a human organism also plays a central role in the biological approach, I would now like to elaborate on why this consequence does not arise, but rather, the biological approach can allow an existing human organism to survive changes in its genetic identity.<sup>84</sup>

To see this, first let us recall that in the biological approach the human organism is not understood as a genetic individual in Wilson's sense (1999, p. 64 f. and 106). The identity of the genome is neither a necessary (mutation, copying error) not a

---

<sup>82</sup>For this reason, Harris's claim (1998, p. 81) that the events that befell the gametes from which a specific organism has evolved belong to the life history of this organism can only be understood as applying exclusively to the factual procedure. As soon as one considers counterfactual procedures having influenced the procreative event then one can only speak of this specific organism *post festum* rather than prospectively, as in the latter case it simply does come into existence and possible organisms are not individuated.

<sup>83</sup>As regards this case Zohar is therefore borne out. There cannot be prospective person-regarding arguments (referring to A) for gene-technical interventions before the moment of the beginning of A's life. But it is naturally not impossible for there to be person-regarding arguments in respect of other factually existing persons (e.g. the potential parents). Such an intervention is as yet necessarily ethically indifferent, because possibly there are indeed impersonal ethical reasons – for which Parfit pleads (whereby such recognition of impersonal reasons in ethics does not pledge a utilitarian position).

<sup>84</sup>This does not mean that there could not also be gene-therapeutical interventions that are incompatible with persistence. If, e.g. such an intervention leads to multiple formation, the fusion of previously distinct individuals, or to a termination of the integrated life process, then the persistence of the original organism would be ended.

sufficient (identical twins) condition for being a human individual. On the contrary, the special role of the activated genome consists in being the central instance of self-regulation and integrative achievement of the human organism in its earliest stages. The thesis that all features that this organism will develop because of its genome are also necessary for persistence is not tied to this.<sup>85</sup> In fact, persistence is itself understood as a causal space-time process. And with this, a gene therapeutic intervention that in any case concerns the material basis of realization that is here understood as a process can be understood as a kind of sidetracking maneuver: Influence is exerted on the specific development through which this organism will pass in the concrete embodiment of her space-time causally organized “track”. In the biological approach, not every change of direction in the life of an already existing human organism means at the same time the end of its existence. In other words: Although the biological approach bases on a specific conception of essence of natural kind, it does not imply individual essence.<sup>86</sup> If one observes both the difference between Q-1 and Q-2 and the function that is ascribed to the individual genome in the biological approach, then the latter is by no means preassigned to the consequences that Zohar has drawn in respect of gene therapy. Even when it does not answer the question of the ethical evaluation of gene therapy, the biological approach can beyond that actually help to sort out the somewhat convoluted debate about the connection between gene therapy and the non-identity problem.

### 3.4.2.3 Genetic Identity and Personality

Hardly any other natural factor exercises such a great influence on the scientific, medical and everyday understanding of personality as genetic identity. At the latest since the issue of whether cloning human beings endangers the “identity” of the original or the copy, it has been clear that many presume a subliminal belief in the genetic determination of personality (Quante 2014b, chapter V). So it comes as no surprise that a connection is also formed between genetic and personal identity in the field of gene therapy.

In the framework of the analysis of persistence, the biological approach avoids using the concept of person or interpreting the person as a different entity from the human organism (elsewise Elliot 1993, 1997, also Persson 1995, 1997). However, a gene therapeutic intervention can cause the personality of the human being treated in this way to be differently constituted from what she would be without the intervention. Even if the biological approach also separates questions of persistence and those of possible personality changes, it is still unarguable that the individual human genome plays a central role for the respective personality. Where does this relation fit in with the conception presented here? First of all, the human genome is an entity

---

<sup>85</sup> Concerning technical interventions, not even the limitation to the natural species/kind would be demanded, as the biological approach does not provide any answers in this context.

<sup>86</sup> Limited to the boundaries drawn through the premise of necessity-of-origin. However, this premise signifies a purely quantitative and not a qualitative conception of the essence of an individual.

that can be grasped from the scientist's observer perspective. This means that an evaluative-descriptive phenomenon such as personality, which can only be understood from the participant perspective, does not appear within the perspective that is constitutive for the genome. If one enquires after the bond between them, one presumes a preconception of personality from which, in the observer perspective, to ask about the causal enabling conditions for personality, or more exactly, certain features of personality. Not only the personality as a whole, but also most of the features of personality, are not graspable from the observer perspective. Such features as the color of eyes or hair that can be grasped from the observer perspective, can as such, though belonging to the appearance of a person, hardly form a constitutive part of her personality. This would only be the case if the person took an evaluative stance towards these genetically determined properties. And even such general character features as aggression or caution are, if at all genetically anchored, too unspecific as behavioral patterns to count directly as constitutive features of a personality. In the framework of the general conception presented here, the following relation thus results: Certain genetic features can be understood evaluatively from the participant perspective and their causal enabling conditions can be identified in principle in the observer perspective. Now if a gene therapeutic measure induces a personality change (an outcome ensuing later in the case of an embryo), the question is whether such an imprint on the personality is desired. Since the personality has not yet developed in a therapy at an early stage and therefore this imprinting will be integrated into the personality from the start, the question cannot be answered from the viewpoint of the person involved. The ethical consideration must therefore be a decision at an intersubjective reasonable-person standard. But with this, gene therapy takes its place alongside many other measures through which a personality is molded (socialization, environment etc.). For this reason there will be one general ethical problem here – every personality is constituted through the specification of natural conditions and in the context of social molding (cf. Chap. 5). An analysis of the spectrum of gene therapeutic interventions one can or can no longer take responsibility for cannot be conducted in the framework of this study. One of the strengths of the biological approach is, however, that it shows where the ontological problems end and the evaluative questions begin.<sup>87</sup> A further central result for ethics consists in showing that the evaluation of gene therapeutic interventions before and after the beginning of life requires different ethical justification strategies. Even if neither suffices to answer the multifaceted ethical problems in the context of the evaluation of gene therapy, in vitro fertilization or other reproduction technologies, the conception developed here does permit the institution of preconditions for a rational discussion of these problems.

---

<sup>87</sup>Due to the implicit essence of natural kind I also consider it possible to effect an initial limitation of the "genetic defect" concept via the specific characteristics of normally developed members of a species. Since the concept of ill health also has a social dimension, however, this does not itself suffice to draw boundaries between the correction of genetic defects and the improvement of human organisms. And the question of whether the difference between ethically permissible and ethically impermissible interventions runs along this boundary, cannot be decided herewith.

## Chapter 4

# Death

In Germany, during the run-up to the passing of the transplantation law in the early 1990s and in the context of actual discussions of ethical problems of transplantation medicine in the last years, a fierce dispute was conducted over so-called “brain death”. Johannes Hoff and Jürgen in der Schmitzen, the editors of the anthology “*When is the human being dead?*”, for example, made it their expressed aim to expose as illusionary the notion that there is a social or scientific consensus about the definition of death. They view the brain death criterion as a new definition of death and fear that for pragmatic reasons the death of a human being is being newly defined for the purpose of maximally effectively winning organs for transplantation. The controversy within this dispute runs between the traditional “cardiovascular death” and the “whole-brain death”. Not as soon as the brain of a human has failed irrevocably, but only when the cardiovascular function has also failed can a human be regarded as dead in the eyes of these critics of the whole brain death criterion.<sup>1</sup>

In a debate that was conducted mainly in the USA, concerning the adequate definition or rather an adequate new definition of the concept of death, the controversy lies elsewhere. As the title of a collection of essays: “*Death: beyond whole-brain criteria*” edited by Richard M. Zaner indicates, it concerns whether e.g. the conception of whole-brain death suggested by the President’s Commission (1981) should not be criticized as being too conservative a strategy and replaced by the “more humane” conception of higher brain death. The person has ceased to exist not only when the whole organism is dead due to failure of the whole brain including the brainstem; the death of a human person has already occurred when the higher brain functions indispensable for the personality of a human being fail through failure of

---

<sup>1</sup>For the connection between the German discussion on the brain death criterion in transplantation medicine cf. Ach et al. (2000), Stoecker (1999) and the contributions in Ach and Quante (1999). The latter volume also contains, alongside the current text of the German transplantation law, the various drafts that have been discussed and voted on in the German Parliament.

the neocortex. For the supporters of this concept of higher brain death<sup>2</sup> it is all about defining the subject of death correctly. And this is said to be the human person, but certainly not the human organism (cf. Gervais 1986 or Engelhardt 1986). According to Robert M. Veatch,<sup>3</sup> for example, one treats a person morally wrongly not only when one treats her as dead although she is still living, but also when one treats her as living although she is already dead. Moreover, the argumentation continues, the application of the criterion of whole-brain death results in relatives being unnecessarily emotionally and financially burdened, important medical resources being wasted and the possibility of transplantation not being realized.

## 4.1 Targets

As in the theme of the previous chapter, in the discussion about human death ontological and ethical questions overlap in a way that is difficult to disentangle. In this chapter I would like to show that the analysis presented here can contribute to clarification. Above all, the differentiation between persistence and personality should help to keep ontological and evaluative aspects of the problem apart in the right way. After a few initial reflections which are indispensable to an adequate discussion of the definition of death (Sect. 4.2), I will propose and explain a concept of death (Sect. 4.3). This ensues from the analysis of human persistence suggested in the second chapter of this study. Thus, for one thing, the suggestion made there will be filled out and detailed. For another, not only can the concept of whole-brain death be philosophically underpinned, but inversely, the conception of human persistence developed in this study can be additionally buttressed through proof of its practical relevance and plausibility. Following that, on the basis of the analysis of personal identity presented here, rival philosophical definitions of death and the criteria of death connected with them will be discussed and reviewed (Sect. 4.4). In the final section of this chapter I will discuss some of the basic difficulties of the conception presented here, which arise through objections or even fictitious problem cases, and attempt to eliminate them as far as possible. Besides this the burdens of proof and questions shall be named which have not been answered or have been left open respectively (Sect. 4.5).

---

<sup>2</sup>When discourse in the following concerns the higher brain death criterion or concept, I am always referring to the so called “higher-brain-death”. The conception that is valid in the UK, which equates the death of a human being with the irreversible failure of the brainstem, but will be defined below as the conception of brainstem death (on the criticism of this concept cf. Evans (1994) and Sect. 4.4.2.2 of this chapter).

<sup>3</sup>Cf. The following statements by Veatch (1976, p. 33): “It is morally wrong to treat a dead man as if he were alive” and “it would be still a moral affront to the dignity of man to treat a corpse as if it were a living person” (ibid. p. 34). The question as to who is the subject of harm here is posed by Walton (1980, p. 44 ff.) who sees it as a central task in the context of the brain death debate to clarify the moral status of the human corpse and provide it with ethical protection. The fears about misuse expressed e.g. by Jonas (1985) are certainly one reason for these reflections.

## 4.2 The Fight over the Definition of Death: Some Clarifying Considerations

### 4.2.1 *The Background*

The last 20 years of debate have shown a distinct tendency to cite transplantation medicine as a plea for a variant of the brain death conception. In this respect, the possibility of winning organs, while not being the original reason, does in the meanwhile enjoy lofty standing in the defense of so-called brain death (cf. McCullagh 1993, Chaps. 1 and 2). This context reveals the first big problem of the concept of death, which one can call the *problem of evaluation and description*. Does death concern a fact that we can discover, or do pragmatic, social or moral criteria, alongside conceptual elements and scientific assumptions, also enter constitutively into the specification of the definition of death?<sup>4</sup> Should we even settle on the time of death according to considerations of usefulness, since this is ethically legitimate (cf. Bayertz 1999)?

### 4.2.2 *Criteria of Death: Three Candidates*

If one disregards such perceptions that see the death of a human being as being that an immaterial soul leaves the body, that the last breath has been drawn, or even that all bodily functions have ceased, then there are vis-à-vis three conceptions that favor the following criteria of death: the cardiovascular death, the whole-brain death and the higher brain death. Against this background the following scenarios are conceivable:

1. all brain functions, including the cardiovascular functions, have failed irrevocably – death according to all three conceptions
2. all brain functions have failed irrevocably, but the cardiovascular functions and respiration are maintained with mechanical support – death according to the concept of whole-brain and higher brain death
3. only the neocortex functions have failed irrevocably; the brainstem is still functioning, respiration and heart activity are spontaneous – death according to the concept of higher brain death
4. some of the neocortex functions have failed irrevocably – death according to some conceptions of higher brain death
5. Total irreversible failure of the brainstem with maintenance of the neocortex functions; cardiovascular functions and respiration are maintained with

---

<sup>4</sup>Cf. Gutmann (2015) for an overview of the actually discussed options in Germany and further European countries.

mechanical support – possibly death according to the concept of brainstem death that is valid in Great Britain.<sup>5</sup>

(Five Scenarios)

These five possible cases highlight the second wide problem area that marks the dispute about an adequate definition of death. One can define this as the *pluralism problem*, which must be divided into two types. The first can be paraphrased by the following questions: Do several definitions of death exist side by side? Or does a human die different deaths? First the death of a person and then the death of the organism? Is a plurality of death concepts permissible because there are different subjects of death? One can define this as the *ontological pluralism problem*. If one asks differently whether a plurality of definitions of death must be permitted because the understanding of death is relative to worldviews and moral values (i.e. irreducibly evaluative) and in this area no consensus is possible, then the second kind of pluralism problem is revealed. This difficulty arises from the combination of a specific answer to the *problem of description and evaluation* with the thesis of irreversible plurality in respect of evaluative assumptions. One can define this as the *evaluative pluralism problem*. Both kinds of pluralism problem lead to a follow-up question: Who should decide – the scientist, the physician, the philosopher, society, the law?

### 4.2.3 Three Levels of the Problem

To obtain some degree of clarity here, it is necessary to distinguish three levels in the discussion: the level of testing procedures, that of criteria and that of the definition of death (cf. Birnbacher 1994, p. 29 pl.). *Testing procedures* belong exclusively in the medical domain and provide the empirical proof that the arguable criteria demanded by the respective overall conception are fulfilled. A *criterion* serves as an indication of the presence of the features required by the definition and must satisfy both the relevant empirical conditions and the scientific standard (cf. Birnbacher 1999). The features included in the *definition* are not ascertainable by means of empirical conditions or scientific standards. Here, the framework can be provided only by reflection on the adequacy of a definition in relation to linguistic intuitions and logical conditions, or to other theoretical (philosophical) assumptions and, if applicable, to theoretical or practical experience. So it is not only scientific or empirical conditions which determine whether a definition of death is acceptable; that decision must be reached before the background of wider (philosophical, religious, cultural) convictions. In contrast to the question of suitable criteria and tests,

---

<sup>5</sup>This condition is called the total locked-in syndrome; cf. Bartlett and Youngner 1988, p. 206; Linke 1993, p. 132 f.; McCullagh 1993, p. 63 or Spittler 1999, p. 44. How long such a situation can be stable is empirically just as unclear as the answer to the question of whether the function of the cerebrum is adequate for the existence of phenomenal or propositional consciousness in total failure of the brainstem (cf. Spittler 1999, p. 44).

the definition of an adequate concept of death thus requires extensive conceptual analyses and is therefore a philosophical matter.<sup>6</sup>

#### 4.2.4 Three Strategies

In efforts to reach an adequate definition of the concept of death, the problem of evaluation and description and that of pluralism outlined above can be tackled by employing three fundamentally different strategies:

*First Strategy* If one accepts that *every* definition of death of necessity contains pragmatic, social or moral elements because one considers death to be a social or moral phenomenon, then one must concede as much plurality as is viewed to be acceptable or unrescindable as regards social or ethical convictions. If no ethical arguments can be introduced which point to a specific concept being worthy of taking precedence, then one must grant the right to an individual choice of the concept of death.

*Second Strategy* This differs from the first strategy in its assumption that one can develop valid ethical arguments that are evidence for the binding character of a particular conception of death. In this way pluralism is avoided through a basis of strong ethical vindication.

The objection raised e.g. by Jonas that the brain death concept has been introduced for purely pragmatic reasons, can be confronted in the first two strategies with comment that in the end every concept of death is pragmatically or morally instituted (cf. Bayertz 1999). The objection raised by supporters of the traditional conception of death is thus not actually invalidated but declared to be irrelevant. It is just unavoidable and therefore legitimate to determine the concept of death from pragmatic, social or moral aspects.

*Third Strategy* If one is not convinced that a specific conception of human death can be deemed universally binding on the basis of an ethical theory and is nevertheless not prepared to concede a plurality of concepts of death, then, as an alternative one can only assume that there is philosophically reconstructible content

---

<sup>6</sup>Alongside this trichotomy of definition, criterion and testing, which has become widely accepted in the literature, Veatch (1976, p. 24) gives a variant four-step array which arises from his further subdividing the definition level into a “purely formal analysis of the word death” (ibid.) and an analysis of “the definition of death” (ibid.). Likewise differing from the above trichotomy, Kurthen et al. (1989) demand as a fourth dimension a level of attribution, i.e. a definition of the subject of death. This submission is misleading, as it implies that there could be several subjects of human death. The contribution by Kurthen and Linke (1994) is an example for this error. The attempt there to provide proof of the inadequacy of various concepts of death rests on the following consideration: In a first step the loss of which functions constitute death is specified. Then, in a second step, one constructs that subject of death which is constituted resp. definable via these functions. Thirdly, one shows that the various subjects of death thus obtained are not identical with the human being as the subject of death. The error lies in the assumption that for every bunch of functions it should be possible to construct an exactly fitting subject of death.



to the concept of death which is of a purely descriptive nature. One must therefore suggest a reconstruction of the concept of death which is independent of pragmatic, social or moral elements. If such a definition of the concept of death succeeds, then it must be regarded as adequate, independently of whether it is revisionary compared to traditional concepts or commonsensical preunderstanding, or whether it favors or impedes specific possibilities such as organ transplantation.

This third strategy takes two things for granted. Firstly, there is only one concept of death; talk of “whole brain death”, of “cardiovascular death” or of “higher brain death” is misleading, as it implies a plurality of concepts of death: with failure of whole brain, higher brain or the cardiovascular cycle *criteria* are being named which enable the death of a human being to be established or predicted. The event of death itself should be distinguished from this at least conceptually. Secondly, this strategy presumes that there is a concept of death which can manage without pragmatic, social or moral components. According to that thesis, the event of death is amenable to a definition that contains none of the “pragmatic” dimensions in the sense explicated above. Pragmatic elements as well as a certain amount of plurality only gain relevance when the criteria are devised. This third strategy contradicts head-on the objection to the pragmatically motivated redefinition of death. Both linguistic intuitions and philosophical considerations and our cultural background demand a non-pluralistic, descriptive conception of death according to the basal intuition of this third strategy.<sup>7</sup>

#### 4.2.5 *Further Differences*

In addition, two differentiations must be undertaken without which an adequate clarification of the concept of death is not possible. For one thing, the death event and the declaration of death must be differentiated; whereas the former is open to a purely descriptive definition, the latter (as a speech act) also has moral, social and epistemic conditions of adequacy. For another, the process of dying must be differentiated from the event of death and this event from the status of being dead. The following concerns neither an analysis of the definition of dying nor a discussion of the different problems posed by the status of being dead from the philosophical

---

<sup>7</sup>Significantly, Hans Jonas shares this fundamental intuition, for his objection against the brain death criterion is already justified by the indication that it concerns a “pragmatic” decision. But such an objection only makes sense when one imputes the possibility of a non-pragmatic decision. If one did not do this, one could at the most only criticize the choice of the brain death criterion as being a poorly justified pragmatic decision.

viewpoint.<sup>8</sup> These questions remain just as untreated as a comprehensive analysis of the conditions under which it is justifiable to declare someone dead.<sup>9</sup>

### 4.3 Death: Human Persistence (II)

#### 4.3.1 Definition

Death: The end of the existence of a specific organism.  
(Definition of Death)

This definition associates the determination of death with the questions concerning the conditions of persistence of a specific sort of entity – namely organisms. Persistence conditions consist not only in those conditions applying to the identity of this life process itself, but also in some of the identity conditions that apply to the body as the material basis of the realization of this life process. This results from my premise, that spatiotemporal instantiation is a necessary condition for the individuation of concrete entities. The differentiation between type and token cannot be undertaken on the level of functional conditions or functionally characterized processes alone. Hence the identity conditions of the basis of material realization must also be incorporated, so that possible cases of fission or fusion can be ruled out.<sup>10</sup>

---

<sup>8</sup> Whilst Walton (1979, Chaps. IV and V) discriminated precisely between the process of dying and the event of death; in his study the analysis of the definition of death is intermingled with an analysis of the status of being dead (ibid. p. 67). “Being dead” for one thing poses the problem of who predestines this status and for another leads to the question of what attitude should be taken rationally to this status; for the discussion on some of these regressing puzzles that can be traced back in the tradition cf. Feldman (1991, 1992) and the contributions in Fischer (1993). Alongside the motive that some of the traditional philosophical questions are linked with the status of “being dead” rather than death as an event, Walton’s intermingling of the two problems thus stems from his migration from the analysis of the meaning of the definition of “death” to the analysis of the relevance of death for us. However, the epistemic and axiological attitude of the subject to his state of being dead cannot be included in the analysis of the death event. It is one thing to determine the event of death philosophically and something quite different to investigate the relevance of this event for our personal life. In other words, death can be grasped in the observer perspective, whereas the relevance of death requires the participant perspective.

<sup>9</sup> The question of when it is legitimate to declare a missing person dead can quite obviously not be answered adequately through recourse to the considerations lined up in this chapter.

<sup>10</sup> Here one must speak of the body as the material realization basis of reality rather than the *Leib* (the body image or the embodied person, as opposed to the body, as the physical enabling condition) because the observer perspective is appropriate for the question of persistence, while the *Leib* is constituted in the participant perspective.

### 4.3.2 Explanations and Modifications

The definition proposed here is complex, its components requiring elucidation (cf. also the comments on the definition of the beginning of life in the previous chapter). And in contrast to my earlier suggestion for a definition (cf. Quante 1995b) it offers modifications which shall also be explained in brief. But first to the basic elucidation:

- (a) “*Specific*” organism: This clause is necessary because death always means the end of a specific entity which must be explicitly named. Thus the biological process which represents the life of an organism as a whole can be at an end without the life processes of part of this organism (individual subsystems, organs or cells) also having to be at an end. Inversely, the organism can survive the death of some of its parts perfectly well. When one is speaking about the death of a human individual, one is referring to cessation of the integrated life process of this human being as a whole organism. The discourse about a “whole brain”, “higher brain” or “cardiovascular” death contains the ambiguity that it can mean the death of a human individual and it can mean the death of the respective organs or parts of organs. But the identification of the death of a part with the death of the whole always requires more extensive justification (cf. below Sect. 4.3.3). At any rate, it must always be clear which biological process is assumed to have ceased, since a biological process can contain other biological processes as integral parts.
- (b) “*Biological process*”: With this the definition of death alludes to *biological* processes. According to widely shared intuition, only biological entities (organisms) can die. In contrast, artifacts can only cease to exist. An organism is – in the observer perspective – a self-integrating life *process* realized in a material body. So the definition indicates that this concerns biological *functions* and not structures of the causal carrier of these functions.<sup>11</sup> In the debate over adequate criteria and testing, the question is frequently asked as to whether the loss of functions or the destruction of the structures and properties of the material basis that supports these functions causally should be decisive.<sup>12</sup> The decision in favor

---

<sup>11</sup> In the talk of “higher functions”, which is widespread in the discussion of the concept of brain death and is also occasionally used below, the following differentiation must be observed: “Higher functions” normally refers to functions such as consciousness, self-consciousness, rationality, verbal comprehension and communication, i.e. those achievements which are necessary for personhood and whose causal basis is thought to be in the neocortex. The term “higher functions” can also have the meaning of a “function of a higher order”. In that case it means that certain functions arise through the collaboration of other more basal functions. Here I assume that the higher functions in the first sense are always realized through functions of a higher order, i.e. by the brain as a whole or several causally supported, interacting brain areas. In this sense, all brain functions are biological functions because their basis of realization can be explained by the collaboration of other basal functions.

<sup>12</sup> For the discussion of this point cf. Lamb (1985, p. 59 ff.). With recourse to the criteria level, Lamb grapples here with critics who believe the destruction of the structure rather than the loss of function to be decisive. In contrast, an argument for the recourse to the loss of function is deduced

of the functions option becomes conceptually necessary through the specification of the entity in question as a biological *process*. It also has the advantage of both being able to take into account the principle variability of the functions compared to the material basis (multiple realizability thesis) and being more easily assessable through testing.<sup>13</sup> Since, in addition, higher functions assume in part the interaction of other functions, the approach via structures is also inadequate, because the functional unity of functions is not adequately grasped therewith. But it is precisely this functionality of a higher order that is decisive for the integrative achievement that leads to a complex integrative life process. These functions are not necessarily bound to a specific material basis, but demand only an adequately complex organization of this basis. Should a biological life process be realized in a silicon basis and be causally carried by the structures of this material basis, then this is also a biological process.

The explicit talk about the life process has two advantages over the general discourse about the organism: For one thing, it avoids the material basis (the body) being centered as the sole determinant for the conditions of identity. For another, it permits the possibility that a specific individual life process (i.e. an organism as a whole), continues to exist even when some of its parts (= the function of organs) are supported by another causal basis (e.g. artificial organs). According to the above definition, persistence consists only in the further existence of the integrated life process of the organism as a whole, regardless of how this is realized in a material basis. The loss of the material basis of a sub-process is unimportant for the integrated life process of the organism as a whole as long as this sub-process can be maintained by a replacement of the material basis. But this applies only with the qualification that such a replacement must not override other conditions for the persistence of the organism in question. The integrated life process of the whole organism therefore persists as long as the integrative function, which integrates the various biological processes into a life process, is performed without interruption. Neither the concrete material basis nor the entirety of all sub-functions is necessary for this. Death is therefore the event of the cessation of the integrated life process and not necessarily a process of the destruction of the causal basis for this integrated life process.

---

here from the definition of the definition of death itself. Different from Stoecker (1999, p. 75 ff.), who defends a morphological definition of the definition of life targeted at the structures (and on this basis reaches a rejection of the brain death criterion), I assume a processual conception of life according to which life consists in an actually exercised capacity. This ensues from the above definition of death, because therein the organism is defined as a life process, and thus I comply with the analysis by John Locke (II, 27, § 4 = 1975, p. 330 f.). On my expressions “causal basis”, “basis of realization” and the “multiple realization thesis” cf. Kim (1993, pp. 309–335).

<sup>13</sup>For one thing, in many functions one does not know the basic causal structure and therefore cannot know exactly what to look for; for another, complex functions are a collaboration of several functions which can each have different causal bases. A precise localization of such higher functions in the brain structures is therefore hard to imagine. Moreover, changes in the structure of the brain during a patient’s lifetime are difficult to verify. And finally, there is the problem of differentiating the changes in structure that caused the loss of function from those changes in structure which have only occurred after the loss of function.

- (c) *A modification*: Divergent to an earlier proposal (cf. Quante 1995b), the present definition of death no longer contains the stipulation of irreversibility, as the recourse to the condition of irreversibility is misleading in the definition of death. Organisms cannot survive either short or longer term phases of interruption of the integrative life process (cf. Olson 1997, Chap. 6). This differentiates them from artifacts such as e.g. an automobile which can survive phases in which it is taken to pieces and stored.<sup>14</sup> To speak of reversibility and irreversibility is therefore not meaningful on the level of the definition of death, but, as regards the functions belonging to the integrative life process, it is. Here, it is also important to keep in mind that the stipulations of reversibility and irreversibility refer only to the function or its loss, but not to how this function is supported causally and wherein it is realized (cf. Gervais 1986, p. 26 ff.). A loss of function is still reversible if there is a possibility of replacing the basis of realization (organ transplantation, artificial organ, prosthesis) or of instigating the exercise of the function mechanically (artificial respiration, pace maker etc.). The only function whose loss cannot be replaced by another basis of realization is the integration of the life process as a whole. This is because such a loss implies an interruption of this integrative life process that is incompatible with the persistence of the respective organism.<sup>15</sup>

### 4.3.3 *The Criterion of Death for Human Organisms*

Whereas the definition of death and the above considerations have up to now not had to take into account what kind of organism is under discussion, a specific criterion of death cannot be formulated without first establishing this. For, depending on

---

<sup>14</sup>Thus it depends on the further course of the integrative life process whether it can be said of the freezing of an organism that the same living creature continues to exist after defrosting as the one that began life before being frozen. If this integrative life process is interrupted by the freezing, then we are dealing with a new living creature of the same kind whose material basis of realization has a great deal in common with the previous organism (cf. Sect. 4.5.2 of this chapter).

<sup>15</sup>The differentiation between a spontaneous and non-spontaneous execution of a function, which is also drawn on in the context of the replacement of individual organs and the definition of death, only indirectly covers these findings. According to the above conception, it is indeed possible to replace the spontaneous execution of the function of individual organs by artificial aids without the persistence of the organism being endangered in principle. Relating to the replacement of that material basis that is causally necessary for the organism to become and remain an integrative life process, the insertion of an artificial replacement organ, while leading to the end of the persistence of this organism, does not do so because the integrative achievement is no longer reached spontaneously, but because it has in the meantime been interrupted (cf. the fifth section of this chapter). The correct intuition on which in my view the notion of spontaneity is based, is that because of such a replacement the difference between organism and artifact becomes dubious. With these quite different criteria of persistence come into play and the definition of death can no longer be used smoothly. These findings are also roped in by the analysis of human persistence suggested here, which assumes that for such cases the criteria written into our parlance do not apply.

the biological species, the conditions demanded by the definition of death of a concrete organism of this species are quite possibly fulfilled by diverse events. Which biological processes count as the loss of the integrative function cannot generally be said independently of the species of the respective organism. The loss of the integrative achievement in humans<sup>16</sup> comes along with – and this is an empirical and contingent fact – the failure of the brain as a whole, since this is the organ that constitutes a necessary causal basis for the capacity of a human organism to integrate. The following criterion for the death of a human being therefore applies:

The irreversible failure of the brain as a whole is, from the moment in which this organ took over the integrative achievement, the end of the integrated life process and therewith the death of the human organism.

(Criterion of Death)

This also applies when isolated subsystems still fulfill their function, regardless of whether this happens spontaneously or non-spontaneously. Such processes that were previously parts of the whole integrated process, can continue to exist even when the whole process has ceased to exist, i.e. the death of the human being has occurred. By “brain as a whole” it should not be understood that all functions of all parts of the brain (e.g. all nerve cells) have failed, but rather, the brain-specific functions by means of which the human organism integrates its life into a unity. At that moment when in a human being the brain as a whole fails, the organism ceases to exist – it is dead. This differentiates the failure of the brain as a whole from the failure of other organs or parts of the brain, since the latter, though resulting in the early death of the organism, cannot be identified with the death event. In as much as it rests on philosophical considerations rather than empirical facts, the justification of this whole brain death criterion results from the alternative conceptions and the criticism thereof that is to be conducted in the following.

## 4.4 Criticism of Alternative Approaches

In her book *Redefining Death*, Gervais distinguished between three strategies for defining the definition of death: the biological, moral and ontological strategy.<sup>17</sup> This trichotomy is misleading and should be replaced by a dichotomy. The clash

---

<sup>16</sup>One problem omitted here is that of the time span for an embryo-fetus whose brain has not yet developed. Here, other contextual distinctions are necessary. In the debate on the brain death conception this point is, however, not immediately decisive, since embryo-fetuses do not receive intensive medical treatment in the development phase in question (cf. the fifth section of this chapter). In contrast, the nonexistence of a criterion of death for embryos does bear consequences in other contexts such as e.g. abortion and winning fetal brain tissue for purposes of transplantation medicine (cf. Ach et al. 2000, Chap. 5).

<sup>17</sup>The differentiation into three *strategies* introduced by Gervais in the context the definition of death must be demarcated from the differentiation of three *approaches* to determining the identity of the human being over time suggested in this study (cf. Chap. 2). In principle, both the simple approach that is factually not represented in the discussion about the definition of death and the

between the moral strategy on the one hand and the biological and ontological strategies on the other is fundamental. According to the moral strategy, the question of an adequate definition of death is genuinely moral and does not apply to the existence of a specific individual but to her worth. This argument thus concerns the problem of evaluation and description.<sup>18</sup> Against this both the biological and ontological strategies assume that when death is to be defined it refers to the end of the existence of a specific individual. So the assumption common to both is that the problem of the definition of death requires an ontological approach. The biological strategy identifies human death with the death of the organism, whereas the strategy designated as ontological by Gervais centers on the death of the person.<sup>19</sup> Thus, two oppositions must be distinguished in following. First of all it concerns the question of whether the definition of death is accessible to a purely descriptive designation or whether it needs prescriptive considerations; I will pursue this aspect in the discussion of the moral strategy (Sect. 4.4.1). Subsequently, the issue in the argument between the ontological and biological strategies will be who the adequate subject of the death of a human being is (Sect. 4.4.2).

## 4.4.1 *The Moral Strategy*

### 4.4.1.1 **The Moral Argument**

Moral arguments are brought in over and over again in the argument about the adequate definition of death and the consistent choice of criteria for the ascertainment of the death of a human being. In doing so, the moral strategy is consulted to justify quite differing positions. Jonas e.g. defends the “traditional” cardiovascular

---

complex approach using the concept of person (cf. Chap. 2) are possible background conceptions for the strategy Gervais calls “ontological” (cf. Sect. 4.4.2.1 of this chapter). The biological approach that I favor corresponds to some extent with the biological strategy, but in decisive aspects does not match up with the positions of Lawrence C. Becker (1981) and David Lamb (1985), which Gervais summarizes and criticizes below (Gervais 1986, Chap. 3). To avoid confusion, in the following sections I will speak of “strategies” when referring to Gervais’ classification and “approaches” when my classification is meant.

<sup>18</sup>Since the analysis of persistence proposed in the second chapter implies the thesis that one can develop a purely descriptive definition of death, the problem of evaluative pluralism no longer arises for me in the following.

<sup>19</sup>If one comes to the same conclusion as Gervais, that with the human organism and the human person there are two numerically different subjects of death, then one arrives at ontological pluralism. This can lead to evaluative pluralism if one asks which subject of death the definition of death has to refer to that could be drawn on for socially relevant purposes such as transplantation medicine or the discontinuation of medical treatments. Whereas Gervais (1986, p. 205 ff.) believes an ethical consensus is attainable in this regard, the problem of evaluative pluralism crops up for everyone who is skeptical on that score. Different from the moral strategy, this crops up only on the basis of an ontological plurality of possible subjects of death. Since the arguments submitted in the first part imply that there is only one subject of human death, namely the human organism, this follow-up problem does not arise for me.

criterion, whereas Veatch, in contrast – at least according to the reconstruction of his position by Hoffmann (1979) – arrives at a defense of the higher brain death criterion on the basis of a moral argument.<sup>20</sup> Veatch considers it to be morally wrong to treat a dead body as a living person (moral dimension) while at the same time assuming that human beings must be understood appropriately not as organisms but as persons. “Appropriately” can mean, and this constitutes the ambiguity<sup>21</sup> of Veatch’s argumentation: ontologically adequate, or just – as Hoffmann thinks – morally suitable. But if it is the case that we are making a moral error when we treat a mere organism as a living person, and moreover, if it is a question of morality whether we should regard human beings essentially as persons or essentially as organisms, then the choice of the definition of death is in that sense a moral-philosophical decision that a suitable subject of death must be designated. Hoffmann pleads for the person, understood as being the subject of a conscious life, to be regarded as the adequate subject of death. According to Hoffmann, this decision pertains to the primary level of moral thinking in that it concerns the designation of intrinsic values and forms of the good (cf. Hoffman 1979, p. 436). In contrast, according to Hoffmann, the secondary level of moral thinking concerns the relation between various such values of the primary level as well as questions of priority or criteria of justice.

---

<sup>20</sup>Veatch’s position does however show ambiguity as regards the classification into Gervais’ different strategies. His phrasing of the traits that are supposed to be “essentially significant to human life” and whose loss means death is significant. This essential significance can have an evaluative dimension (“what we value most in our existence”), but also an ontologically constitutive meaning (“necessary and sufficient conditions for the persistence of an X of species A”). Hoffman (1979) has interpreted Veatch’s argument in the sense of the first reading, so that the question of whether to regard a human being as an organism or a person is a moral question. The second reading leads automatically to the ontological strategy and the question of an adequate analysis of the persistence conditions of human persons. Veatch (1988, p. 174) himself does admittedly contradict this, but in doing so bases his argumentation on the mix-up of questions pertaining to the conditions of personality and those of personal identity. He supposes that the central question is the one pertaining to the essential conditions that must be fulfilled in order that we can regard someone as being alive (ibid. 175). But in order to clarify who or what this someone is and what his essential properties are (in the constitutive sense), a theory of identity over time has to be consulted.

<sup>21</sup>One can also understand Veatch’s thesis that the definition of death always includes a moral dimension in the entirely fundamental sense that every choice of a basic conceptual scheme represents a value decision. However, this reconstruction of Veatch’s argument, which was suggested to me by Johann S. Ach in discussion, is unspecific for the above question due to its generality. Even if the choice of a basic conceptual scheme always contains a value decision, the relevant question for the definition of death remains whether the designation of the subject of death has normative or descriptive conditions of adequacy. The differentiation between value and fact questions is always an internal question, in relation to a chosen basic conceptual scheme. A further thesis also hinted at by Ach states that the division between normative and descriptive elements for purposes of practicing individuation is also internally (at least in the case of human beings) impossible. Questions of the identity of human beings would then always be an unresolvable unity of normative and descriptive elements. This seems to correspond to a reconstruction of Veatch’s arguments undertaken by Green and Wikler (1981, p. 63, fn.28). I do not agree with this reconstruction of Veatch’s considerations, nor do I consider the thesis itself to be plausible as it clashes with the linkage of identity over time to causal regularity.



Contrary to the reflection of Veatch and Hoffmann, Jonas assumes that we cannot determine the exact moment of death and warns against bringing forward the moment of death for pragmatic reasons e.g. in the interests of winning organs. This is not acceptable, since in doing so we are taking for granted a certainty of the knowledge which we do not have and, according to Jonas, which is on principle not available to us. In addition, Jonas argues, a human body that is still alive is being treated in a morally impermissible way when one e.g. removes any of its organs. Jonas firstly assumes that we treat the body of a brain-dead individual and therefore the person herself (cf. Currie 1978, p. 179) morally wrongly when we declare death according to the brain death conception and, secondly, states in form of a slippery-slope argument what morally catastrophic consequences can arise from pragmatically bringing forward the moment of death.

#### 4.4.1.2 Criticism of the Moral Strategy

The misgivings expressed by Jonas (1985, pp. 219–239) regarding a slippery slope of misuse of brain-dead human bodies contain, firstly, a *petitio*, since in the case of a brain-dead individual he speaks of a living body in the sense of an organism. Apart from this, it is secondly apparent that the social implementation of the brain death concept at the same time demands ethics of the adequate handling of the human corpse (cf. Walton 1980). Such adequate handling should also be claimed independently of the brain death debate, since in other contexts diverse “recycling interests” exist towards the human corpse, whose legitimacy must be weighed up against the moral status of the corpse. This argument put forward by Jonas applies to the problem of the responsible integration of the brain death concept in a society and her moral concepts. Although a relevant aspect is also being mentioned herewith, this aspect is irrelevant for the question of the definition of death. Veatch’s thesis, that it is morally wrong to treat a dead person as a live one, and Jonas’ thesis that it is morally wrong to treat a brain dead but otherwise living individual as a corpse, beg the question against one another: A decision as to which of the two describes the case of a (partly) brain dead individual with (spontaneous or) non-spontaneous cardiovascular functions and respiration correctly, requires additional non-moral considerations, for one thing, the issue of how the organism aspect and the person aspect interact must be clarified. What is more, clarity must be gained as to whether questions of personal persistence are ontologically indetermined and therefore only accessible to an exclusively pragmatic or moral determination. Jonas, at least, argues only an *epistemic* indeterminacy of the moment of death, and on no account an ontological indeterminacy in the sense in which e.g. Parfit speaks of “empty questions” in respect of the persistence of persons. According to Parfit, questions about the persistence of persons are empty questions because all discussants are in agreement about the facts and can nevertheless be of differing opinions as to whether the person in question has or has not ceased to exist. But this is an ontological thesis one

does not have to agree with.<sup>22</sup> From Jonas's thesis it does at least not follow that a descriptive approach to the definition of death must be wrong in principle, and this holds even when the exact moment for implementation on the level of the criteria and tests cannot be established due to epistemic constraints. Such a situation must never be misunderstood as meaning that the death of a human being is merely a social construct. Such a conclusion is based on mixing up the conditions of the death event with the conditions for justification of the verbal speech act "declare dead".

So left over from the moral strategy is the reflection that it concerns the adequate designation of the subject of human death: Is it the person that dies or the human organism, whose death is under discussion? If this were an ontological option unavailable to further reduction, then a moral-philosophical argument would possibly have to decide who is to be regarded as the morally adequate subject of death. This also seems to be one of the main bones of contention in the dispute about the traditional understanding and the variants of a brain death conception (cf. Kurthen and Linke 1994; Steigleder 1994 or Wils 1994 on this). But if it can be shown that the "organism – person" opposition is not of *this* kind, then this point, which is the only one that speaks for a moral-philosophical dimension of the definition of death, also drops out. In all, one can agree with Gervais (1986, p. 77) that there are indeed moral and social reasons for conducting the debate on an adequate definition of death. Besides, as the question of the permissibility of organ removal or discontinuation of treatment show, different criteria, if they are made the basis of social practice, also entail ethically diverging consequences. But the actual criteria, which are to be considered in such a definition or redefinition of human death, must be geared to the conditions of human persistence.

#### 4.4.2 *The Subject of Human Death*

Both the approach presented here and the attempts Gervais classifies as biological or ontological strategies share the premise that the definition of human death is a descriptive undertaking that reflects on the conditions of persistence. The biological strategy in Gervais' sense thereby attempts to define human death solely through recourse to the human organism, i.e. while fading out personal aspects, whereas the ontological strategy, in total opposition, focuses on the death of the person. In the following, this opposition will be avoided, since a premise that is accepted equally

---

<sup>22</sup>The context in which Parfit (1989) proposes this thesis in the third part of *Reasons and Persons* is, however, distinguished by his opposition to a theory he calls the "simple view" according to which personal identity consists in an ontological "deep further fact" – e.g. an individual soul. As a reductionist, Parfit firstly insists as regards personal persistence that this persistence is reducible and secondly that in some cases the facts do not permit us to pass explicit judgment concerning persistence, cf. here the second chapter of this study and Herrmann (1995) on Parfit's overall theory. In the context of the debate on the pragmatic and moral basis of the definition of death, these reflections quite generally play an important part: if a purely descriptive decision could not be justified, moral considerations might tip the scales.

by the biological and ontological strategy will be disputed: the possibility of a division of organism and person into separate substances. Inasmuch as the approach developed here also bears in mind the personhood of the human being, it hooks up with the ontological strategy. But since personhood is understood as a normal though not naturalizable achievement of the human organism, the approach proposed here shares essential features with the biological strategy, because the natural functions in which personhood is realized are taken into consideration. Even if personhood and personality cannot themselves be captured in the framework of a biological approach, their natural bases of realization must be taken into account. This applies because personhood and personality are achievements of normal human dispositions.

Advocates of the concept of higher brain death raise the following objection equally against both the conception of whole brain death and the biological strategy:

A biological approach confuses the death of the human organism with the death of the human person. The concept of death attained thus captures the wrong subject of death and is therefore inadequate in principle.

(Objection)

The standard criticism imputes that human death is really about the question of when the person ceases to exist, this already being the case when the functions causally supported by the neocortex fail irreversibly. The function of consciousness, which is regarded by most as the essential function of personal and mental life respectively, is especially marked out here (cf. e.g. Gervais 1986; Green and Wikler 1981). In a first step I would like to portrait and review the strategy referred to by Gervais as “ontological”, which operates with the concept of person (Sect. 4.4.2.1). After that I will address the strategy Gervais calls “biological”, which shows the greatest similarities to the approach I favor (Sect. 4.4.2.2).

#### 4.4.2.1 The Ontological Strategy

Supporters of the higher brain death criterion who pursue the ontological strategy are committed to the thesis that the loss of some or all higher functions suffices for a person to cease to exist. A “purely biological survival” as a “human vegetable” without consciousness is no longer a personal life. In this way, two fundamental assumptions are linked together. For one thing, the brain functions of the cerebrum are viewed as the basis of realization of the achievements and capacities of a human organism necessary for personhood. And for another, it is assumed that the concept of person is a sortal which can provide conditions for persistence. However, this latter assumption is not tenable.

At this point one must differentiate between two meanings of the question about the conditions of personal identity: For one thing, the conditions of persistence for a specific spatiotemporally individuated single entity can be sought. For another, with this question one can also inquire after the properties that any entity must exhibit in order to belong to the class of persons (cf. Quante 2007a, 2012, Chapter 1). So this

second sense can be referring to the *status* of a single entity: the status of personhood. If one observes the case of interest here, i.e. human persons, one realizes that a human individual does not fulfill the conditions *prima facie* at all moments of her existence, so as to belong to the category of persons. “Personhood” is therefore a “phase-sortal” i.e. the kind of sortal that does not simultaneously provide the conditions for the persistence of a single entity. In comparison with many other sortal definitions that can apply to a human being (e.g. club member, teenager or sportsman of the year), personhood applies only to specific sections of the temporally extended existence of a human being. A first consequence to be drawn from the differentiation of these two questions and that of constitutive and phase sortals is that the conditions for the persistence of a specific entity do not necessarily coincide with the conditions necessary to allow this entity to fall under a specific sortal. The outcome of this is that the entity in question then persists with loss of status if the conditions that are determined by the sortal that is no longer applicable do not yield the conditions for the persistence of the entity in question. Even if one assumes that persistence is always dependent on a *specific* sortal,<sup>23</sup> this does not mean that every sortal the individual in question falls under provides the conditions for persistence. But ultimately, which sortal it is that provides the conditions for persistence can only be shown by an analysis of our practice of identification and reidentification and a stipulation of the conceptual nature of the sortal in question.<sup>24</sup> If one accepts the thesis that “the death of X” implies “the cessation of the existence of X”, it follows that in the case of the death of an X the conditions for the persistence of this X can no longer be fulfilled. But this means that the existence of this X has ended because his persistence, which is determined by the constitutive sortal, no longer exists. Supporters of the higher brain death criterion who pursue the ontological strategy have only two *prima facie* plausible possibilities of justifying their thesis that the death of the human person is the irreversible loss of higher functions that are supported by the neocortex.<sup>25</sup>

*Option 1:* The loss of these higher functions, above all the loss of consciousness, must be equated with the cessation of the existence of a specific entity (see e.g. Green and Wikler 1981 or Gervais 1986).

*Option 2:* The sortal “personhood” provides the essential conditions for the persistence of human persons (see Veatch 1976 – in the ontological reading).

(Two Options)

---

<sup>23</sup>In this sense, Gervais also criticizes the attempt by Green & Wikler, to define persistence without recourse to sortal definitions: “Kind-essentialism and individual-essentialism are opposite sides of the same coin” (Gervais 1986, p. 120).

<sup>24</sup>For an analysis of this practice cf. Rapp (1995); a more precise analysis of the logic of identity and sortal definitions can also be found in Lowe (1989, p. 61 ff.).

<sup>25</sup>A third option would be to allow the loss of status rather than the cessation of the existence of an entity to be sufficient for death. I will not pursue this further, as such a manner of speaking can only be regarded as “metaphorical”. Taking this metaphorism literally rests on conceiving of death as a social construct and regarding the (social) declaration of death as the event of death (exemplary for this is Wartofsky 1988).

### Option 1

The first option implies substance dualism, since it must be postulated there that a specific substance – the person – ceases to exist although a different substance – the human organism – continues to exist.<sup>26</sup> The foundation of this variant of the ontological strategy is normally, though not necessarily, a neo-Lockean conception of personal persistence in which conscious mental life (memories, intentions etc.) alone provides necessary and sufficient conditions of the identity of persons over time (cf. Gervais 1986 or Green and Wikler 1981). But this approach is confronted with severe problems. Firstly, the Lockean criterion is only suitable for the first-person perspective on personal identity: According to the Lockean analysis, it is for an *x* itself that *x*'s 'identity' stretches through time (cf. Chap. 2). This perspective, which is partly the basis of our moral practice, does not, however, coincide with the approach to personal persistence from the observer perspective (cf. Chap. 5). Moreover, such a conception would also have to clarify the relation between the two substances in question. Under the plausible assumption that higher brain functions are, just like other brain functions, realized in biological functions, a categorical difference seems doubtful. And because of the holistic nature of the human brain and its flexibility, the supposition is implausible that the two processes "personal life" and "biological life" can be distinguished from one another through the separation of their causal basis, the brain regions. Different brain areas can causally support various functions, and if one does not interpret "consciousness" in a limited sense as explicit self-conscious mental activity, it is actually arguable whether the brainstem should also be added to the causal basis of consciousness.<sup>27</sup> But such a limited understanding of consciousness as would be necessary to locate consciousness strictly in the neocortex confuses our first-person experience of our own persistence with the actual existence of this persistence. A limitation of personal persistence to temporal elongation of a conscious stream of conscience is an inadequate reconstruction of our concept of personal identity (cf. Chap. 5). Alongside this unclarity that adheres to the concept of consciousness, one can accuse this defense of the higher brain death criterion of above all not embracing our understanding of persons as human and hence biological entities. The decisive point is that the subject of death comprises both biological and mental subprocesses. The flaw in this conception, which is ultimately due to an inadequate theory of personal identity, is to equate the loss of a component of the integrated life process with the cessation of the existence of the entire life process.

---

<sup>26</sup>Weaker forms of ontological dualism (e.g. aspect- or property-dualism) and epistemic dualism (of perspectives) are not an option here due to the premises of the ontological strategy itself.

<sup>27</sup>The approaches of Gervais (1986) and Green and Wikler (1981) differ in the interpretation of their psychological identity criteria for persons. Whereas Green and Wikler proffer a concept of consciousness oriented on the Lockean model, which is oriented on the propositional stance, Gervais argues a weaker concept of consciousness, albeit without a clear elaboration of this concept (this corresponds to the oscillation between demanding and less demanding concepts of person described in the second chapter). This point is not only central to the question of the time of death, but also gains significance in the framework of the meaning of a whole brain death conception since it concerns the question of the morally adequate treatment of irreversibly comatose patients.

One reason for being ensnared by this flaw is the confusion between ontologically necessary conditions and such features as seem to make continued personal existence *desirable* to us. But our “identifying” ourselves with a specific future status does not coincide with the preservation of our persistence. Inversely, it does not hold that our persistence ends when our future state does not seem worthwhile to us (cf. Martin 1993). The dispute over Parfit’s analysis of the conditions relevant for survival has shown that the constitutive and desirability use of the phrase “what matters in survival” must be differentiated here (cf. Unger 1990, p. 92 ff. and Quante 2012, Chapter 7). Supporters of the higher brain death criterion who wish to defend their approach with the first option confuse the constitutive level with the level of values with which we appraise our persistence (cf. Chap. 5 and Sect. 4.5 of this chapter).

## Option 2

The second option consists in postulating that the sortal in question, relative to which the persistence of human individuals is fixed, is the concept of personhood. Loss of personhood is hence not only a loss of status but also at the same time the end of the existence of a specific human person (cf. Veatch 1976). This argumentation also reveals an inadequate analysis of our practice of identification and reidentification. Persons do not form a natural kind, but rather, the concept of “personhood” defines a merely nominal kind. There is no conceptual reason why others than human creatures, or even artifacts, should not be able to belong to the class of persons. However, the conditions under which e.g. computer persons or even amoebae persons lose their identity are dependent on the structures and properties they share with other computers and amoebae respectively. In this sense, “personhood” is a nominal sortal, because it puts together a class of entities according to criteria fixed by us, without this class having to have uniform conditions of persistence such as exist in natural kinds through scientifically detectable laws that apply to the respective species.

This has been proven by the various thought experiments in the debate on personal identity; there are no conceptually or scientifically detectable regularities that cover the class of persons as such. In fact, the sortal relative to which we have developed our practice of identification and reidentification is that of the “human being” (cf. Wiggins 1980, Chap. 6; Wilkes 1988, Chap.1 and Johnston 1987). This does not mean that only human beings can be persons, though in the question of whether a specific human individual has ceased to exist we allow ourselves to be led by the structures and properties attributed to human beings as such. And humans are constituted by organic subsystems and (as distinguished among them) a mental subsystem whose integration is causally supported by the whole brain. The loss of a part of these subsystems is not sufficient for the persistence conditions to no longer be fulfilled. Hence neither the loss of organic integration through failure of the brainstem, nor loss of mental integration through failure of the neocortex is synonymous with the end of a human being’s persistence. While this entails conceding to supporters of the higher brain death criterion that in defining human death the question is about when a human person has ceased to exist i.e. when the conditions for the persistence of human persons are no longer fulfilled. But it is, however, not conceded that the person is identical with the subsystem “mental life” or that the conditions of persistence are provided by the nominal sortal “personhood”.

#### 4.4.2.2 The Biological Strategy

While criticism is rightly made by the advocates of the higher brain death criterion against someone who, like David Lamb, pursues the biological strategy, it does not apply to the whole death criterion being defended here, since it explicitly refers to the human personality. In the biological approach, in contrast to the biological strategy, the issue is not the death of the human being qua organism, but the death of a specific human individual. This becomes clear when one contrasts this approach with Lamb's position.<sup>28</sup>

According to Lamb (1985), the death of the brainstem is tantamount to the death of the human being, as the brainstem causally supports the integrative function for the life process of the whole organism. The failure of this function not only leads to the failure of other higher brain functions, but also means the end of an integrated life process. Inversely, the retention of the integrated life process during failure of the higher brain functions is enough to preserve the persistence of the human being. While I will presently justify my acceptance of Lamb's second thesis, I would first like to unfold my criticism of his equation of brainstem death and the death of the human being.

From the viewpoint of the whole brain death conception represented here, the objection must be raised that a brainstem death conception ignores the integrative achievement accomplished by the neocortex in the form of (personal) consciousness. This can be illustrated by two examples, a fictive one and a possibly real one.

- (a) A thought experiment: Imagine a human individual, whose actual stream of consciousness is being constantly maintained<sup>29</sup> and is meanwhile implanted with a brainstem prosthesis. In such a case there is no reason to doubt the persistence of this human being. The persistence of the body (with the exception of the replaced brainstem) linked to the stream of consciousness certainly suffices to guarantee the persistence of this human organism. The thesis that before and after the implantation two numerically different human individuals are involved is implausible. Neither is the information that the integrating function of the vital processes previously maintained by the brainstem will now no longer be achieved spontaneously by the organism itself convincing. For an adequate measure of integrative achievement is generated as long as this human is able to develop consciousness. In our thought experiment, only the material basis in which the integrative achievements are realized has been partly replaced. The pointer to the lack of spontaneity is presumably to suggest the deduction that such a replacement leads to the entity in question not being a human organism at all after the implantation.

---

<sup>28</sup> Becker's position (Becker 1981) is not considered in the following, because he integrates evaluative aspects in his conception and thus does not represent a purely biological strategy. Gervais criticized this inconsequence of Beckers convincingly, whereby she welcomes the evaluative aspects by virtue of her own premises and criticizes the biological aspects (cf. Gervais 1986, p. 58 ff).

<sup>29</sup> How this condition can be fulfilled technically need not be discussed here. In such thought experiments, as the literature shows (cf. Unger 1990), the sky's the limit.

In this context, Lamb (1985, p. 61) proposes the following thesis: “The idea of brainstem transplant is both a conceptual and practical impossibility.” However, this statement is too strong. It is undisputed that at present and maybe for all time such a prosthesis is beyond the scope of human technology. The factually technical unrealizability cannot, however, suffice on the level of conceptual elucidation. But the claim that such a prosthesis is conceptually impossible is misleading on the one hand and wrong on the other. It is misleading because it is not the concept of such a prosthesis that is contradictory (why should it be?). At the most, the claim that the persistence of a human organism could be maintained by means of such a prosthesis could be contradictory. But this would only be the case if the implantation of such a prosthesis led to the conceptual consequence that we are then no longer dealing with a human organism. But in my view our intuitions regarding this thought experiment do not speak for this assumption.<sup>30</sup> In contrast, should it be the case that the condition demanded of the implantation of such a prosthesis with maintenance of the actual consciousness is unrealizable in the thought experiment because of natural laws, then that would really make it impossible to maintain the persistence of a human individual by means of a brainstem prosthesis. But this would be a case of necessity that is not conceptual but a law of nature necessity.

- (b) The real case: If one takes the total locked-in syndrome as a real, occurring situation in which after total failure of the brainstem function the integrative functions of the neocortex are still being maintained and achievements of consciousness are present, then at least for a short period of time exactly the same situation exists as has been outlined as permanent in the thought experiment. Assuming that the total loss of the brainstem functions also quickly leads to a breakdown of the higher brain functions, then one can view the brainstem death criterion as a prognostic criterion for the imminent occurrence of death (= failure of the whole brain). But this must not be confused with the occurrence of death. The imminent occurrence of death belongs to the phase of dying and thus to the life of an organism and must not be identified with the death event itself. The error in the biological strategy consists in equating a partial loss of the integrative function of the brain with total failure thereof. The reason for this error is the same as in the ontological strategy. Both assume the premise: “either organism or person”, and then opt for one of the two alternatives. In a second step they identify the part of the brain that counts as the basis of realization of

---

<sup>30</sup>In fact, these intuitions argue for the maintenance of mental ‘life’ with replacement of the brain stem and sundry persistence of the body being adequate for the persistence of the human individual. And indeed, without further argument, Gervais (1986, p. 32) infers from this finding that consciousness is at the same time also a necessary condition for human persistence. Through this error she mistakenly leads over from the justified criticism of Lamb’s biological strategy to her own ontological strategy. Her error becomes understandable when one bears in mind that Gervais starts from the premise “either organism or person” – her (implicit) argument then takes on the following form: If consciousness is adequate for the persistence of a person, then the organism cannot be the adequate subject of death. For the organism can continue to exist without consciousness. So it must be the person to whom the definition of death refers. Without the questionable premise, which I do not accept, this argument is not conclusive.



the respectively selected option and in a third step they ignore the integrative achievements attained by the respectively omitted part. In their reciprocal criticism the biological and ontological strategies point out that the lapse of the integrative function assumed by the respective opposing side cannot be equated with the death of the human individual because the ignored part is sufficient to maintain persistence.

From the viewpoint of the biological strategy and its concept of whole brain death represented here, the criticism applies to both alternatives. Neither the lapse of the brainstem nor that of the neocortex is sufficient for the death of a human individual. This position becomes possible because it rejects the premise “either person or organism” that is shared by the biological and ontological strategies as a false alternative. Thus, the equation of the biological strategy and my biological approach imputed by the supporters of the ontological strategy is unfounded. The justified criticism of the concept of brainstem death is not at the same time appropriate to the concept of whole brain death. Only if one accepted the premise rejected here that one must choose between the person and the organism as two distinct subjects of death would the criticisms coincide. But since this alternative does not exist as such for reasons that have been shown in this study, the conception of whole brain death remains untouched by the objections being bandied between the biological and ontological strategies. However, further objections and open questions still remain, which shall now be discussed.

## **4.5 Objections and Open Questions**

One aim of this chapter was to present the dispute about human death and to characterize and criticize the most important argumentative strategies being followed in the debate. A second aim consisted in the analysis of the conditions of adequacy which had to be met by a definition of the concept “death”. The answer I suggest thereto is open to certain objections which should now be discussed (Sect. 4.5.1). In addition, some questions must remain open in this study, the answer to which is indispensable for dealing comprehensively with the problem. Since this cannot be achieved here, the deficits and burdens of proof shall at least be explicitly named (Sect. 4.5.2).

### **4.5.1 Objections**

In the following I would like to discuss six objections, whereby I formulated the first two whilst the third and fourth were raised by others. The last two objections rest on thought experiments for which the proposal developed here seems to provide inadequate answers.

1. *Who is the corpse?* The definition of death proposed here implies that no persistence remains between an organism and the corpse that remains as so-called “mortal remains” after the death of this organism (also Lockwood 1985, p. 11). Our everyday talk that is used, for instance, in crime films when someone identifies a dead individual by means of the corpse cannot therefore be analyzed as a statement of identification. So whoever glances at a corpse and says: “Yes, that’s XY” cannot make a statement of identity according to the definition of death proposed here. This objection is sound, but it shows that my proposal has a counterintuitive trait and that the limit of the mere reconstruction of our everyday concept has been reached. The assumption of an identity between the body in which the integrated life process of an organism has taken place and the corpse might be suggested on the surface of our everyday verbalization. But it is neither necessary for the justification that a human corpse is entitled to a certain ethical protection (cf. Ach et al. 2000, Chap. 3), nor does the surface structure of our everyday utterances withstand philosophical analysis at this point (cf. Rosenberg 1983, p. 27 ff.). What is more, this assumption of identity can also be explained psychologically in that the body of a human being and the relevant corpse are similar from the qualitative point of view. The willingness to identify the corpse with the deceased person dwindles to the same extent as that similarity (think of e.g. an urn filled with ashes or a corpse that has been almost completely dissolved in acid). Hence the conceded counterintuitivity of the proposal introduced here is by and large justified.
2. *Mental integration?* In the above reflection, two propositions which are prima facie incompatible have been put up and defended. On the one hand I claimed that death can be understood from the observer perspective and should therefore be defined free of evaluative aspects. On the other hand, the continuation of a conscious or even self-conscious ‘life’ is viewed as an integrative achievement of the organism that is adequate for its persistence. But quite obviously the so-called unity of consciousness or self-consciousness is only understandable in the participant perspective. The objection that this is a contradiction suggests itself, but can be eliminated. While it is correct to say that the concept of integration cannot be analyzed without evaluative aspects i.e. without recourse to the comprehending participant perspective, according to my prerequisite that these human capacities and achievements are realizations based on naturally describable causally enabling conditions, it now applies that the integrative achievements of the neocortex are in fact these same ascertainable natural achievements in the observer perspective. If one speaks of the integrative achievement of the neocortex, one does not refer to the evaluatively tinged unity of consciousness or self-consciousness comprehensible in the participant perspective but to the integrative achievement that is explainable and functionally graspable in the observer perspective and constitutes the natural basis for this unity (cf. Hurley 1998).
3. *Brain dead expectant women?* In opposition to the whole brain death criterion, it is often argued that in brain dead expectant women the pregnancy can be sustained with intensive medical care (cf. Stoecker 1999, p. 87 ff.). This objection can either be understood in a way such that one views the functions and pro-

cesses that have to occur during a pregnancy as adequate to form an integrated human organism. But this is implausible, since the remaining integrating functions of the brain dead woman are not sufficient for counting as a self-integrating organism itself. Or one means that the entire apparatus alongside the staff caring and coordinating the various mechanical supports together with the life processes still occurring in the brain dead expectant woman form this unity. However, in the latter case, it is, for one thing, misleading to say that such a brain dead expectant woman is herself in a position to sustain her pregnancy. And for another, the entire constellation of caregivers, machines and life processes that occur in a corpse are not an entity that can be described as an integrative life process. When Stoecker deduces from the possibility of intensive medical continuance of certain functions and life processes that the brain dead individual is herself “in a position to maintain her structure” (ibid. p. 87), then he transfers the location of the integrative achievement incorrectly to the corpse, without taking the entire constellation into account.<sup>31</sup>

4. *Brainless fetuses?* It is incontestable that a human individual does not have a functioning brain in the first stages of its individual existence. But since the embryo must be taken to be an integrated life process even before this time, this objection cannot be parried in the same way as the third objection, although Stoecker (1999, p. 87 fn. 47) groups them together. In its pure form, the problem arising for the proposal developed here through the stage without a developed and functional brain in the existence of a human organism is formulated by Ach (1999, p. 132 pl.) thus: If the whole brain death criterion is defended in the framework of an organism-centered conception of human persistence, it follows that the human individual only starts to exist when the brain has developed and takes over the integrative achievement. This objection does not have as its target that the whole brain death criterion is an inoperative diagnostic criterion for the first stage in the life of a human individual. On the contrary, this objective points out a counterintuitive trait and “a more (...) than merely terminological curiosity” (ibid. p. 134) of my conception. Even if I agree with Birnbacher that the answer to the end of life of a human individual “does not prejudice the difficult question of the beginning of life” (Birnbacher 1994, p. 37 f.), I am of the same opinion as Ach that the above objection, should it be the case, shows a defect in the theory presented here. For, in my view, it is one of the strengths of the biological approach as opposed to psychological ones that it does not have to answer the question of the beginning of life of the human individual in a counterintuitive way (cf. Chap. 3 and Olson 1997). But this objection goes astray for several reasons.

*For one thing*, three questions must be differentiated: Firstly, one can ask if a specific life process is human life. Secondly, one can ask whether a specific human life process is one integrated organismic life process or an interference of several life

---

<sup>31</sup> It would be a different case if a human, in failure of the brain stem, were to organize this coordination through own actions and instructions with the aid of technical means. However, here the difference between spontaneous and artificial is not decisive, but rather, the fact that independent organisms cannot be a proper part of another organism.

processes. And thirdly, one can ask what the conditions of persistence are that make it one and the same organismic human life process over various temporal intervals. As already became apparent in Chap. 3, the theory proposed here does not provide an answer to the first question, but concentrates solely on the interlinked second and third questions. Totipotent human cells are e.g. clearly human life, but presumably not human organisms (this depends on the activation of their genome). If I, like Ach, were committed to the thesis that a specific human individual only starts to exist after the brain has developed and started to function, then my approach would imply a consequence that is hard to accept. However, this implication does not exist, as, *for another thing*, the thesis that from the moment a human organism has developed a brain and it has taken over the integrative function, the whole brain death criterion applies, is compatible with the thesis that before this moment the integrative function is achieved by the organism in a different way.<sup>32</sup> The definition of death demands only that the organism effects this integrative achievement, not that this achievement is effected by the same organ at every moment of her existence. The thesis of continuity requires only that there is no interruption in this total integration during the change-over. Beyond this, the bond of the persistence to the causal laws pertinent to the respective species requires the change-over of organs effecting the integrative achievement to be law like. But all this seems to me to be implicit in the case of normal development of the human fetus.<sup>33</sup> It is simply an empirical fact that in human beings, when the brain has developed it is not the case that a different organ can compensate the failure of the brain as a whole. And further empirical data is needed to determine how the integrative achievement of the human organism is achieved before the brain develops and takes over this function. The unacceptable consequences of my proposal feared by Ach do not follow from all this.

5. *Science-Fiction – first scene*: Imagine, as can be seen in many essays on the problem of personal identity, that humanity has developed the technique of constructing prostheses that functionally equate the brain, onto which – thanks to yet another revolutionary technical miracle – one can transfer all information “stored” in the brain of the recipient shortly before the replacement of the imminently failing original brain by the thus prepared prosthesis takes place.<sup>34</sup> While the operation is carried out, the integrated life process including the actual stream of consciousness has to be interrupted. After the operation a person wakes up

---

<sup>32</sup>The formulation of the criterion of death in my earlier proposal (cf. Quante 1995b) is not quite blameless for the false impression arising here, since the definition “from the point in time at which the organism took over the integrative achievement” was not contained in the original version.

<sup>33</sup>The Anenzephalics, likewise also cited by Ach (1999, p. 133) as a problem, do not in my opinion fulfill the conditions necessary to allow them to be counted as a human organism. But of course it is not impossible for the abnormal course of development of a fertilized human egg cell not to lead to the formation of (at least) one human individual, but only to human life in the form of inadequately integrated human cells (hydatidiform mole).

<sup>34</sup>Whoever can moreover imagine a procedure to cleanse a used organ of its data trash, can replace the prosthesis with a ‘recycled’ donor brain. However, it will have to be a living donation (with fatal consequences), since the failure of the whole brain does here count as criterion of death.

who – with the exception of the brain prosthesis – has the same body as the person who underwent the operation. Moreover – according to the prerequisite – she possesses the same knowledge, attitudes and abilities.

If there were an unlikely chance that it could be possible technically or maybe even according to the law of nature, then this would be a remarkable state of affairs, since, according to the conception proposed here, the pre- and postoperative person would be numerically different entities.<sup>35</sup> Now, in view of this example, the general inference that according to my approach the implied assumptions of persistence are put out of action by the technical intervention, so that our intuitions are uncontrolled in such scenarios, is indeed valid. But in addition, this example shows vividly just how important it is to differentiate between the evaluative identification and the question of persistence. For from both the viewpoint of the preoperative person who asks herself whether she should allow this operation to be carried out and that of the postoperative person, all the valuable aspects that make our personal life so important – with the exception of persistence – are ensured (cf. Chap. 5). If one keeps the levels of causal analysis of persistence and that of evaluative personality separate, one can resolve the apparent counterintuitivity that inveigles the theory that identity is not the point (Parfit 1989), or survival is a question of values rather than causality (Martin 1998).

6. *Science-Fiction – second scene*: The consequence that it concerns two numerically different ‘patients’ was forced in the first scene by the assumption that the implantation of a brain prosthesis is only possible with an interruption of the integrated life process. What happens when one imagines the even more utopic scene in which a gradual replacement of the brain is effected while maintaining the integrative function?<sup>36</sup>

Presuming that during the gradual replacement there is factually no fusion or fission, that this replacement is therefore not a copying process, in which the master copy is preserved or in which several copies are made, then all the conditions of persistence required by my approach are present. Again, in this case it also applies in general that the scope of our persistence criteria is vacated during the technical intervention and other intuitions, which in the end rest either on the particularities of the first-person perspective or on evaluative ‘identification with’, ‘entice’ us to our judgments of identity. The loophole, which consists in asserting that an entity with an artificial brain prosthesis is no longer an organism, because the integrative achievement does not occur spontaneously, is not open to me. For one thing, the spontaneity is an achievement of the organism qua life process, and for another, the structures necessary for these achievements can be realized in various material

---

<sup>35</sup>The question of whether the postoperative entity is in fact an organism will be discussed in the context of the second thought experiment.

<sup>36</sup>Pollyannish, one could imagine that every brain cell being replaced one after another with brain cell prostheses which execute their substitutive function straight off. Such an operation, which would admittedly take rather a long time, could then even be carried out with full consciousness and without anesthetization.

bases. If there are artificial bases by means of which an organism can conserve its integrative unity, that does not contradict its being an organism. It must hence be conceded that this second scenario of the approach proposed here does reach its limits. However, I consider a philosophical model which only reaches its limits through such a scenario to be nevertheless very powerful. Moreover, my approach permits the systematic reason for this limitation to be specified. Under the conditions of the above thought experiment, the boundary between organism and artifact becomes suspicious. However, this differentiation belongs to the essential prerequisites of our discourse on the persistence of human persons. If the technical options imputed in the thought experiment really did exist, then these basal categories of our process of identification would lose their footing, because contingent conditions are written into them which would be annulled by these technical options.

### 4.5.2 *Open Questions*

In his criticism of my first attempt to develop an appropriate definition of death, Ach pointed out that the terms “organism” and “integrative life process”, while “they might be well anchored in the *Lebenswelt*,<sup>37</sup> [but] they are theoretically difficult to handle” (Ach 1999, p. 133). This broaches a problem which cannot be dealt with further in this study. My argumentation commits me to the theory that one can render the concepts of organism and integration in the observer perspective precisely enough for them to be “manageable”. Ultimately, this thesis bases on my conviction that these concepts are indispensable not only for our manifest image but also for biology.<sup>38</sup> At this point, however, further clarification of these concepts must be omitted. Alongside the concepts of organism and integration, I also consider the concept of life to be indispensable if a definition of death is to be established. In contrast to the closer analysis of what constitutes “human” life, the above theses commit me to specific conceptions of life which can likewise not be rendered more plausible here. The definition of the organism as a self-integrating life process and the assumption that organisms cannot survive the stoppage of this process, articulate that life is not understood as a capacity but as the performance of this capacity. With this processual and actualistic conception, neither the syntropic conception proposed by Rosenberg (1983, Chap. 4) nor the morphological conception proposed by Stoecker (1999, p. 78 pl.) is compatible, as both relate to capacities.<sup>39</sup> Ultimately, whether, in determining the cri-

---

<sup>37</sup>“*Lebenswelt*” means the physical and social world understood from within, i.e. taken from the participant perspective.

<sup>38</sup>Cf. on this also the literature quoted in Chap. 3.

<sup>39</sup>As far as I can see, in Stoecker the capacity conception of life is a main reason for rejecting the brain death criterion. As one can see from his treatment of “discontinued life” (Stoecker 1999, p. 80 ff) he considers it possible for an organism to survive the interruption of the integrated life process. Thus, due to his definition of life as regards the freezing of organisms, he reaches a radically different conclusion from mine (on the problem of interrupted life cf. also Feldman 1992, p. 60 ff.).

terion of death, one refers to the performance of functions or to the structures that form their material basis of realization depends on the concept of life on which it bases. Whilst I, like Stoecker (ibid. p. 68), am of the opinion that a suitable definition of death cannot be given independently of the definition of life, the determination of “life” in my view does for one thing only supply the necessary constraints for answering the question of when a human being is dead. And for another, the definition of “life” and “death” also relates to the question of what an organism is. The processual conception assumed implicitly here is incompatible with the morphological conception proposed by Stoecker and leads to conflictive positions in respect of the definition and criterion of death. Thus my reflection as regards the problem of a definition of “life” is not neutral. However, an explanation and justification cannot follow in the framework of this study, so that a central issue remains open at this point.<sup>40</sup>

## 4.6 Persistence and Biomedical Ethics: A Brief Conclusion

The use of the theoretical framework developed in the second chapter for questions concerning the beginning and end of life of the human individual lead to other philosophical problems. But even if the attempt proposed in the second chapter to display the relevance of the biological approach for biomedical ethics does leave a string of questions open, in the past two chapters three things have nevertheless been achieved: *Firstly*, my biological approach could be rendered textually more precise through the relation to the questions concerning the beginning and end of life. *Secondly*, it was shown that both the differentiation between human persistence and personality and the more specific definition of the former in the form of a biological approach are of interest for questions of biomedical ethics. For one thing, the differentiations introduced in the first chapter can help to structure the partly rather convoluted debates. For another, the suggestions for solutions developed thereby result in consequences in the field of biomedical ethics and are therefore ethically not without consequences. And *thirdly*, the theoretical conception of a purely descriptive analysis of human persistence is supported by just these ethical implications. For, on the basis of the biological approach, it can be shown that there are good reasons for not treating the questions of the beginning and end of life as social and moral problems. Hence the basal intuition that the beginning of life and death are purely descriptive facts is allowed for, just as the no less basal intuition that there can be no pluralism and relativism in these questions. Such pluralism would, at least in our culture, have to lead to extensive revisions of our social practices. But it cannot be envisioned that these would provide overall better conditions for a good life. So the biological approach is argued for not only by greater coherence with our everyday intuitions but also in the end maybe even by an ethical argument.

---

<sup>40</sup>Stoecker (1999, p. 75, Fn. 30) draws on John Locke as a historical representative of the thesis that life consists of specific capacities. This seems to me to be historically incorrect, since Locke understands the organism to be a continuous life process.

# Chapter 5

## Personality and Autonomy

In the first chapter, I differentiated four questions hidden behind the label “personal identity”, including the distinction between the question of persistence and the specific way in which persons know about their own existence over time. This knowledge specific to persons will now also be discussed within the conception of personality. Thus the answer to the question of persistence proposed in the second chapter will be augmented by an aspect of first-personal biographical ‘identity’ that is specific to persons. This is based on the assumption that the specific way in which persons behave towards their own existence over time is manifested in the person’s personality. So the relation between personhood and personality can be determined in two ways. For one thing, one of the capacities necessary for personhood to be attributed will form the core of the personality: this is the knowledge of persons about their own identity over time and their ability to take a stance towards their own temporally extended existence. For another, one can understand the personality of an individual as the respectively individual way of using the properties and capacities accompanying the personhood to cultivate an own biography and to lead an own ‘personal’ life. *Personhood* denotes a status attributed equally to every individual qua person, if he adequately fulfills the required conditions. In contrast, *personality* denotes the respective individual shaping of the personhood which is manifested in an individual biography (cf. Quante 2012, Chapter 8).

### 5.1 Targets

The reflections in this chapter do not claim to be a comprehensive theory of personality or autonomy. Springing from everyday pre-understanding, some of the general and ‘formal’ aspects relevant to the overall course of argumentation in this study will be discussed in the following. The aim is to illuminate this aspect of personal life to the extent needed to facilitate adequate handling of the question of the person’s identity over time. Thereby, proving the evaluative nature of personality



should serve as a first step towards answering the question of whether the identity of the person can be a principle of biomedical ethics. For this, the evaluative dimension of personal identity in the sense of personality will be unfolded in the second section. Subsequently it concerns the connection between personality and autonomy (Sect. 5.3) and the proof that the principle of respect for autonomy which is accepted as far as possible by medical ethics refers to the evaluative dimension of personality (Sect. 5.4).

## 5.2 Personality

In this section, the first step (Sect. 5.2.1) is to underscore the clarifying function and objective relevance of the distinction between personal identity in the sense of persistence and personality respectively, on the basis of Parfit's assertion that personal identity is not what matters. Subsequently, in the second step, the aspects of persons' biographical identity which are rightly claimed by the simple view compared to the complex strategies and which cannot be grasped in the observer perspective will be integrated in the conception of personality with her underlying participant perspective (Sect. 5.2.2). In the third and final step, two forms of identification are differentiated, whereby "identification with" proves to be the core of the evaluative-voluntative self-understanding (Sect. 5.2.3).

### 5.2.1 *Is Identity Important?*

Possibly the most provocative thesis formulated by Parfit in his overall provocative answer to the question of what constitutes the identity of a person over time states that „Identity is not what matters“ (Parfit 1989, Chap. 12). The question of whether identity is even important for our understanding of ourselves as persons surfaces in the framework of Parfit's defense of a complex theory of personal identity in which he defends a psychological criterion of persistence against both a physical criterion and the simple view. Parfit thus formulates a variant of complex theories which employ the concept of person within the analysis of persistence. Hence, for him, in contrast to the biological approach, the thought experiments about fission and fusion are relevant. On the other hand, the first-person perspective and evaluative aspects also enter into his analysis.<sup>1</sup> Parfit's argument can be presented as follows (cf. also Brueckner 1993):

(P 1) Identity is a relation which is always unambiguous.

---

<sup>1</sup> Even if Parfit's approach is inadequate for the reasons explained in the second chapter, his reflections are nevertheless instructive, because in them the tension between the level of persistence and that of personality becomes visible; for a comprehensive presentation of Parfit's theory cf. Herrmann (1995) in which its ethical consequences are also discussed.

- (P 2) A psychological relation PR is constitutive for the identity of the person over time.
- (P 3) PR allows cases of fission and fusion.
- (P 4) In thought experiments, situations of fission and fusion can be described in which everything relevant for person A as regards her existence over time is fulfilled. These are the cases in which PR exists.

- 
- (C 1) PR is not identity. [from P 1 and P 3]
- (C 2) Identity does not matter to a person as far as her existence over time is concerned, but PR does. [from P 2, P 4 and C 1]

(Parfit's Argument)

Parfit can argue this way only because, as will be shown presently, in the framework of a complex view, he employs the concept of person which transports both the first-person perspective and evaluative aspects. This leads to an ambiguity in Parfit's conclusions. Unger (1990, p. 93 ff.) has – following Perry (1976, p. 85 f.) – shown that Parfit's question: "What matters in survival?" is ambiguous. He differentiates three meanings of "what matters": first the sense of desirability, second that of prudence and third the sense of constitution. In the desirability sense, which e.g. Lewis (1976, p. 17) applies, "what matters" is understood as everything which is desirable and makes the continuation of a person's existence more valuable as opposed to the non-continuation of her own existence. In the prudence sense, "what matters" subsumes everything considered relevant to the rational self-interest in the person's survival. In contrast, in the constitutional sense, "what matters" denotes precisely those conditions which must be fulfilled so as to guarantee the person's persistence.

In his discussion of Parfit's thesis that identity does not matter, Perry (1976, p. 85 f.) distinguishes between the sense of desirability and that of constitution. My proposed division between persistence and personality and Perry's differentiation aim at the same point. The sense of desirability and that of prudence both belong to the dimension of personality made accessible through the participant perspective and essentially involve the first-person perspective. In contrast, the sense of constitution aims at persistence and is thus dependent on the observer perspective. Even if one can for certain purposes differentiate even further between the senses of desirability and prudence, the decisive ambiguity of Parfit's reflections still lies between the evaluative level of personality and the descriptive level of persistence.<sup>2</sup> So I will

---

<sup>2</sup>At least, this applies to the way of structuring the problem area suggested in this study. The option proposed therewith, that there should be differentiation between the evaluative dimension of personality and the dimension of persistence is often overlooked in the analytical debate. For Parfit there is only the alternative of a complex theory oriented on the physical criterion or the simple view, alongside his complex theory oriented on the psychological criterion. Lewis, Perry and Unger also remain within this spectrum and use the concept of person in the framework of their respective complex theory. In so doing, they integrate evaluative aspects into their analyses of persistence (e.g. Unger 1990, p. 97) and thus merge the various levels I suggested above. In this

differentiate in the following between an evaluative and a constitutional sense of what “matters”.

Unlike Lewis, Perry or Unger, who all operate with the concept of person in the analysis of persistence and thus cannot definitively separate the question of the identity of the person from the persistence of the human individual, the ambiguity of Parfit’s thesis that identity is not what matters to a person as regards her own survival can be disambiguated in my theory as follows:

- (Q 1) Does identity in the sense of persistence matter for personality in the constitutive sense?
- (Q 2) Does identity in the sense of persistence matter in the evaluative sense?
- (Q 3) Does (the identity of the) personality matter for persistence?
- (Q 4) Does the identity of the personality matter?

(Four meanings of Parfit’s thesis)

Parfit himself understands his thesis that identity does not matter for the identity of the person as a negative answer to both the first and second questions equally, because he does not distinguish between the two.<sup>3</sup> For him, they are both cases of the identity of person which are in no way cases of human persistence (teletransport etc.) and cases in which everything evaluatively important for the person is present without human persistence being present. He affirms the third question since he conducts his analysis of persistence with the help of the concept of person. As far as I can see, he does not make the fourth question a subject of discussion in his theory. However, one can deduce implicitly from Parfit’s reflections that he affirms it, since he considers the constitutive psychological properties and capacities of a person, her beliefs, wants, intentions and memories to be evaluatively relevant. It is indeed the survival of these personality traits which allows a person to evaluate a future

---

framework, Lewis (1976) and Perry (1972) try to submit analyzes of persistence which also allow the identity to be retained in cases of fission (cf. also Parfit’s counterarguments (in Parfit 1976) and the discussion of alternative options in Maddy 1979 and Quante 1995a). Differently from Unger or Lewis, Perry (1976, p. 75 ff.) sees clearly that the evaluative level of personality, for which Perry uses the concept of identification, must be distinguished from the descriptive level of persistence. His reflections intimate that the unambiguousness of the persistence relation in the derivational sense is important for the personality (ibid. p. 84 ff.). Nozick’s „closest-continuer-theory“ provides a further alternative. Whilst I submit the proposal that the level of persistence and that of personality should be isolated from one another, Nozick (1981, p. 29–70) goes in the opposite direction and develops a concept of persistence in which the evaluative and descriptive aspects are correlated in a complex way and integrated into a genuine concept of identity over time. But I doubt whether Nozick’s theory is really the best that „Parmenides can do in an almost Heraclitean world“ (ibid. p. 46). Whether his approach or the one proposed here is preferable as a whole will have to emerge from the probation of my proposal in all the contexts discussed here.

<sup>3</sup>Thus, Sosa (1990) can rightly assert against Parfit’s argumentation that there is a flaw in it. This is that the proof that a certain fact is not relevant in the constitutive sense does not already carry proof that a fact, which is insignificant in the constitutive sense, must also be irrelevant in the evaluative one. There is always the possibility that a fact that is insignificant in the constitutive sense must be considered vital in the evaluative one. With this objection, Sosa demonstrates that Parfit jumps from the analysis of the first question to a response to the second one without making this explicit.

condition as positive, even if she is convinced that she is not identical to the individual instantiating these traits (cf. Quante 2013a).

Objections to Parfit's negative responses to the first two questions are meant to show that identity in the sense of persistence is of importance for the personality in the sense of the biographical 'identity' of the person. Indeed, both Unger (1990) and Perry (1976) or Doepke (1990) share Parfit's assumption that personality and personhood are relevant when answering the question of persistence. So it is not always clear from their answers whether they want to criticize Parfit's thesis as a negative response to the first or the second question.<sup>4</sup> Thus, Perry or Doepke defend the identity of the person in the sense of persistence for the reason that a person can, in the case of persistence, anticipate to a high degree that her future plans will be realized. In addition, says Perry, there are not just future projects one would like to see actually coming into fruition, but also those one really even wants to implement oneself. Whoever wants to become a grandmaster of chess, an Olympic champion in synchronized swimming or soccer world champion wants to achieve this in person and not leave it to someone else, even if the other person is in a position of psychological relatedness. Whilst one can understand Perry's argument primarily as an answer to the second question, since he says nothing about why the identity in the sense of persistence increases this probability, Doepke's defense of the importance of identity in the sense of persistence amounts to the control of the person over future events being guaranteed or increased therewith. So he is answering the question more in the sense of the first interpretation.<sup>5</sup> Unger's argument (1990, p. 269 ff.), that identity in the sense of persistence matters to the person because it preserves the "focus of one's own life", switches between the two interpretations. At times he seems to want to claim that the possibility of "leading" one's life as a person (ibid. p. 269) depends in the constitutive sense on there being no cases of fission. But somewhere else he seems to think that the condition of identity (in the sense of persistence) has rather more weight in the evaluative sense, which is part of our understanding of how to lead a life in a personal way (Unger 1990, Chap. 9).<sup>6</sup> The reason both Parfit in the formulation of his thesis that identity does not matter for the existence of a person over time and those supporters of complex theories who object

---

<sup>4</sup>Lewis (1976) seems to be only interested in the proof that there is a logical possibility of interpreting the problem cases described by Parfit in such a way that one can hold on to the concept of identity. So he only counters Parfit with a "non sequitur" without making the differences I consider decisive the subject of discussion. Thus, his contribution is, just like those of Perry (1972) or Noonan (1991), primarily interested in the logical questions associated with the concept of identity applied to intervals of time.

<sup>5</sup>Wiggins (1979) argues that the worry about existing at a future time cannot be present without the worry about existing at every time in between. Hence he asserts the constitutive function of persistence for the evaluative worry about one's own existence over time.

<sup>6</sup>In all defenses, the first-person perspective thereby plays a decisive role. But even Parfit has to pit the first-person perspective of the person against the analysis of persistence in order to make his converse thesis plausible. So it is significant that his inference that not identity, but only the conservation of the central psychological personality traits is decisive is only plausible when one looks into the past after the fission. It's a totally different state of affairs for the individual worried about his future before fission (cf. Sosa's criticism 1990, p. 319).

against this thesis do not differentiate clearly between the interpretations (Q 1) and (Q 2) is because they assume, like Parfit, that the concept of person is suitable and relevant for an analysis of persistence (thus affirming Q 3). In this way the first-person perspectives and evaluations come into play via the concept of person and cases of fission or fusion cannot be excluded.

This differs on the basis of the biological approach proposed here. On the persistence level neither evaluations nor the first-person perspective, nor yet the concept of person are to be used. Moreover, cases of fission and fusion are not possible with retention of persistence, so that these thought experiments make no clear sense in the framework of the biological approach. In contrast, on the level of personality, one is dealing with an evaluative phenomenon constituted in the participant perspective. So what form do the answers to the various questions which can be given from this approach take?

In fact, the biography of a person constituted from the first-person perspective in a social context cannot be fully grasped in the observer perspective. However, one can meaningfully understand the first question as asking which constitutive contribution the factual nature of human persistence makes to the human personality. The answer to this question is unequivocal: The constitutive conditions for human persistence are causally enabling conditions for having adequate properties and capacities for personhood and personality. In addition, they also provide the framework in which the individual and social shaping of the personality is realized (cf. Chap. 9). The second question, which refers to “important” in the evaluative sense, can only be understood as the question of how the factually valid conditions of human persistence are relevant for the human understanding of personhood and personality from within the participant perspective. At this point only a formal answer can be given. On the one hand, our understanding of personhood and personality are affected by facts pertaining to human persistence. One sees this e.g. in that a complex theory of personal identity permitting fissions conflicts with both our ethical practice of ascribing responsibility and our conceptions of death. Further areas in which the factual constitution of our biological existence affects our understanding of personal life are e.g. sexuality and reproduction, or even quite generally the concept of ill health.<sup>7</sup> In the following chapters some contexts in which the relevance of the conditions of human persistence is shown in the framework of the evaluative concept of personality will be discussed. In my view, the fact that in the framework of the approach proposed here, (Q-2) can be reconstructed as a question of the relevance of the biological nature of human beings for his self-understanding as a person supports this segmentation of the problems, whereby it should be noted that it is a descriptive question which must be distinguished from the normative question to

---

<sup>7</sup>These examples can be accommodated in the following general pattern: The evaluative relevance of the factual nature of human persistence can be felt everywhere where the hitherto valid facts can be overridden by technical possibilities. The ethical conflicts ensuing from this can be understood as an erosion of our moral ontology. They lead to the question of whether we really want the changes in our understanding of personhood and personality that are brought about by the invalidation of the hitherto valid natural base.

what extent the factual biological basis, e.g. in biomedical ethics, should also have ethical significance (cf. Chap. 9 on this). The answer to the third question is in the negative: the concepts of person, personhood or personality cannot be “important” for the analysis of human persistence either in a constitutive or evaluative reading, because their character is essentially tied to the evaluative and participant perspective. This follows directly from the limitation of the analysis of persistence to the observer perspective, so that not only can a differentiation of the third question in both interpretations of “important” be passed over, but neither does the question of whether the identity of the personality might be important for persistence have to be debated. The fourth question, which is only mentioned implicitly in the other approaches, can only be specifically discussed here at all because of the clear division between persistence and personality. It can then be understood as the question of whether any value is attached to having personhood and a personality in the participant perspective. In view of the central meaning of the concept of person for ethics, an affirmative answer is inevitable. But even the interweaving of personhood and autonomy with a personality speaks for the development of personhood and personality in the life of a human individual being attributed high positive value.

### ***5.2.2 Personality and the Question of the Identity of the Person Over Time***

The biological approach waives both the first-person perspective and the concept of person in the analysis of the persistence of a human individual. The analysis of human persistence, tied as it is to natural kinds and causal laws, can be conducted entirely in the observer perspective i.e. purely descriptively. In this way, the concept of person which refers to the designations of personhood and personality is reserved for an analysis oriented to the participant perspective into which both evaluative and first-personal aspects of personal life can enter. Whereas answer of the biological approach to the question of human persistence is complementary to the question of the special way in which persons know about their own existence over time, the objection must be raised to the other variants of the complex view that in the framework of the observer perspective they attempt to operate with the concept of person which is unusable for this perspective because it is tied to the evaluative and first-person perspective. So it is no wonder that these attempts to answer the first question with the aid of the concept of person result in the identity of the person being underdetermined or even indeterminate. Psychological, physical and combined theories are therefore, in contrast to the biological approach, not compatible with the problem segmentation proposed here, as they require a quasi naturalistic understanding of the personal aspects.<sup>8</sup> The dissent between the simple and complex

---

<sup>8</sup>A comparable distribution underlies the discussion by Wolf (1986) and the criticism by Siep (1987) of the analytical discussion on personal identity. Viz also Warnock (1994, Chaps. 6 and 7), who also differentiates between the biological level of persistence and the essential evaluative and social dimension of a biography constituting identity over time (ibid. p. 125).

views, but also within the complex camp, is nurtured by mixing up the two questions being differentiated here. As opposed to these complex theories which continue to use the concept of person, simple theories can always insist that the first-person aspects of the phenomenon and the properties and capacities specific to persons must be included. They can object at least *prima facie* to the biological approach that it can in no way embrace the identity of the person over time. The first objection by the simple view is applicable, but the consequence to be drawn from it is not to claim a Cartesian perspective analog to the observer perspective. The second objection to the biological approach would only be justified if the latter claimed to answer not only the first question concerning persistence, but also the second one as to the specific way in which persons know their own identity over time and how they behave towards it. But this is not the case, at least in the variant proposed here. The consequence drawn from the justified criticism of some complex theories must be to integrate into the participant perspective the first-person aspects of the identity of a person i.e. her knowledge about her own existence over time, and to segregate the questions of personality and persistence. Then, an analysis of the personality can integrate some of the central elements and some of the intuitions and phenomena which make the simple view plausible.

Between the conception of personality proposed here and the simple theories discussed in the second chapter there is, admittedly, in two respects a decisive difference. For one thing, the first-person perspective is understood as a non-separable moment in the social participant perspective: self-consciousness is an essentially social phenomenon. For another, the cognitive aspect of the first-person perspective i.e. self-consciousness as special knowledge of itself, is not separable from the volitional aspects of self-consciousness i.e. self-consciousness is understood as volitional self-relation. The therewith postulated volitional nature of self-consciousness, which is required as a premise for my further deliberations, is responsible for both the evaluative and the social nature of the first-person perspective. For one thing, a volitional relation to oneself implies the respective states and characteristics attributed in a first-personal way. For another, the evaluations manifested in these self-attributions are constituted through social processes of interpretation and recognition. Thus the participant perspective rather than a purely 'theoretical' Cartesian perspective mimicking the observer perspective is the suitable one for persons to get characteristic knowledge about their own existence over time.<sup>9</sup>

In their criticism of complex theories, the various variants of the simple view have reproved the first-person character of memories and anticipations. But instead

---

<sup>9</sup>The evaluative and social dimension of self-consciousness that is anchored in the volitional nature of self-consciousness is not made the subject of discussion just as much in the Cartesian perspective, which is focused on the epistemological particularities, as in wide areas of the analytical philosophy of the psyche. One reason for this, as Korsgaard (1996, Chap. 13) shows, is the dominance of Hume's model of experience in which the active character of the mind takes a back seat. In contrast, the proactive and evaluative aspects of self-consciousness as an evaluating self-relation have been emphasized in the traditional line reaching from Kant via German Idealism to Heidegger; cf. e.g. Siep (2014, Part IV and 1992, Chap. 4), Sturma (1997, Chap. IV, above all p. 124 ff.) or Tugendhat (1979, 7.-12. lecture).

of emphasizing the evaluative aspects of these phenomena, which – even for the respective subject herself – can be made accessible in a comprehending interpretation i.e. in the participant perspective, the simple view concentrates on the special epistemic aspects contained in the knowledge of the ‘mineness’ (Jemeinigkeit) of the moot first-person episodes. However, it is obvious – above all as regards anticipations – that an evaluative reference of the person to her respective own future is manifested in this phenomenon.<sup>10</sup> The thought experiments based on anticipations, which serve as objections to the cases of under-determination of the identity of a person at a future time, rest on involving the concern about one’s own future i.e. an evaluative self-reference, as an argument for under-determined cases of personal identity being unacceptable.<sup>11</sup> In these scenarios, it becomes clear that an evaluative reference to one’s own existence over time is constitutive for persons. This generates irritation and tension in some cases allowed by the persistence conditions proposed in complex theories.<sup>12</sup> The evaluative self-reference of a person, that is normally aligned to persistence, comes to nothing in these cases because – e.g. in fission scenarios – either it is not clear with which future individual the anticipating person is identical, or – in clear cases of the breakdown of persistence – all the essential aspects for positive evaluative self-reference are present in the anticipated situation. Supporters of the simple view take this result as a general argument for the inadequacy in principle of complex suggestions for solutions, but then concentrate on the special epistemic relations of self-consciousness which putatively show that the identity of the person over time is ‘strict’. In the thereby adopted Cartesian perspective, the evaluative dimension of the self-relation of persons over time is overlooked or suppressed as irrelevant. In contrast, my deliberations show that it is precisely that evaluative aspect of persons’ relation to their own existence over time which makes it possible to describe and solve emerging problems. The biography of a person is constituted in a participant perspective the respective individual adopts

---

<sup>10</sup>That in the case of memories, the evaluation and ‘interpretation’ of past experiences also always emanates from the current self-understanding of the person, is shown e.g. in the studies by Wollheim (1984, Chap. IV) or Hacking (1995). Therefore, the participant perspective of the remembering subject is also always constitutive for this form of knowledge of one’s own existence over time. On the evaluative aspects of personal temporal consciousness cf. also Lloyd (1993).

<sup>11</sup>This kind of thought experiments goes back to Williams, who formulated them as an objection to those complex theories which work with a psychological criterion in order to make a complex theory based on the physical criterion plausible (cf. my detailed description in Quante 1999a). The problem thrown up by these cases is, however, not limited to those theories which work with a psychological criterion, but of a general nature. It concerns the question of how persistence and evaluative self-reference interfere.

<sup>12</sup>The biological approach proposed in this study excludes a large number of these problem cases (e.g. fission in the personal phases of human life), because they are incompatible with the regularities pertinent to the human species. However, in the framework of the biological approach, situations can also be described in which persistence and evaluative understanding of self-conflict with one another (e.g. the state of irreversible coma). In contrast to the fantastic scenarios used in the debate on the identity of person over time, here it concerns genuine cases. But it is important to identify the source of this problem as the conflict between persistence and evaluative self-reference. This point, which is also important for biomedical ethics, is misjudged by the simple view.



in a social environment, in which her own existence over time is ‘told’ as a biography.<sup>13</sup> Even if in this way it is possible with the first-person perspective to integrate a central element of the simple view into the conception of personality outlined here, the decisive difference must be noted, namely that this first-person dimension of personality is not being viewed from the Cartesian but the participant perspective.<sup>14</sup>

### 5.2.3 *The Double Meaning of Identification*

The concept of identification plays a central role in the framework of handling the problem of personal identity over time in which the first-person perspective is relevant: the evaluative trait of anticipations and memories is due to the fact that a person identifies herself with past or future states.<sup>15</sup>

To render this thesis plausible, it is necessary to differentiate between two forms of identification – the descriptive form of “identifying oneself *as*” and the evaluative form of “identifying oneself *with*”.<sup>16</sup> If someone looks at a photo of a school class and after searching says “that’s me”, or if someone recognizes herself as someone who is requested in a search notice to report to the police, she is identifying herself – in a descriptive way – *as* that someone whose juvenile face is recognizable on the photo or who is requested to get in touch. In contrast, if someone just about remembers his or her inappropriate behavior at a party the previous night and is ashamed, she is identifying herself *with* herself just as someone who fears the imminent dentist appointment in anticipation. The identification *with* contains an evaluative

---

<sup>13</sup>Here, the thesis is being argued that personality has a narrative structure which is constituted both for the respective individual and for others in the hermeneutic participant perspective (cf. Jacobs 1984). This narrative conception of personal ‘identity’, which traces back above all to Herder and romanticism (cf. Izenberg 1992), is developed e.g. in the studies by Taylor (1976, 1991, 1992 and 1993, Chap. 1, 2 and 10) and Rorty (1989).

<sup>14</sup>Ricoeur (1996) also suggested that the question of the identity of person in the context of an hermeneutic approach should be treated as a question of personality and differentiated from the question of persistence. The evaluative aspect of the first-person self-reference also appears every so often in representatives of the simple view (e.g. Chisholm 1970a, p. 36 f.). But it no longer plays a role in the treatment of the question of the identity of person over time.

<sup>15</sup>I also consider the more general thesis to be applicable, that every act of self-consciousness i.e. including the self-attribution of a present mental episode, contains an evaluative identification and is volitionally ‘colored’ (cf. the third section of this chapter).

<sup>16</sup>The “self” indicates that “identification” is limited here to the first-person perspective in which a person identifies herself as somebody or with somebody. But since, in my view, the context makes this clear enough, I speak of identification. Speaking of aspects expresses the thesis that the central way for personality to know about an own existence over time and to interact with this existence always contains both ‘forms’ of identification. So it is consistent that there are pure forms of “identification *as*” and “identification *with*” outside of self-identification (in these cases the evaluative dimension is silenced). The identification-free self-reference conceded for present self-attributions (cf. Chap. 2) is thereby limited to “identification *as*”, as Shoemaker’s contributions also suggest.

response to the remembered or anticipated situation which a person attributes as applicable to herself – this response can be positive (affirming) or negative (negating).<sup>17</sup> *Identification with* is always involved in the phenomena of memory and anticipation that are important for personality. It is precisely this form of identification which is expressed in the ability to assume a first person relation to one's own existence, which constitutes the core of personhood and personality.

These two forms of identification have not been clearly differentiated in the debate about the correct analysis of the identity of a person over time.<sup>18</sup> So it is no wonder that in the framework of an analysis of persistence the “future-great-pain-test” brought into the debate by Williams and extensively discussed e.g. by Unger (1990) causes irritation in cases where the persistence question is not settled so that identification as cannot be used. Normally, both forms go together in the case of first-person identification. But in the constructed thought experiments on fission and fusion, the two aspects of identification go asunder. Whilst representatives of the simple view see this as a sign of the principle unsuitability of complex theories, since they do not differentiate between the two aspects of identification, one can, on the basis of this differentiation, grasp the source of the irritation and see what problem is. And the aspect of identification with one's own existence over time that is central to the evaluative nature of personality comes clearly into view.

Various constellations are possible between these two forms of identification. For instance, a person can identify herself in anticipation as the individual existing in a future situation and at the same time identify positively or negatively with this future life situation.<sup>19</sup> In contrast, the cases in which the following happens are problematical: (i) According to the underlying theory of persistence, it is unclear whether the anticipating person should understand herself as the future person or not. But a distinct positive or negative identification with the person whose state is being anticipated comes into play. (ii) It is clear to the anticipating person that the conditions of persistence are not fulfilled, but nevertheless she considers all the aspects relevant for her own existence to have been fulfilled and hence identifies herself *with* this future person although she does not currently identify *as* her.<sup>20</sup> In

---

<sup>17</sup>In the following, one must therefore keep “identification with” free from the resonating everyday language connotation of an exclusively positive rating.

<sup>18</sup>Martin (1987, 1990, 1991 and 1993) takes the credit for also having emphasized the aspect of identification as an evaluative reference of a person to the future and for defining it as an essential feature for the identity of person over time. The more he also emphasizes the aspect of identification in his theory, the more distinctly the causal moment relevant for the identity of person takes a back seat for him (Kolak and Martin 1987). On the one hand, Martin sees thereby that the conflict based on Parfit's question of whether identity even matters is generated by the tension between persistence and personality. On the other hand, Martin tries to answer the question of persistence itself through recourse to evaluative identification (cf. on this the criticism by Hanley (1993) and Martins (1998, p. 85 ff.) partial agreement).

<sup>19</sup>The negative identification with an own future manifested in the wish for assisted suicide or active euthanasia will, just as the question of the binding character of advance directives, be discussed in the next chapters with the help of the distinctions determined here.

<sup>20</sup>Even when the identification is limited to the case of one's own personality in this study, there is no intention here either to dispute that this relation could not also exist to other actual persons, or

the debate about the correct analysis of identity of the person over time it was not, however, the everyday cases that were important, in which a person e.g. identified herself with the activities of her children, parents or role models, but in point of fact, such scenarios in which after fissions or fusions identity is no longer present, but certain forms of ‘ramified’ relatedness still are.<sup>21</sup> It is these problematical constellations of identification as and identification with, within the first-person relation of a person to her own existence over time, which have prompted Parfit’s thesis that identity is not what actually matters – or should matter – in our personal existence. On the basis of the division between persistence and personality, the close connection between personality and autonomy shall now be revealed.

### 5.3 Personality and Autonomy

In its original meaning, the concept of autonomy refers to political structures and denotes their right to give themselves laws. Only later was it transferred from the political to the ethical domain and there it denotes a special capacity of persons, a special quality of will, decisions or actions. In analogy to its use in the political domain, autonomy means the capacity of persons (or their wills) of self-determination, which is manifested in decisions and actions. Since the concept of autonomy is more of a philosophical ‘artifice’ than a permanent feature of everyday language, its meaning cannot be grasped comprehensively through recourse to everyday parlance. On the contrary, every interpretation will contain an element of construction. So the conditions of adequacy of such a construction are, alongside observing certain common sense guidelines, above all the assumed theoretical and possibly also practical objectives which by and large govern the analysis.

So in the following I shall set out from the broadly uncontroversial prior understanding that persons are autonomous if and when they are capable of deciding and acting in self-determined ways. In addition, I shall proceed on the assumption that autonomy is a characteristic feature of persons in the ethical context. But in order for persons to be able to be autonomous, they must possess certain psychological capacities such as that of rational thinking or appropriate evaluation of situations and possible options for action as well as knowledge of their own wants, ideals and (longer term) intentions. In other words, our everyday understanding of personal autonomy imputes an appropriate psychological state. A further important trait both of the everyday understanding of autonomy and of every conception suitable for biomedical ethics is that autonomy deals with a complex capacity which, though it can be present in degrees, is nevertheless as a rule acquired to an adequate degree

---

to claim that this self-identification-with-another-person is irrelevant for ethics. As e.g. Hare frequently emphasizes, our faculty for adopting perspective or imaginative role reversal creates an important basis for ethics (cf. e.g. Hare 1981, p. 96 ff.).

<sup>21</sup> Martin (1998, Chap. 5) shows how in these cases the emotional and cognitive evaluations are ‘superimposed’ on the knowledge about the failure of persistence.

by average developed members of our society in the course of normal socialization.<sup>22</sup> It is this largely uncontroversial prior understanding on which biomedical ethics is based and whose various aspects are deepened, refined or maybe even amended in the framework of philosophical debate. What a person's being autonomous means exactly and what the necessary and sufficient conditions for autonomy are, need not be decided in biomedical contexts as long as both the clear and undisputed cases can be identified in each individual case and the problematical cases are also recognizable as such. As I am not dealing here with a comprehensive theory of autonomy, in this section only the thesis shall be justified that our understanding of autonomy is dependent on the concept of personality taken as the biographical identity of a person.<sup>23</sup> For this the first step (Sect. 5.3.1) will show that our concept of autonomy refers to personality. Subsequently I will discuss in the second step whether, inversely, autonomy is a necessary precondition of having a personality (Sect. 5.3.2).

### ***5.3.1 Personality as Precondition of Autonomy***

Harry G. Frankfurt and Gerald Dworkin have presented a series of papers from the early 1970s on, in which they have developed a hierarchical model for the analysis of freedom of will and autonomy.<sup>24</sup> Particularly Frankfurt's proposal provoked a number of objections, which led to a refinement of Frankfurt's and Dworkin's original theories as well as further development of the hierarchic models in general. As I have described and discussed these objections and developments elsewhere, I will only outline the basic hierarchic model (Sect. 5.3.1.1) and what are in my opinion the decisive extensions by the biographical (Sect. 5.3.1.2) and externalist dimensions (Sect. 5.3.1.3).<sup>25</sup>

---

<sup>22</sup>The everyday concept of autonomy that is appropriate for biomedical ethics is a threshold concept: if the capacities necessary for personal autonomy are present to an adequate extent, then autonomy is recognized in its full sense (Beauchamp and Childress 1994, Chap. 3 and Faden and Beauchamp 1986, p. 235 ff.). The ideal of personal autonomy aiming at perfection and thus unattainable for human beings must be distinguished from this understanding of autonomy (cf. Feinberg (1986, Chap. 18) for an analysis of the various autonomy conceptions, and Quante (2014b, Chapter VII) for a discussion of the same in the context of biomedical ethics). In contrast to Benn, who reserves the term autonomy for the ideal and refers to the realist conception as autarky (Benn 1976), in the following I will use "autonomy" in the less demanding sense, thereby concurring with Faden und Beauchamp.

<sup>23</sup>For an overview of this see Atkins (2008) and the papers included in Atkins and Mackenzie (2008).

<sup>24</sup>Their most important work can be found in Frankfurt (1988 and 1999) and Dworkin (1988).

<sup>25</sup>For further development of Frankfurt's theory cf. Quante (2000c) and (2011b).

### 5.3.1.1 The Hierarchic Model

If one asks whether a person, her will or one of her actions is autonomous, the basic idea of the hierarchic model consists in the suggestion that one can only answer this question if one differentiates between wants of various order.<sup>26</sup> The wish “I would like to donate blood” is a first order wish, because it is directed at an action to be executed by me. In contrast, a second order wish is directed at a first order wish, so in our example: “I would like to have the wish to donate blood.” Generally speaking: one obtains a wish of order  $n+1$  when this wish refers to a wish of order  $n$ . Clearly, people have lots of wants, of which they cannot convert all into action at the same time. So it is sensible to define, as Frankfurt does, the will of a person as the effective wish of a person, i.e. as that wish that initiates an action. Correspondingly, one can then also further differentiate on the level of second order wants between mere second order wants in which the wish is to have a first order wish, and second order volitions in which I not only wish to have a specific first order wish, but also wish that this specific first order wish initiates action i.e. should be my will. When this first order wish becomes my will by virtue of the corresponding second order volition, then I have the will that I wish to have.

Like Frankfurt and Dworkin, I believe that with this hierarchic structure a central aspect of autonomy is being named. Concordantly, Frankfurt and Dworkin describe the relation of second order wants and volitions to first order wants as identification. In contrast to Galen Strawson (1986, p. 243 ff.), who takes identification only to be the theoretical act of identification as, one must in my view include the relation as identification with: when a person refers to her first order wants in this way, she herself enters an evaluative self-relation. In other words: the hierarchic model can count as an illustration of the central capacity of persons to develop a personality and a biographical identity.<sup>27</sup>

If one understands the hierarchic model as a model for the biographical identity of persons i.e. for their evaluative-volitional self-understanding, then one can avoid some of the standard objections raised in particular against Frankfurt. Thus, it is clear from the outset that the autonomy guaranteeing power of *identification with* is not due to an occult quality of the wish, or to the volition of the respective higher order, but to the interplay of the wants of various orders that have been brought together by identification. But also, when through the hierarchical analysis a usable model is voiced for an aspect of the evaluative-volitional self-understanding, further conditions are nevertheless still needed, since in the hitherto existing form, only two isolated wants are correlated. Let us imagine a person who has several such each time internally connected first and second order wants. But in this person there are conflicts both on the level of the first order wants and on the level of second order wants and volitions. Such a person who is furnished with several identifications that

---

<sup>26</sup>In the following I will use “want” in the application of the term in the literature as a generic term for all volitional mental attitudes.

<sup>27</sup>Thus e.g. Gerald Dworkin (1988, p. 15) speaks explicitly of „identify with“ and views this faculty as a characteristic feature of persons which distinguishes them from other living creatures.

are not compatible among one another would be lacking something essential to autonomy: an adequate amount of coherence. Who or what this person wants to be is manifested in this coherent amount of beliefs and wants required above all by the second order; in other words: it concerns the coherence of the respective personality which is attributed to a person at a point in time.<sup>28</sup> For one thing, this extension makes allowance for autonomy being less of a local quality related to individual wants or actions and more of a global quality. For another, in this way an initial explicit reference to personality ensues.<sup>29</sup>

But regarding the question that interests me here, of whether the biographical identity of the person resp. personality is or should be a relevant principle for bio-medical ethics, the objection of excessive demand or inadequacy must still be faced (cf. also Beauchamp's criticism in the fourth section of this chapter). It would lead to a utopian ideal of autonomy that is unattainable for mortals if one demanded that a person is only autonomous when she has identified with each and every one of her wants. In the same way as one has to weaken the condition of coherence, one must also make allowance for the objection of excessive demands. The appropriate strategy consists not in demanding the actual identification, but in a person's general capacity for this evaluative state. The personality of a human being can be manifested in her lifestyle and in her emotional and volitional states, without this person having to place herself factually at a reflexive distance from her wants, basic evaluative states or personality. In such a case, a non-reflexive form of identification with is present, which is expressed in the lifestyle of this person (cf. Sect. 5.3.2 below). If, as demanded by the condition of coherence, no significant disruptions occur, one can speak of an autonomous person if one at the same time awards her the capacity e.g. of carrying out such a critical, positive or negative identification with in a conflict situation. Not the actual execution, but adequate coherence plus the capacity for reflexive identification are therefore required as necessary conditions for the autonomy of persons. Thus I support the thesis of Gerald Dworkin (1988, p. 15) that

it is not the identification or lack of identification that is crucial to being autonomous, but the capacity to raise the question of whether I will identify with or reject the reasons for which I now act.

---

<sup>28</sup>Waddell Ekstrom (1993) sees the true or central self as constituted through the positive identification with a subset of the first rank wants, and Berofsky (1995, p. 10) characterizes this required coherence as „personal integrity“. It is important not to overtax mortals with the condition of coherence. Thus Christman (1993, p. 287) suggested that coherence is adequate for autonomy as long as the subject in question does not perceive any conflicts within his mental states which could significantly impair his capacity to act; cf. also Christman (2009, Chapter 7) and my discussion of his conception in Quante (2013b).. Positively expressed, one can, as Dieter Birnbacher suggested to me in discussion, also demand that a subjective feeling of freedom must be present. The amount of coherence that is thereby necessary for autonomy and personal integrity will not be ascertainable independently of social standards (for these external elements of autonomy cf. 5.3.1.3 below).

<sup>29</sup>Correspondingly, Dworkin (1988, p. 15 f.) states: „Autonomy seems intuitively to be a global rather than local concept. It is a feature that evaluates a whole way of living one's life and can only be assessed over extended portions of a person's life“.

### 5.3.1.2 The Biographical Dimension

This extension, effected by the condition of coherence and the more realist version of the hierarchic analysis geared to a general capacity does not, however, suffice to reveal the internal relation between autonomy and personality. The reason for this is that up to now it has been about a purely synchronic analysis in which only the structure of a person's wants and beliefs present at a particular point in time is inspected. For this reason the hierarchic model outlined hitherto is exposed to a mandatory objection targeting the acquisition of wants and beliefs: Imagine an individual in whom e.g. through perverted socialization or the unnoticed intervention by mighty and malicious brain surgeons, a personality in the form of a coherent set of beliefs, wants and second order volitions has been 'installed'. This thought experiment makes it clear that internal coherence and the capacity for reflexive identification cannot be enough. For it cannot be ruled out a priori that not only first order wants and beliefs, but also second order wants and volitions are the result of manipulative influences which are incompatible with the presence of personal autonomy.<sup>30</sup> This negative result does not arise from thinking in this thought experiment primarily of the socialization products of racist, dictatorial or otherwise ethically unacceptable orientation. Even an otherwise ethically admirable personality such as that of Mother Theresa will not be counted as adequate for personal autonomy if the way of engendering this personality has been manipulative.<sup>31</sup>

In accordance with John Christman<sup>32</sup>, I would therefore like to extend the hierarchic model of personal autonomy by a biographical dimension which can accommodate the problem of the acquisition of beliefs, wants and volitions: A person P is only autonomous in relation to mental attitude M

- (i) when P either becomes aware of the process of the development of the personality of M and does not offer resistance to this development, or when P would not have offered any resistance had she become aware of M's development; and
- (ii) when P's resistance against M's development does not fail (or would not have failed) to materialize because thereby influential factors are (or would be) at work, which make reflexive identification impossible; and
- (iii) when the factual or counterfactual reflexive identification imputed in (i) is (minimally) rational and free of delusion.

(Necessary Conditions for Personal Autonomy)

---

<sup>30</sup>At this point, the relation between the personality and the process of her previous development is therefore central; in the context of the justification of medical paternalism, the opposite line of vision i.e. the shape and form of the development of future wants and beliefs, will become relevant (cf. Chap. 8).

<sup>31</sup>I will shortly be addressing the question of whether an external element in the form of a perception of the good life is thereby being written into the conception of autonomy. (Sect. 5.3.1.3).

<sup>32</sup>The following conditions refer to Christman (1991); for a detailed description cf. Quante (2013b).

This further necessary condition for personal autonomy refers back to the central capacity of persons to develop an evaluative-voluntative self-relation and also respects the cutback of coherence to the minimal i.e. not logically perfect rationality necessary for mortal creatures on the one hand, along with the recourse to the (ability to) execute acts of reflexive identification. By way of the recourse to reflexive identification and the negative condition of the absence of self-deception, the connection is made between the individual mental disposition M and the entire network of these dispositions in which the personality of P is manifested.<sup>33</sup> Through the addition of the biographical dimension as condition of personal autonomy, an explicit relation is established between autonomy and personality, since on the one hand, the capacities specific to personality are indispensable, whilst on the other hand, what has to be counted as preventing autonomy in a particular case will also depend on the respective personality and the other capacities of the respective individual. Certainly, there will also be conditions ascertainable from the observer perspective, which are either necessary for autonomy (e.g. functioning cognitive faculties), or adequate for the absence of autonomy (e.g. massive brain damage). And presumably, some of the general social standards of rationality and coherence conceivable from the participant perspective could be formulated. However, in order to answer the question of the autonomy of a person or of an individual action, the concrete personality of the respective individual must also be taken into account.

Overall, it has thus been shown that our concept of autonomy refers to that of the personality, whereby the developed conditions hitherto are formally phrased to allow the contextual substance of the concept of autonomy to be undertaken to the greatest possible extent through recourse to individual wants and beliefs. However, the last two sections have already made it clear that autonomy cannot be determined without the inclusion of an external dimension. Not only the contextual determination that must count here as perverted socialization, but also the weakened condition of coherence or the evaluation of the counterfactual elements refer to social evaluation standards. It would seem that it is in fact the recourse to the personality with its character of social recognition that leads to autonomy becoming an unattainable ideal.

### 5.3.1.3 The Externalist Dimension

Extending the concept of personal autonomy to include a biographical dimension thus reveals a difficulty that arises out of the tension between the ideal of autonomy as self-determination and the social nature of the human being. Does it not follow from the fact that human persons are shaped decisively by their socialization that no

---

<sup>33</sup>All in all, my strategy for the definition of autonomy rests on the compilation of a list of factors that endanger or debar autonomy. This procedure stems from my conviction that such concepts as freedom or autonomy cannot be analyzed by means of a list of positive conditions; in Quante (2013b) I therefore suggest a default-and-challenge strategy for such an analysis.



personal autonomy can be accorded to them? In other words: Is there not incompatibility between autonomy and socialization?<sup>34</sup>

Christman (1991, p. 23 f.), just like Thalberg or Young, ascribes this inferred incompatibility to traditional liberalist political philosophy which is owed to its 'atomism', i.e. the thesis that all aspects of social and political reality must be traced back to the characteristics and capacities of individuals. Although Christman himself acknowledges that he sympathizes with the communitarian criticism, according to which this liberalist atomism is an illusion, and although he rejects liberalist theories because they "rest on false or illusory presuppositions about human independence" (ibid. p. 23), he nevertheless tries to adhere to a purely internalist approach.<sup>35</sup> On the one hand he acknowledges that his biographical model of personal autonomy points in the same direction as the communitarian criticism of liberalist atomism. But on the other hand, his theory is still 'atomist', since the "linchpin of the definition of autonomy remains the perspective of the agent himself" (ibid. p. 24). I believe that in this sense, an internalist conception of autonomy is based on misunderstandings. In the following it should become clear that an externalist conception of personality, in which the social environment and intersubjective processes of acknowledgment are ascribed a constitutive function, is a suitable base on which the thesis of liberalist incompatibility can be invalidated.

- (a) *The errors in the thesis of liberalist incompatibility*: Feinberg (1973, p. 12 f.) distinguishes four different kinds of constraint of individual freedom which, if the metaphysical problem of free will is ignored, can be understood as constraints of personal autonomy. For one thing, Feinberg differentiates between negative and positive constraints, whereby "negative" denotes the absence and "positive" the presence of such a factor. For another, he differentiates between internal and external constraints, whereby "internal" denotes all factors belonging to the "total self" (ibid. p.13) rather than the central self (Feinberg speaks of the "intimate 'inner core' self"). Accordingly, all factors which do not belong to the respective human person are "external" (Feinberg speaks of "outside a

---

<sup>34</sup>This problem is discussed by Thalberg (1983, Chap. 5) and Young (1980 and 1986, p. 37 ff.). The strength of this conjecture over incompatibility is e.g. clearly noticeable in existentialist analyses of freedom and autonomy such as those of Sartre, which regard any kind of social molding as 'improper'. At the same time, this conjecture about incompatibility is not only a problem in the context of freedom of will, but also, as one can see e.g. in Young, arises in the framework of a compatibility approach in the context of the debate on personal autonomy, which is understood as social or political freedom (cf. on this Kristjánsson 1996 and my discussion in Quante (2007b)).

<sup>35</sup>Here, one must differentiate between two different meanings of "external" and analogously "internal". In one sense there are the causal enabling conditions of autonomy which are overall external to the participant perspective. In this first sense, the differentiation of internal versus external refers to that between the observer and the participant perspectives. In a second sense, one must differentiate within the participant perspective between the internal and the social factors relating to an individual and the social factors which cannot be reduced to this individual's internal factors. The following concerns the differentiation internal versus external in the second sense i.e. within the profound participant perspective.

person's body-cum-mind").<sup>36</sup> If one combines these two differentiations, one arrives at four kinds of constraint.

So the thesis that socialization and personal autonomy are incompatible owes its plausibility either to the assumption that personal autonomy in the sense of self-determination is irreconcilable with the existence of external constraints in principle, or it rests on the assumption that socialization consists only in such external constraints as have to be counted as autonomy jeopardizing constraints. As it became clear earlier, the first assumption bases on an illusory model of individual independence and cannot therefore, as Christman (1991, p. 23) rightly discerned, be made the basis of an appropriate understanding of personal autonomy. But the second assumption is, concurring with Young (1980, p. 575 f.) – without implicit recourse to the first assumption – implausible in respect of socialization, as constitutive structures (e.g. education), limitations of various degree (e.g. social starting conditions) and restrictions are not differentiated.<sup>37</sup> If, on the other hand, one considers the social nature of the human being and her dependence on upbringing and the social community, without which a formation of the capacities necessary for autonomy would be impossible, then the assumption that any external molding of the psychological constitution of a self is irreconcilable with personal autonomy leads directly to the conclusion that human persons cannot be autonomous. Even if one were to accept these considerations as the ideal of autonomy, one would not thereby acquire a useful concept for the analysis of concrete phenomena and the treatment of concrete ethical problems. A realist conception of autonomy will therefore have to differentiate between such external social factors which can be regarded as constitutive structures and tolerable constraints and demarcate those two groups from such social constraints as are irreconcilable with the shaping or possession of personal autonomy. This is backed by the assumption that social and in this sense external factors are not per se irreconcilable with personal autonomy.

As the discussion in Young (1980 and 1986) makes clear, such a realist theory of personal autonomy can use reflexive identification with as a criterion for social moldings and constraints. Socialization processes which facilitate the capacity for critical self-reflection and then, when the subject recognizes (or were to recognize) them and are (or were to be) evaluated positively, can be regarded as promotional for autonomy or at least reconcilable with personal autonomy. Of course, this criterion is not adequate on its own. But in combination with the condition of coherence through which the person is considered in her entirety, most cases of “perverted” socialization can, in my view, be excluded.<sup>38</sup>

---

<sup>36</sup>Feinberg's terminology does not correspond to the terminology I have just introduced (cf. the previous remark).

<sup>37</sup>An illuminating discussion of the error of generally identifying restrictions with coercion can be found in Kristjánsson (1996, Chap. 2). For a broad discussion on the liberalist conception of autonomy in respect of communitarist criticism cf. Benn (1990, Chap. 12).

<sup>38</sup>In my view, a completely satisfactory answer cannot, however, be obtained without the addition of an objectivist-externalist conception of the good life which also incorporates factors beyond individual and intersubjective valuations. But all in all, I think it would be wiser to pursue a formal

- (b) *The external dimension of the personality*: On this base, the prima facie paradoxical reflections by Christman, who on the one hand supports a communitarian criticism of the liberalist thesis of incompatibility while on the other hand wanting to adhere to an atomist and internalist concept of personal autonomy, can be freed of its apparent contradiction in terms. Firstly, one must differentiate between the level of constitutive conditions of socialization which are indispensable for autonomy in an ontological sense and the level of objectives aimed at in the frame of such socialization. An apposite theory of personal autonomy must be externalist in the sense that it acknowledges indispensable aspects of socialization. But this is fully reconcilable with a 'liberalist' ideal, according to which socialization should yield human persons that are capable of critical reflection. On the basis of an externalist conception of personality, nothing hinders the support of a liberalist ideal of personal autonomy.<sup>39</sup> Secondly, on this basis one can even justify the commendation of the respective individual aspects within a theory of personal autonomy. The rejection of certain anti-liberalist interpretations of the concept of autonomy and the ethical preference for (hypothetical) critical reflection do not contradict an externalist conception of personality in which the social disposition of human persons is not seen per se as jeopardizing autonomy.

Decisive for this option is the adherence to the concept of identification, in which a subject behaves judgmentally to both herself and the social influences that mold her.<sup>40</sup> If one does not draw the borderline between internal and external factors as Feinberg does, orienting on the psycho-physical unity of the personified subject, but takes it in an evaluative sense, then those social aspects with which a person identifies herself positively can be understood as "internal", as they become an integral part of her personality. Even such factors as are rejected by a subject will of course aid the constitution of her personality – more precisely: the rejection of and critical debate on these factors will mold the personality. In other words: identification

---

and limited conception of personal autonomy rather than integrating the question regarding the good life into the concept of autonomy.

<sup>39</sup>In Quante (1997 b, p. 69 ff.) I discuss the differentiation between the ontological and the ethical levels in more detail taking the example of Hegel's theory of will.

<sup>40</sup>In contrast to Arneson, I consider the concept of critical reflection to be an indispensable core feature of personal autonomy. As opposed to that Arneson would like to make the contextual substance of the concept of "real self" dependent on which concept of self is most likely to enable a person with a given natural disposition and in a given social environment to lead a good life (Arneson 1994, p. 54): „The best conception of the Real Self component of autonomy is the one such that if people guided their choice of values and preferences and life plans by it, would best enable people to adopt and fulfill the values and preferences that they would affirm in an ideally considered manner“. With this, the evaluative concept of "real self" and therewith indirectly also the concept of autonomy is related to the question of a good life. In contrast to this, I hold on to the model of identification as a core feature of personal autonomy and dissociate the link to the question of the good life. Connected with that is the assumption that in certain social situations or when specific psychological needs are present personal autonomy is not the best way for a human individual to lead a good life.

facilitates the internalization of constitutive external factors in a person's evaluative self-understanding and in this way the integration of these factors into personal autonomy. Of central importance here is that the concepts of personal autonomy and the good life do not coincide: the criteria for personal autonomy proposed here remain largely formal and are therefore inadequate for a definition of the good life.<sup>41</sup> The conception of the 'real self' is not a benchmark for which wants or beliefs are the strongest in a specific human individual, but rather, an evaluative concept oriented on an ideal of personal autonomy. And since this ideal is defined purely formally, there is also no guarantee that a personal life geared to identification will be a good or happy life. For it to be that, more social parameters and an appropriate psychological constitution of the respective individual must be present amongst other things. The objections of Thalberg (1983), Friedman (1986) or even Watson (1987 and 1989) either assume the notion that the 'real self' must contain the factually most powerful wants of a person, or assume that the development of a real self and a good life go hand in hand. In accordance with my submissions, the former is neither a necessary nor a sufficient condition for the latter, even if one can establish that in modern Western societies the concept of a real self that is bound up with the ideal of personal autonomy and which builds on second level volitions normally belongs to the constitutive conditions of a good life. In other words: personal autonomy and leading a personal life belong to the core assets of our culturally most widely accepted values or ideals.

### 5.3.2 *Autonomy as Precondition of Personality?*

If the previous argumentation is correct, then a plausible concept of personal autonomy refers to the conception of personality proposed here. The question is now whether the concept of personality also refers to the concept of autonomy. Do the features of identification with and of the evaluative self-relation that are inscribed into personality imply that only beings with personal autonomy can also have a personality?

---

<sup>41</sup>This would only be the case if one were to see the good life as being defined entirely through the individual wants and beliefs of the respective subject. Such a subjectivist conception appears unsatisfactory to me for reasons which cannot be discussed here. Nor can I deal exhaustively at this point with the question of whether the acceptable conception of the good life implies the ideal of personal autonomy (but cf. the next section). Ultimately, I argue for the combination of a formalist concept of personal autonomy and material value ethics which, in the framework of biomedical ethics, can be implemented through pluralism of principles in the sense of Beauchamp and Childress (1994).

### 5.3.2.1 Borderline Cases of Personal Autonomy

In his review of Berofsky's theory of personal autonomy, Dworkin discusses a borderline case of personal autonomy. Should one continue to regard a person who has autonomously decided to forgo independent decisions in the future course of her life as autonomous (Dworkin 1997, p. 214)? Whereas Berofsky ranks such a person – in the background is a Jehova's witness – as no longer autonomous (Berofsky 1995, p. 127 f.), according to Dworkin, personal autonomy can still be attributed to her. Although Dworkin rightly says that one can only compare different theories of personal autonomy with each other in terms of their problem-solving powers, at this point I do not wish to discuss the questions possibly thrown up by the example, but instead, use the dissent between Berofsky and Dworkin to show the problems of such borderline cases. This will not only further clarify the previous analysis of the relation between personal autonomy and personality, but also allow the relation between autonomy and personality to be defined more precisely.

Berofsky and Dworkin both assume a concept of personal autonomy in which the capacity for critical reflection and the identification with wants and beliefs play a central role. Further, both presume that these capacities are practiced as regards the way in which a person wants to lead her life. In other words: Both integrate a component into their theory of autonomy, which is here embraced by the concept of personality. As regards the controversial borderline case, both also presume that the member of the sect in question is capable of reflecting critically on wants and beliefs. So there are no physiological or psychological stumbling blocks. Before the backdrop of this idealizing assumption, the problem case is now that the person in question decides at some point in her life not to lead her own life, but to allow all important decisions to be made by a "guru" (Dworkin 1997, p. 214). She may give as a reason for this decision that she is only in this way able to lead a good life as she sees it. On the basis of these mutual beliefs, Berofsky comes to the conclusion that e.g. a decision relevant in the context of biomedical ethics made by this person does not merit respect because her wants and beliefs are not "really" her own (Berofsky 1995, p. 129). This counts for at least as long as personal autonomy is seen as a necessary condition of respect (*ibid.* p. 128). In contrast, Dworkin comes to the conclusion that this member of the sect also has the right to be acknowledged as a person whose wants and beliefs are to be respected (Dworkin 1997, p. 215).<sup>42</sup>

Before I can comment on this dissent, various ways in which a person can relinquish her autonomy in an autonomous way must be distinguished. Above all, one must draw the line between local (a.) and global (b.) renunciation.

ad (a.) One can speak of *local* renunciation either when a person in respect of specific content (e.g. claustrophobia, smoking etc.) decides to implement therapeutic means (conditioning etc.) which are incompatible with the practice of critical

---

<sup>42</sup>Both Berofsky and Dworkin are only interested in personal autonomy and don't discuss the question of an autonomy of action and decision respectively, which can possibly be conceived independently of personal autonomy. Since I shall not be making this problem the subject of discussion until the fourth section, I won't pursue the issue in this section.

reflection (a.1). Or, as in the mythical case of Ulysses and the sirens, one can also speak of local renunciation when a person forgoes her freedom of action and decision completely for a limited period and subsequently regains them (a.2).<sup>43</sup> Whilst the first kind of local renunciation can be irreversible (because ‘thematically’ limited), the second must be temporally limited and therefore reversible (because “thematically” unlimited). Both cases are straightforwardly reconcilable with personal autonomy because they are integrated in the personality’s important wants and beliefs, or through the temporal limitation i.e. through the relation to the person’s biography.

ad (b.) Two forms must also be differentiated for *global* renunciation, which contains neither thematic nor temporal limitation. In one case (b.1) the capacity for critical reflection or capacities necessary for the practice of critical reflection is lost (e.g. when a person destroys these capacities by permanent drug consumption). The other case imputed by Berofsky and Dworkin in the above example is characterized by the person in question still possessing the capacity for critical reflection but due to an autonomous decision no longer wishing to use it (b.2).<sup>44</sup> Whereas the former case is clearly irreconcilable with the possession of personal autonomy, the latter is not clear cut. In the case of a still existent and merely not called on capacity for critical reflection (b.2.1) the viewpoint could more easily be argued that it is a borderline case of personal autonomy than in the case of a capacity which can only be regained with difficulty (b.2.2). In both cases, one would have to tie the attribution of personal autonomy to a counterfactual statement or an apposite disposition of the person in question. In the former case (b.2.1) one could thereby draw on the counterfactual updating of her actual capacity as a basis (“if she wanted, she could reflect critically”); in the latter case (b.2.2) one would have to refer to the remaining capacities to regain the capacities necessary for critical reflection “if she wanted, she could regain the capacities necessary for critical reflection”).<sup>45</sup>

---

<sup>43</sup>This will become important in the context of so-called „Ulysses contracts“ (cf. Chap. 7).

<sup>44</sup>As almost every other example, the above one is also underdetermined. For one thing, it is not clear whether the capacity for critical reflection itself or that for the performance of the same necessary functions are still present. For another, it is not made clear whether these capabilities must actually be present, or whether it must at least be possible for them to be redeveloped. Bearing in mind above all psychological factors, it is clear that it will most likely be about a continuum of cases. The longer lasting renunciation of the practice of critical reflection will almost always have the effect that a subject must first rehearse the respective capabilities. However, this complication can be disregarded for my discussion above.

<sup>45</sup>The lack of differentiation between the two cases seems to me to be a reason for the dissent between Berofsky and Dworkin. These differences are systematically important, not only because conflicting intuitions can be clarified better therewith, but also because the latter case can, analogue to itself, be understood as developing (socialization) and dwindling (loss of competence) personal autonomy. So, just as in socialization the capacity for autonomy can accrue from non-autonomous influences, it can also be destroyed or suspended on the basis of autonomous actions. Here, the different treatment due to the biographical structure of the individual cases is also important: whereas e.g. a young person in the process of socialization, who every so often exhibits personal autonomy is still protected and not constantly regarded as fully autonomous, the tables are turned

In respect of the difference between local and global renunciation, two things can be put on record. Firstly, this differentiation does not simply coincide with that of reversible and irreversible processes. Whereas reversibility is necessary for (a.2) and (b.2.2) if the assumption of reconcilability between renunciation and personality is to be at all plausible, this is not the case for (a.1) and (b.2.1); in contrast, case (b.1) is in no acceptable way reconcilable with personal autonomy. The missing reversibility is compensated in one case (a.1) by integration in the coherent set of wants and beliefs of the personality and in the other case (b.2.1) through the – presumed permanently available – capacity, which is no longer updated because of a one-time autonomous decision.<sup>46</sup> Secondly, the main difference is based on local renunciation having the overall aim of strengthening, reestablishing or securing the personal autonomy of the person in question. In contrast, global renunciation cannot be tied to this aim, but is ultimately geared towards the conception of a good life accepted by the respective person.

To return to the dissent between Berofsky and Dworkin, if one asks oneself with whom one should agree, one must on the one hand keep in mind that the under-determination of the example could be a reason for the dissent. But on the other, it is crucial to differentiate clearly between the two interwoven questions:

- (Q 1) Should one grant personal autonomy to a person who autonomously decides to forgo critical reflection?  
 (Q 2) Should one respect the wants and beliefs of such a person?

(Two Questions)

The dissent between Berofsky and Dworkin over the second question is factually independent of their dissent over the first question. Berofsky himself says that the negative answer to Q 2 only follows from Q 1 when one regards personal autonomy as a necessary condition for respect. Berofsky and Dworkin seem to share this assumption. Consequently, Dworkin construes the affirmation of Q 1 out of his positive answer to Q 2, whilst Berofsky construes the negative answer regarding Q 2 from the negative answer to Q 1. However, if – in contrast to Berofsky and Dworkin – one does not regard autonomy as a necessary condition for the respect of wants and beliefs, then one can say with Dworkin that the wants and beliefs of our exemplary person are worthy of respect because of their coherent interrelation (Dworkin 1997, p. 215). And one can hold the same view as Berofsky, that such a person should not be attributed personal autonomy. This seems to me – in particular for (b.2.2) – to be the more plausible solution, as otherwise one must either argue a kind of transfer principle of personal autonomy, according to which personal

---

in the case of ‘autonomy flaring up’ in temporarily no longer autonomous persons (e.g. in long-term care). As long as these phases of personal autonomy occur, incapacitation should not be permitted: Autonomy quasi ‘reverberates’. The above distinctions also make it possible for a patient e.g. to relinquish her autonomy in certain areas she can maybe no longer handle, while retaining her overall personal autonomy.

<sup>46</sup>The problem of having to postulate the existence of a capacity that has never been updated does not arise here, as our cases presume an autonomous decision to relinquish.

autonomy is handed down to all wants and beliefs which in their turn trace back to an autonomous decision.<sup>47</sup> Or one must rely on counterfactual statements which herald the development of certain capacities. This seems to me to be the less attractive strategy, since a plausible demarcation is hardly likely to be possible.<sup>48</sup>

If, as I would now like to suggest, one defines the concept of personality and the element of identification contained therein in such a way that personal autonomy is not a necessary element for the possession of an evaluative self-image, the source of our respect for the wants and beliefs of a person who forgoes critical reflection, hinted at by Dworkin, can be reconstructed as respect for the *integrity* of that person, as respect for her personality.

### 5.3.2.2 Autonomy, Authenticity and Personality

First of all, two forms of authenticity must be differentiated (cf. Taylor 1991). In one sense, authenticity means that a person has verified and identified herself with all her wants and beliefs on the basis of critical reflection. One could also in this regard speak of the personality becoming reflexive. In this meaning, authenticity and the ideal of personal autonomy coincide. A person is authentic if and only if she has made all her wants and beliefs 'her own' through identification. This ideal demands that there are no elements in the personality of a person which are not the result of a process of identification. A single conviction or a single valuation is authentic when it has survived the critical verification of the person and has been made 'her own' through a positive evaluation by the person. At this point one could also speak of a reflexive competence for autonomy of agency. In the following, I shall call this form of authenticity *reflexive authenticity*. One speaks of a different kind of authenticity when e.g. one says of a musician that he plays blues 'authentically'. His 'authentic' does not mean critical reflexive identification but an unbroken relationship. I will call this form of authenticity '*prereflexive*'. Since the element of critical reflection is missing in those beliefs and valuations which are classified as prereflexive authentic, they are not attributed the attributes of personal autonomy, but rather, they can be the result of processes of socialization or even the expression of psychological needs of the respective person. Reflexive and prereflexive authenticity share the feature of being constitutive for an evaluative relationship to self of the respective person i.e. constitutive for the personality.

In both forms of evaluative self-relation, the 'identification with' that is necessary for the possession of a personality comes into effect. In the prereflexive

---

<sup>47</sup>Cf. Fischer (1995, Chap. 2 and 3), Fischer and Ravizza (1998, p. 163 ff.) and my discussion in Quante (2007b) on the transfer principle that plays a central role in the debates concerning freedom of will.

<sup>48</sup>When the chief motive, namely the assumption that personal autonomy is necessary for respect, is left out and the fact is acknowledged that the concept of autonomy is not determined by its meaning in everyday speech, but must always be defined in the framework of a theory, then the decision is no longer serious. It is important to observe that this waiving of the counterfactual strategy is not incompatible with the analysis of the identification constitutive for personality as a capacity.



authentic personality it is manifested in that the wants and beliefs of the respective person are aligned to this prereflexive authentic basis and overall coherence is thus generated. The evaluative self-relation of the person, which can also be manifested in basal forms of an emotional or affective kind, is factually directive in the actions and valuations. One can thus say that the respective person will, to the degree she – possibly unconsciously – makes her prereflexive authentic ‘core’ the basis of her lifestyle, identify herself with this ‘core’ via her lifestyle. This is also the reason why the ‘identification with’ does not generally imply a reflexive distance from one’s own personality. In contrast, the required ‘identification with’ does have an effect in a reflexive sense in reflexive authenticity. Those evaluations and beliefs which direct the lifestyle do themselves become the object of critical reflection and thus of identification.

To possess personality, it is necessary to have an evaluative self-image and with it at least one identification with that is manifested in the lifestyle. To possess personal autonomy, this evaluative self-image constituting personality, or the evaluations and beliefs constituting it must moreover themselves be the subject of critical reflection and thus become the subjects of identification. This means that prereflexive authenticity suffices for the possession of personality and thus personality does not require the presence of personal autonomy. A personality can constitute itself through an evaluative self-image which is manifested in the lifestyle as a coherence generating and agency directing core of prereflexive authentic wants and beliefs, without becoming thematic of the respective person. Such an individual is then e.g. firmly and continuously anchored in a traditional *Lebensform* and aligns his actions, wants and beliefs to this evaluative self-image without questioning it.<sup>49</sup>

Thus, just as a factually<sup>50</sup> unattainable ideal is formulated with the ideal of complete personal autonomy on the side of reflexive authenticity, since nobody can ever subject all her wants and beliefs to critical reflection, on the other side, prereflexive authenticity can also embrace the borderline case of a ‘happy’ person who never has to evaluate singular actions, wants or beliefs critically in relation to her evaluative self-image. A continuum of cases distinguished by an increasing amount of critical reflection stretches between these two poles. It is generally impossible to establish from when on a person draws on personal autonomy and this surely also depends on social norms. But, as the discussion about personal autonomy in this chapter has

---

<sup>49</sup>The Jehovah’s Witness that Berofsky and Dworkin made a subject of discussion would then be an example of this, if it had been brought about through upbringing rather than an autonomous decision to become a member. Since it can at any rate evaluate individual actions with reference to the evaluative self-image of a Jehovah’s Witness that for them doesn’t come into it, autonomy of agency can by all means be attributed to her. But this only earns respect because it is an expression of the integrity of this person (cf. Sect. 5.4 on this).

<sup>50</sup>If one were also to demand for reflexive authenticity that a person in critical reflection of a want may only draw on such wants and beliefs as have themselves already been the subject of critical reflection, then the ideal of reflexive authenticity and perfect personal autonomy seems to become inconsistent (for a conception that approaches this, cf. Sartre’s theory of freedom in *Being and nothingness*). Here the only demand for the ideal of personal autonomy is that a person – step by step – subjects all her wants and beliefs to critical reflection.

shown, it is important to differentiate in the personality with personal autonomy between the unattainable ideal and the realist conception. The latter demands – and this is how the personality without personal autonomy differs – that the person in question knows that in principle she can and must subject every part of her personality to critical reflection when the circumstances necessitate it. With this the personality of the person possessing personal autonomy becomes increasingly thematic.

It is this knowledge and having the capacities necessary for critical reflection at her disposal that allow personal autonomy to be attributed to a person and not the ongoing critical reflection itself. Thereby, it is the identification with specific wants or beliefs as manifested in agency which must be added to the identification being cultivated in critical reflection, which allows the dispositional analysis proposed by Dworkin (1988, p. 15 ff.) and Christman (1991) to become plausible. Permanent critical reflection and identification is not required for personal autonomy because identification with an evaluative self-image is being manifested in the lifestyle. If the capacity for critical reflection is present, this self-image is sufficient for personal autonomy. But this critical reflection need not be directed at the ‘core’ of personality as long as the person in question is happy with the way she leads her life in the framework of processes of acknowledgment.<sup>51</sup>

My deliberations on the relation between personality and personality are here-with concluded. If the submissions in this chapter were successful, then they have yielded a material content of the concept of personality and personal autonomy proposed here. But the evidence of the formal link between personal autonomy and personality proffered in this chapter does not yet suffice for my central question as to whether personal identity is a principle of biomedical ethics. Rather, in the following section clarity must be gained as to whether the principle of respect for autonomy refers to personal autonomy and thereby implicitly to the concept of personality, or whether it is possible to understand the principle of respect for autonomy exclusively as respect for the autonomy of agency and decision.<sup>52</sup>

## 5.4 The Principle of Respect for Autonomy

In biomedical ethics, the principle of respect for autonomy has entered primarily in the form of the recognition of informed consent (or refusal). Such decisions are individual acts of patients, which explains why the focus on autonomy of agency

---

<sup>51</sup> Just as the „wholeheartedness“, which Frankfurt later (1988, Chap. 12) added to his hierarchical model refers to the criterion of coherence, his subsequently added criterion of „satisfaction with one’s self“ (Frankfurt 1999, p. 104) can be understood as the recognition of a non-reflexive identification being manifested in the execution of life; on the more recent development of Frankfurt’s theory cf. Quante (2000c and 2011b).

<sup>52</sup> In the following, the autonomy of action, choice and decision will be subsumed under the concept of the autonomy of agency and differentiated from personal autonomy. Whilst the form refers to concrete individual occurrences in the life of a person, the latter refers to a specific quality in the biography of a person by virtue of which her life can be understood as define and ‘led’ by her.

seems to suggest itself for biomedical ethics. But in addition, it has been suggested, above all by Tom L. Beauchamp, that the principle of respect for autonomy in the context of biomedical ethics should refer in general not to personal autonomy but to autonomy of agency.<sup>53</sup> For this reason, an altercation with Beauchamp's objections and arguments must follow here. After portraying his position (Sect. 5.4.1) I will try to prove that autonomy of agency – with the exception of the borderline case of prereflexive authenticity – cannot exist independently of personal autonomy (Sect. 5.4.2).<sup>54</sup> The conclusion concerns proof that a principle of respect for autonomy that is of use to biomedical ethics cannot manage without recourse to autonomy of agency, but must also refer to personal autonomy or the integrity of the personality (Sect. 5.4.3). The evidence aspired to in this part of the study has hence been supplied, that personal identity qua personality is an essential element of both autonomy and the principle of respect for autonomy and thus a basic principle of biomedical ethics as a whole.<sup>55</sup>

### 5.4.1 *Beauchamp's Theory of Autonomy of Agency*

Beauchamp's thesis is that not personal autonomy but autonomy of agency should be drawn on as the basis for the principle of respect for autonomy.<sup>56</sup> His argumentation proceeds on two levels. On the one hand, he is concerned with the connection between autonomy of agency and personal autonomy, and on the other, with the question of which of the two is and should be conceived as being the target of the principle of respect for autonomy in biomedical ethics. In addition, he develops a conception of autonomy of agency which can serve as the basis for the principle of respect for autonomy. In the following, Beauchamp's thesis, which is pertinent for my question, shall be discussed in more detail (Sect. 5.4.1.1). After that, the concept

---

<sup>53</sup> Beauchamp developed this thesis in two different texts, which were composed jointly with different authors (Faden and Beauchamp 1986, Chap. 8 and Beauchamp and Childress 1994, Chap. 3). If in the following I refer only to him, this is to simplify things stylistically. Neither the contribution by Ruth R. Faden nor that of James F. Childress are being depreciated or ignored thereby.

<sup>54</sup> I will discuss this special case later (cf. Sect. 5.4.3).

<sup>55</sup> Please note that in this chapter there is no intention to either defend the principle of respect for autonomy or render a concrete contextual definition of this principle or its scope and weight as opposed to other principles in biomedical ethics.

<sup>56</sup> My interpretation that Beauchamp is pleading for autonomy of agency instead of personal autonomy as the basis of the principle of respect for autonomy maybe even marks his position coming to a head. But, as will be shown presently, Beauchamp's argumentation does at least suggest this coming to a head even if, as far as I can see, he does not phrase it explicitly (cf. 5.4.1.1). The fact that he develops a model of autonomy of agency that should be adequate for the treatment of problems in biomedical ethics supports the above interpretation in my view. The observation, that his theory does not imply that only autonomous actions are worthy of respect, is, however, not an objection, since here I am only concerned with the question of what the principle of respect for autonomy refers to. And with reference to this principle Beauchamp considers only autonomous actions to be „worthy of respect“ (Faden and Beauchamp 1986, p. 265).

of autonomy of agency, which according to Beauchamp delivers an adequate basis for biomedical ethics and the principle of respect for autonomy, shall be presented (Sect. 5.4.1.2).

#### 5.4.1.1 From Personal Autonomy to Autonomy of Agency

Beauchamp's thesis that it is autonomy of agency rather than personal autonomy which should be drawn on as a basis for the principle of respect for autonomy can be understood as the conclusion ensuing from three premises which in turn he tries to make plausible. His argumentation can be presented as follows:

- (P 1) Autonomy of agency is independent of personal autonomy.
- (P 2) Personal autonomy is neither necessary nor sufficient for the implementation of the principle of respect for autonomy in biomedical ethics. The respect for the autonomy of the patient, which is reflected in the concept of informed consent relates only to autonomy of agency.
- (P 3) The attempt to draw on personal autonomy as a basis for the principle of respect for autonomy leads to ethically unacceptable consequences in the context of biomedical ethics.

---

(C) Autonomy of agency is the applicable basis for the principle of respect for autonomy in the context of biomedical ethics.

(Beauchamp's Argument)

Provided that the three premises can be justified plausibly, conclusion (C) is plausible: If personal autonomy is not required for the implementation of a principle of respect for autonomy either in the logical or material sense, but in fact to the contrary, its implementation is harmful from the ethical aspect, then it is obvious that the principle of respect for autonomy should rest solely on the basis of autonomy of agency.<sup>57</sup> In his discussion on the connection between personal autonomy, autonomy of agency and the principle of respect for autonomy, Beauchamp expresses and justifies his premises as follows:

*ad (P 1)*: After developing his own theory of autonomous actions, Beauchamp poses the question of whether a hierarchical model of personal autonomy is needed for this approach. His answer is that it is not (Faden and Beauchamp 1986, p. 264). To be sure, he admits that it is impossible to conceptualize a comprehensive theory of autonomous actions which e.g. can deal with problems of multiple personality or manic depression, without developing a "satisfactory theory of the

---

<sup>57</sup>Even under the assumption that only a concept of autonomy can be a basis for the principle of respect for autonomy, this conclusion is admittedly not mandatory, as there could be a third option alongside personal autonomy and autonomy of agency. Such a view is not held, so that I can restrict myself here to the above argument.

self' that is "capable of distinguishing alien forces on the self from the core self or 'real' self" (ibid. p. 269). He nevertheless maintains that the model of autonomy of agency he has conceived suffices for a reconstruction of the concept of informed consent. But this implies that it is logically independent of a concept of personal autonomy in which a satisfactory theory of the self is developed. Alongside this constraint, which does not invalidate (P 1), Beauchamp adds as a further weakening qualification that at least the conception of personal autonomy, in the form he made the object of analysis and criticism, cannot be a necessary component for an analysis of autonomous actions (ibid. p. 264). But even this weakening leaves (P 1) untouched, as Beauchamp subsequently also tries to show for modified concepts of personal autonomy that they are superfluous to his theory of autonomy of agency (ibid. p. 266 ff.).<sup>58</sup> In all, he claims that his conception of autonomy of agency, which rests on the three pillars of "intentionality, understanding, and noncontrol" (ibid. p. 269) is independent of personal autonomy.

*ad (P 2)*: Beauchamp estimates the consideration of "how well the analysis would function in the moral life, where it will inescapably be connected with the principle of *respect* for autonomy" as one, maybe even "the most important test of the adequacy of an analysis of autonomy" (ibid. p. 265). This test serves him as evidence for the thesis that personal autonomy is neither necessary nor sufficient for the implementation of this principle.<sup>59</sup> Beauchamp's first objection against personal autonomy is that this is insufficient for the emergence of informed consent. Since personal autonomy is aimed at the general capacity of an autonomous lifestyle, but not at the individual practice, it is *not sufficient* as the basis of informed consent. Informed consent is only worthy of respect because it is the actual practice of these capacities. If e.g. someone who is generally in a position to give informed consent does not speak out in a concrete situation, or consents in an erroneous way, then the conditions for respect are not present. Moreover, according to Beauchamp's second objection, due to the demanding conditions of personal autonomy (e.g. reflexive identification), many actions we would of course normally consider autonomous do not qualify as such. In principle, according to this model every conscious decision could if anything be judged not

---

<sup>58</sup> Thus a further restricting condition for the merit of Beauchamp's argument arises. It quite obviously imputes specific conceptions of personal autonomy. His concern is not general proof that personal autonomy as such i.e. without further qualification is an unsuitable candidate. What form should such proof even take? Thus, in principle the possibility of developing a theory of personal autonomy remains open, though Beauchamp does not have that in mind. But since he takes into account the main features of the model of personal autonomy that I have unfolded in the last two chapters, his objections are directed at the conception proposed here and must be cleared up if my thesis, that personal autonomy in the sense developed here is the base of the principle of respect for autonomy, is to hold its own.

<sup>59</sup> Although Beauchamp does not always separate the two levels distinctly, it is a good idea to apply this test descriptively i.e. to ask if the proposed analysis meets our moral practice. The question of whether it is ethically desirable to bring in the debatable principle will be raised by the third premise.

to be autonomous, if the agent had neglected to make his wants and volitions his own critically beforehand (ibid. p. 264).<sup>60</sup> But since one acknowledges these actions as factually autonomous, the condition described through the concept of personal autonomy can be *unnecessary*. Conversely, it is, as the first objection shows, just so, that only by way of an appropriate conception of autonomy of agency can respect for informed consent become intelligible. But autonomy of agency is, as this consideration shows, not only a *necessary condition* for the implementation of the principle of respect for autonomy. As the treatment of patients in psychiatric institutions who cannot mold their general lifestyle according to the yardstick of personal autonomy, autonomy of agency is also a *sufficient* condition (Beauchamp and Childress 1994, p. 121). For on that view it is the case that we do indeed respect these patients' individual autonomous decisions, e.g. their choice of clothing or meals and their refusal of medication.

*ad (P 3)*: According to Beauchamp, it is not that the attempt to establish the principle of respect for autonomy on a concept of personal autonomy describes our practice factually erroneously. In a third step in the argumentation the implementation of such a concept is said to lead to unacceptable consequences from the ethical point of view. The theory of personal autonomy as the basis of a principle of respect for autonomy that is feasible for biomedical ethics is not acceptable because it is oriented on an ideal of autonomy that overtaxes the capacities of normal deciding and acting persons (Beauchamp and Childress 1994, p. 124).<sup>61</sup> With this a critical yardstick is being applied which can lead to extensive pluralism, since rational and competently performed autonomous actions are deliberately being ignored in the name of this ideal which they cannot satisfy. In all there is a risk of degrading the ethical relevance of such actions in our "moral life" through the implementation of such a utopic yardstick, since they would be deprived of autonomy (Faden and Beauchamp 1986, p. 265).

#### 5.4.1.2 Beauchamp's Model of Autonomy of Agency

Beauchamp analyzes autonomous actions as actions of normal or average decision makers which act firstly "intentionally", secondly "with understanding" and thirdly "without controlling influences that determine their action" (Beauchamp and Childress 1994, p. 123). Whereas the definition of intentionality cannot be gradualized, the second and third parameters can be fulfilled to a varying degree. Thus actions can be autonomous to various degrees, whereas the degree of their

---

<sup>60</sup>In principle, Beauchamp adheres to this objection even for a modified conception of personal autonomy such as the one I propose, when he writes that e.g. actions knowingly decided on by a person, but executed reluctantly, cannot – inacceptably – be recognized as autonomous (ibid. p. 267).

<sup>61</sup>Elsewhere, he concedes that the model of personal autonomy he criticizes could possibly be suitable for an analysis of morally autonomous persons, but even there he upholds his objection that this may be too demanding (Faden and Beauchamp 1986, p. 273 fn. 37).

autonomy can be understood as the function of the degree of fulfillment of the second and third definitions. Elsewhere he (Faden and Beauchamp 1986, p. 241–262) elaborates further on these three conditions, while, in accordance with his line of questioning, keeping an eye on the action theoretic and bioethical aspects in equal measure (the account of his concept of autonomy of agency is restricted to the theoretical dimension in the following).

- (a) *Intentionality*: The conception of an intentional action, supported by Beauchamp, is summarized in the following definition: “An intentional action is an action willed in accordance with a plan” (ibid. p. 245). Three important definitions are contained therein. *Firstly*, following Davidson, Beauchamp differentiates between the intention, the formation of an intention and the intended action. Only when an action is performed that corresponds to a formed intention and is evoked by it in the right way, which the agent takes to be the attempt to put his intention into action, only then is it an intentional action. (ibid. p. 242) This condition precludes intentional actions being events happening coincidentally or accidentally which befall the actors. The physical initiation by third persons is also excluded: The agent must have the true belief that he has caused the action.<sup>62</sup> *Secondly*, the intentional action is defined as an action executed in compliance with a plan. This means that the agent has knowledge in advance or accompanying the performance that the individual steps are realizations of his intention. With this definition, that of intentionality moves close to the second definition of autonomy of agency, as intentionality demands an elaborate understanding of the individual steps and interconnections (ibid. p. 243). For an action to count as intentional, its performance must “correspond” to the agent’s plan, (ibid. p. 242) even if the aspired end is not achieved. Without a “conception how X is to be performed” (ibid. p. 243), an action cannot be intentional. *Thirdly*, Beauchamp argues a broad conception of “intentional” in the sense that not only can “intrinsically” or “instrumentally” aspired events or aspects be included, but also so-called tolerated events or aspects (ibid. p. 244). Such part actions or aspects required for performing an action are named as the latter, to which the agent does not actually aspire, but maybe even just tolerates. So Beauchamp’s conception of intentionality is not limited to the intrinsically or instrumentally positively evaluated, but allows all parts or aspects of an action to be counted as intentional if they have been grasped by the agent in a cognitive sense. Consequently, at the same place he defines the intentional action as an “action willed in accordance with a plan even though it may not be wanted.” (ibid. p. 245) A part action can thus, even if the agent is antipathetic, count as intentional if it is a known part of a plan of action.<sup>63</sup> Finally, Beauchamp dis-

---

<sup>62</sup> Here, Beauchamp not only masks the relation of intentionality and autonomy to the question of responsibility (ibid. p. 269 fn. 5), but also ignores alternative definitions related to the designation “intentional”.

<sup>63</sup> The question of whether they must at least be interpreted as a means for (or part of) a positive volitional overall result is not dealt with by Beauchamp at this point (cf. 5.4.3 for the relevance of this problem).

cusses the question of whether intentionality can be a question of degree and comes to the conclusion that this criterion of the autonomy of agency does not allow for any gradualization. The semblance of graduality comes about for one thing, because an agent can decide more or less explicitly and knowingly for the execution of an action by e.g. weighing up possible alternative actions. For another, the concept of gradualized intentionality can also be generated in that the plan involved can be cognitively more or less elaborated. But these two aspects do not apply to the intentionality of an action but to the degree of deliberateness. The latter can be defined in degrees of deliberation, awareness and explicitness, but intentionality itself cannot (ibid. p. 248).

- (b) *Understanding*: With “understanding how”, “understanding what” and “understanding that” (ibid. p. 250) Beauchamp differentiates three kinds of understanding. Whilst the first, as practical competence, is left out by him as a component of autonomy of agency, the “understanding that” constitutes an essential criterion for autonomy of agency. The third form of “understanding what” is, in contrast, a special form which is primarily relevant in the area of communicative action. It measures whether an addressee has adequately understood the content of a communication attempt.<sup>64</sup> In relation to the analysis of autonomy of agency the central question now is under what conditions a person understands “the nature and implications of his or her actions.” (ibid.) Here, Beauchamp develops the conception of complete understanding and remarks that an action is less autonomous to the degree that it represents a case of less complete understanding (ibid. p. 252). Full or complete understanding of her own action is given when a person has correctly grasped all propositions or statements which describe, firstly, the essence of her action or secondly, the unforeseeable consequences and possible results that arise through the performance instead of the omission of her action (ibid. p. 251). Since this definition alludes to the predictable consequences, whilst no omniscience is implied here, nevertheless, complete understanding is certainly an inaccessible ideal for humans. So the objection can be raised against the proposed definition that *for one thing* it is too strong. For this reason, Beauchamp ties his conception of “complete understanding” to an “evidentiary standard” i.e. a standard of justified beliefs, rather than to the “justified true-and-full belief standard”. Thus the standard of the justification of a belief attainable for us is inscribed into complete understanding. Of course, ignorance and false beliefs can constrain or even preclude autonomy of agency. But positively one can only demand that the beliefs put forward satisfy generally recognized social standards (ibid. p. 254). *For another thing*, one can object to Beauchamp’s definition of complete understanding on the grounds that the claim for completeness is factually not even stipulated. Beauchamp also accepts this objection and duly modifies his defini-

---

<sup>64</sup>This form of understanding is above all important for an analysis of informed consent. In all, it can be said that Beauchamp is indeed very closely oriented to the concept of informed consent as regards the debate on the criterion of “understanding”. Merely the exclusion of understanding as practical competence is not plausible beyond this tight framework (cf. below, 5.4.2).



tion to the effect that now an adequate understanding of all those statements and propositions which are “relevant” is required (ibid. p. 252). The basis of autonomy of agency is thus the capacity of being able to differentiate between relevant and irrelevant aspects and consequences of one’s own action. According to Beauchamp, in connection with the ability to adequately understand and to check the relevance of information communicated in a clarifying colloquy, the second condition necessary for autonomous actions and decisions, which forms a basis for respect for autonomy, is thus being formulated.<sup>65</sup>

- (c) *Absence of external control*: Beauchamp understands the condition of „noncontrol“ as “independence from control” by others (ibid. p. 256) and, since it belongs to the core assets of our concept of autonomy, he regards it as the third condition necessary for autonomy of agency. He equates this condition with that of the absence of “external controls on the action” (ibid.).<sup>66</sup> In a *first* step, in a more detailed explanation of this third condition for autonomy of agency Beauchamp differentiates between a completely noncontrolled and a completely controlled action. Completely noncontrolled actions “have either (a.) not been the target of an influence attempt, or (b.) if they have been the target of an attempt to influence, it was either (b.i) not successful or (b.ii) it did not deprive the actor in any way of willing of what he or she wishes to do or to believe“ (ibid. p. 258 – insertions in brackets by MQ). This definition permits – due to (b.ii) – noncontrolled exertion of influence. An action is completely controlled when it is dominated solely by the will of another person and the aims and wants of the controlled person are not taken into consideration in any way (ibid.).<sup>67</sup> As Beauchamp continues, this control is manifested by the controlled person being induced to execute an action for which she herself possibly has no predilection. This illustrates how he wants the definitions of ‘resist’ and ‘irre-

---

<sup>65</sup> Since Beauchamp’s deliberations on autonomy of agency are above all tailored to the concept of informed consent, at this point he once again discusses the moment of “understanding” under the key words “accurate interpretation” and “effective communication” (ibid. p. 255). He hereby explicates that on this level the concept of persons’ mutual understanding, as it has been investigated and developed above all in German philosophy, is in its main features a role model for the “akkurate Interpretation”. Thus, at this point a hermeneutic moment of understanding persons enters into Beauchamp’s conception.

<sup>66</sup> With his conception, Beauchamp wants to avoid defining autonomy of agency via the indistinct and variously applied definition of voluntariness (ibid. p. 257). It is important to note that in his conception, control is restricted to third party actions. Thus, the classical question concerning the problem of voluntariness, that of whether strict causal laws i.e. determinism, are incompatible with freedom resp. autonomy, are blended out at the outset (cf. on this more comprehensive usage of the concept of control, viz Quante (2007b)).

<sup>67</sup> Firstly, this concept permits the aims of the controlling and controlled persons to coincide factually, but it excludes any intention by the controlling person to fulfill the targets of the controlled person. Secondly, paternalistic actions are thus by definition not completely controlled exertions of influence, since the well-being of the controlled person is intended by the controlling person. Beauchamp’s formulation does not demand that the controlled person explicitly name her targets, so that if one may impute her interest in her own well-being, a paternalistic action cannot be completely controlling.

sistable' (ibid.), through which he explains his concept of control, to be understood. In the *second* step he elaborates on the other definition, through which he more closely defines the condition of absence of external control, by differentiating between three kinds of "influence": "coercion", "manipulation" and "persuasion". Coercion is a form of complete control (ibid. p. 259) and bases on the credible threat of unwanted and avoidable harm (ibid. p. 261). In contrast, belief is restricted to exertion of influence by appealing to reason (ibid. p. 261), with the aim of voluntary consent to the beliefs or instructions for action suggested by the person influencing.<sup>68</sup> By manipulation it is referred to all "intentional and successful influence of a person" that, without coercion, either alters the actual options of decision of another person, or else, without appealing to that ..., alters the perception the person influenced has of these decisions. (ibid.). Beauchamp names deception, seduction and indoctrination as possible forms of manipulation. While belief is always compatible and coercion generally incompatible with autonomy of agency, manipulation is a form of influence which must not per se be incompatible with autonomy of agency. Beauchamp adds three explanations to his argument. *First*, one should interpret the exertion of influence according to a subjective standard (ibid. p. 260). This means that bearing the respective influenced person in mind one must decide whether an action is e.g. to be construed as coercion or manipulation. An objective "reasonable person standard" (ibid. p. 32 f.) is unsuitable for the definition of autonomy of agency. *Second*, there is a gray area between the three kinds of influence. So-called "hard cases" (ibid. p. 260) cannot be fully resolved by any theory, as the concepts of everyday speech to be drawn on here permit only demarcations that are just as inexact as social sciences (ibid. p. 262). And *third*, while coercion and belief are not gradualizable categories, manipulation, in contrast, permits various graduations. So in this way, i.e. via this form of exertion of influence, a further gradualistic element enters into Beauchamp's conception of autonomy of agency. Autonomous actions can thus be autonomous to a different degree, because they rest on a different degree of understanding and come about under different degrees of manipulative influence.

### 5.4.2 *The Tacit Preconditions for Autonomy of Agency*

The aim of this section is to invalidate the first two of Beauchamp's premises in order to reject his conclusion that autonomy of agency is the only appropriate basis for the principle of respect for autonomy within the context of biomedical ethics. For one thing, Beauchamp's thesis that his concept of autonomy of agency is

---

<sup>68</sup>With this analysis Beauchamp for one thing avoids the false conclusion of having to assess communicative actions as not generally being instrumental actions. For another, he makes it clear that certain rhetorical elements do not belong in the area of belief but in that of manipulation (ibid. p. 262).

independent of a model of personal autonomy will be discussed. Here it will be shown – thus my counter thesis – that the latter constitutes both a direct and an indirect basis for the former and that the postulated independence is nonexistent.<sup>69</sup> For another, my aim is to prove that Beauchamp's second premise is ultimately also unconvincing. My counter thesis reads here that in biomedical ethics the respect paid factually to autonomy not only refers to autonomy of agency, but also and always to personal autonomy.

Although Beauchamp's conception of autonomy of agency is also tailored to general validity in the area of biomedical ethics, he develops it primarily in the context of an analysis of informed consent. It is not insignificant that he thereby imposes a restriction on explicit informed consent, although he knows other forms which are likewise kinds of consent. Thus, he discusses implicit and putative consent as two forms of tacit consent (Beauchamp and Childress 1994, p. 128) which are recognized in our ethical practice and are also relevant for biomedical ethics. Even if Beauchamp – as his discussion of the permissibility of HIV tests even without the explicit consent of the affected party proves exemplarily (*ibid.* p. 130) – by and large prefers not to resolve these conflicts by means of the concept of tacit consent, but rather, identifies it as an ethically legitimate overruling of the patient's autonomy, he concedes that in other areas forms of tacit consent are relevant. Admittedly, as he demonstrates, this form of justification bears serious risks, since here the possibility of extensive abuse would be opened up. Thus he fights against e.g. constructing a general theory of human goods or rationality (*ibid.* p. 128) as a yardstick for putative consent. As for implicit consent, it is the respective person's actions and omissions and our knowledge of personal ideals and beliefs that should be drawn on as a yardstick.<sup>70</sup> Beauchamp's treatment of these alternative forms of consent is important for our question for two reasons. For one thing, he himself points out that ethically acceptable handling of implicit and putative consent must relate to the patient's personality. The standard of our assessment must be "what we know about a certain person" (*ibid.*). Thus the personality is acknowledged as an indirect basis at least for these forms of consent. For another, by acknowledging these other forms of consent, Beauchamp erodes his strongest argument in favor of the thesis that autonomy of agency is the sole beneficiary of respect for autonomy. Differentiating between the capacity of autonomous actions and the execution of these capacities in fact attests that our respect must refer to the actual execution of these capacities, not just to the capacity in principle (Faden and Beauchamp 1986, p. 237). If a patient has the general capacity of performing autonomous actions, but does not exercise these capacities correctly in a concrete situation, then the attribution of the general

---

<sup>69</sup> In doing so, I differentiate between a direct and an indirect basis, whereas the first can be either a semantic or a material basis. A is an indirect basis for B if and only if X's A-being provides criteria for deciding whether X is B or not. A is a direct semantic basis if and only if the characterization of X as A implies logically that X is B. And A is a direct material basis for B if and only if the A-being of X is necessary for X to be B.

<sup>70</sup> Cf. on this also the concept of the value-base-line as a means of being able to respect personal autonomy as extensively as possible (Caplan 1992, Chap. 16).

capacities does not provide an argument for respecting the concrete decision or action. The far-reaching conclusion Beauchamp draws from this certainly valid estimation is that for informed consent it is not the general capacities but individual executions of these capacities that are necessary. Beauchamp weakens his own point of departure with the concession that there is respect for implicit or putative consent which is indispensable for biomedical ethics, but which does not in fact consist in the actual execution of capacities relevant for autonomy of agency. Furthermore, one must ask whether, with his differentiation between the attribution of the general capacity of autonomy of agency and its execution, he is really describing a difference which can support the thesis of the independence of autonomy of agency from personal autonomy. Beauchamp's train of thought is that personal autonomy in the form of capacities can be present without their actual execution. But respect for explicitly voiced informed consent (or refusal) is respect for the executed capacity. Thus the respect for personal autonomy cannot be adequate to explain our respect for explicitly informed consent. That much is conceded. But, to justify the thesis that personal autonomy is no necessary condition, he must either argue that our respect for an autonomous action does not refer to the agent's capacities at all, but concerns only the feature of the concrete act token (Option 1). Or he must show that the capacities underlying the autonomous action are independent of those that constitute personal autonomy (Option 2). The *first* option is clearly implausible. If one envisages a patient who performs an action which coincidentally exhibits all the features required of an autonomous action, then it is clear that one will not pay respect to this action itself. The action will be respected either because it coincides with our wants or our understanding of what one should or could do rationally in that specific situation.<sup>71</sup> Or one will respect it for quasi pedagogic reasons, to promote the capacities for autonomous action.<sup>72</sup> If we respect a concrete autonomous action in the full meaning of the word, then we respect it because the capacity of the agent in question to perform autonomous actions is manifested therein. But without the underlying capacity, i.e. if it has occurred by chance, we do not respect the autonomy of the action.<sup>73</sup> Thus, only the *second* option remains to Beauchamp,

---

<sup>71</sup> One has only to think of the easing of requirement conditions for informed consent in mentally sick patients with the aim of avoiding the problem of compulsory sterilization.

<sup>72</sup> This is e.g. the case in children whose decisions one respects in this perspective, so as to make the acquisition of necessary capabilities possible and attractive.

<sup>73</sup> So my thesis is that the respect does not refer to the actual properties of a concrete action, but to these properties qua manifestation of the agent's capabilities. This becomes obvious in coincidentally autonomous actions. These are possible because the logic of "knowing" permits the following:

- (i) X performs an action h.
- (ii) This action can be described as a case of action type A.
- (iii) The assertion that X can perform actions of type A is false.

If someone once manages under certain circumstances to shoot a goal from 100m distance, we do not in an everyday sense attribute to him the capacity of shooting goals from this distance. If, in contrast, a tennis player is able to get 99 out of 100 serves in, then we attribute this capacity to him even if he has just served into the net. This feature of our attribution practice has evolved above all in the context of the analysis of freedom of will and responsibility (cf. Kenny 1973 and Fischer

which he takes by defining autonomy of agency through the capacities of the agent. But is it really the case that the capacities on which he draws are independent of those capacities that constitute personal autonomy? There are three reasons why Beauchamp, given his conception of autonomy of agency, cannot successfully opt for this second alternative. For one thing, the required capacities of the agent either presuppose the capacities required for personality or they coincide with them (a.). And for another, his endeavor to substantiate the three pillars of autonomy of agency in a way that will render the concept ethically sustainable shows that it refers to the personality of the agent. This becomes clear both in the integration of the subjective elements in Beauchamp's approach (b.) and in his allowance for the biographical dimension of human personality (c.).

*ad (a.):* Beauchamp's definition of an intentional action as an "action willed in accordance with a plan" (ibid. p. 245) already takes recourse to the capacities of the agent which are also pertinent for personality.<sup>74</sup> He must be in a position to grasp first-person propositions and in addition have knowledge of his own existence over a certain period of time. Otherwise it would be impossible for him to recognize a process carried out over time as the realization of a plan of action. Since Beauchamp subsumes not only single events, but also sequences of events qua plan of action under the concept of intentional actions, his conception incorporates this knowledge about one's own existence through time periods. Thus the respect for a person's autonomous actions takes recourse to capacities that are constitutive for personality.

*ad (b.):* In his endeavors to develop a conception of autonomy of agency which yields ethically acceptable results, Beauchamp suggests that the basic categories of his model should be interpreted according to the subjective standard. This subjective standard requires of the treating physician that he adjusts his treatment to the specific needs, ideals and beliefs of his respective patient (Faden and Beauchamp 1986, p. 33). With this, even on the most generalized level of Beauchamp's theory, the patient's personality comes into play as a relevant parameter. But in the context of autonomous actions that is of interest to us here, the personality also becomes a direct basis for the autonomy of an action. Thus Beauchamp asserts for one thing that the third pillar of autonomous action i.e. the absence of external control, is to be interpreted in the sense of the subjective standard. Therefore there is no general material list of conditions that could count equally for all persons as coercion or impermissible manipulation. Rather, the individual beliefs, wants and ideals must be taken into account in order to determine whether and to what degree a specific action is autonomous (ibid.

---

1995, p. 25–29). I don't think it's a coincidence that it's also about an evaluative context. As in the concept of freedom and responsibility, in the analysis of the autonomy of agency, the evaluative aspects are also unavoidably implicated in the constitutive criteria.

<sup>74</sup>The thesis supported in the second chapter of this study, that the capacity of being an agent is not sufficient to justify a simple conception of personal identity, is compatible with the above formulated thesis that the capabilities necessary for actions include also capabilities which are necessary for personality.

p. 261). For another, Beauchamp has also integrated into the second pillar of autonomous actions an element which likewise refers to the agent's respective personality. Since the ideal of complete understanding of one's own action is unattainable, Beauchamp restricts this understanding to the relevant aspects of the action in question (*ibid.* p. 252). As he himself then explains further, actions do not only have generally relevant aspects (e.g. being life-threatening), but also aspects whose relevance results from the specific personality of a patient. According to Beauchamp's conception of autonomous actions, an action is not autonomous if the agent overlooks or does not adequately register an aspect that is relevant as measured by his own beliefs, wants and ideals. Because both these criteria filled with content according to the subjective standard are constitutive for the autonomy of an action, the personality of the agent is a direct basis for the autonomy of an action. Actions are not generally autonomous *qua act type*, but only relative to the respective personality of the agent. In the context of the question of whether informed consents (or refusals) covering a longer time period or given by a patient for her own future in the form of advanced directives merit respect as autonomous actions, it appears that the concept of autonomy of agency cannot get by without a model of personal autonomy.<sup>75</sup> As Beauchamp himself explains, whether it fits the respective personality is a measure of the respectability of such 'long-term decisions'. In view of the fact that humans can change their minds and ideals in the course of time, two basic conflicts arise here, above all relating to advanced directives. For one thing, if the patient expresses no actual opinion, one can ask whether that decision is still the expression of her present wish. For another, one can, when e.g. there is a currently different patient's decision, under certain circumstances also ask which of the two decisions, the present one or the earlier one, is the expression of her 'real' want (cf. Chap. 7). For such cases, Beauchamp allows the fact that a decision is "in character or out of character" to be a criterion for the autonomy of the decision (Beauchamp and Childress 1994, p. 130). So the personality, understood in the sense of evaluative self-understanding, is again made a basis of autonomy of agency. Beauchamp himself admittedly restricts this result in two respects which, given his basic thesis that autonomy of agency rather than personal autonomy is decisive, is no wonder. Thus he says, for one thing, that the autonomy of the actions *can rest partly* on this coherence. And for another, he would like this criterion of the coherence of a personality to be understood only as an indirect basis:

Acting in character is not a necessary condition of autonomy. At most, actions that are out of character can be caution flags that warn others to seek explanations and probe more deeply into whether the actions are autonomous (*ibid.*).

---

<sup>75</sup>In the seventh chapter I will try to show that the conception of personality developed here is able to testify to the ethical worth of advance directives as a case of the principle of respect for autonomy. In these advance directives the explicit evaluative self-understanding of a person i.e. her personality is discernible.

However, his claim that this does not involve a necessary condition can be interpreted in two ways. To be sure, his observation that we sometimes act autonomously “out of character”, to have a “new experience” are hardly contestable (Faden and Beauchamp 1986, p. 266). But then, this wish to have new experiences now and then or to give one’s own life surprising twists belongs to that person’s personality. For without this background and this general expectation we would not be able to understand the concrete change of direction as an autonomous decision. His submission that new and unfamiliar circumstances can lead to decisions which do not tally with the previous personality is also ultimately unconvincing. One can understand a person’s actions as autonomous because one can see them as that personality’s attempt to orient to those circumstances. Such actions are therefore less „out of character“ and in fact strategies for preserving one’s own personality in the new situation. If, with his thesis that an action can be both „out of character“ and autonomous, Beauchamp means cases in which the personality undergoes partial revision, then his description is certainly acceptable. But neither does this inference argue against personality as the basal category for our assessment here. A personality must not be understood as an invariant set of wants and beliefs; but rather, personal autonomy is also manifested in the flexibility with which new and possibly unexpected circumstances are integrated into one’s own biography. It is precisely our understanding of biographical identity that allows partial revisions of a personality as autonomous decisions to become understandable and respectable. The complete revision of a personality, which generates a break so that there is no way we can draw on the biography of this person to interpret the new personality, can, however, no longer be interpreted in this way. Consequently, neither is the conversion from Saul to Paul portrayed as his autonomous action.

*ad (c.):* In the context of his rejection of the condition of authenticity, for which he always assumes reflexive authenticity, Beauchamp tries to make a plausible case for an action not only becoming autonomous by way of an act of reflexive authenticity, but solely due to her intentionality, the understanding manifested therein and the absence of external control (Faden and Beauchamp 1986, p. 265). In an elucidative comment he then explains that this third condition not only includes the absence of actual external control, but should also preclude that the wants and beliefs active in the current action are the result of an impermissible manipulation or coercion at an earlier point in time (*ibid.* p. 273 fn. 36). As Beauchamp rightly observes, this does not imply that the agent has acquired this want or belief through reflexive identification. But nevertheless, this shows that the autonomy of an actual action cannot be guaranteed without recourse to specific conditions which the development of the agent’s personality must satisfy. With the concession that there are ways of acquiring wants and beliefs which endanger autonomy of agency, Beauchamp at the same time recognizes that the autonomy of an action cannot be analyzed satisfactorily without involving the biographical dimension of the personality.

### 5.4.3 *The Principle of Respect for Autonomy*

The third premise of Beauchamp's argument states that the attempt to take personal autonomy as the basis of the principle of respect for autonomy leads to unacceptable consequences in the context of biomedical ethics. He blames the condition of (reflexive) authenticity anchored in hierarchical models for this. It is rejected as too demanding a condition for autonomy of agency (Faden and Beauchamp 1986, p. 273 fn. 37). Whilst the possibility is conceded that theories of personal autonomy do not have to be so demanding that the autonomy of an action is understood as being "conform to an autonomous person's elected life plan" (ibid. p. 236), nevertheless, a glance at our factual implementation of the principle of respect for autonomy shows that in the implementation of the condition of authenticity, many actions would not have to be designated as worthy of respect which do earn recognition (ibid. p. 265).<sup>76</sup> But such judgments would be "ethically unacceptable" (ibid.) Elsewhere, the objection of excessive requirement and the resulting inadequacy of the conception in question are expressed even more sharply:

One problem with all such exacting requirements for autonomy, including second-order-desire theories, is that few choosers, and also few choices, would be autonomous if held to such standards, which in effect present an aspirational ideal of autonomy. No theory of autonomy is acceptable if it presents an ideal beyond the reach of normal choosers (Beauchamp and Childress 1994, p. 123).

To show that this objection does not really apply, I will formulate two general counter objections in the following. Subsequently, I shall discuss the examples described by Beauchamp, which should prove that the factual implementation of the principle of respect for autonomy does not depend on personal autonomy but on autonomy of agency. This will show that these examples cannot be counted as proof of Beauchamp's thesis.<sup>77</sup>

The first counter objection against Beauchamp's critique is that even his own conception of autonomy of agency is not shielded *per se* from the objection of excessive demand. As he himself explains, one must not infer from the fact that actions are never completely informed, voluntary or autonomous that they are never adequately informed, voluntary or autonomous (Beauchamp and Childress 1994,

---

<sup>76</sup> Beauchamp equates authenticity with reflexive identification. It is important not to confuse this with the pre-reflexive authenticity I brought in earlier. It should also be noted that Beauchamp certainly recognizes that in his sense i.e. reflexive, authentic decisions merit respect (Faden and Beauchamp 1986, p. 265). His objection is not that personal autonomy in this strong sense does not merit respect, but rather, that personal autonomy is an unsuitable yardstick for the implementation of the principle of respect for autonomy.

<sup>77</sup> The invalidation of Beauchamp's third thesis is still relevant independently of the results of my discussion in the second paragraph of this section. If, firstly, the thesis of independence from personal autonomy and autonomy of agency cannot be sustained, but secondly the objection of excessive demands cannot be invalidated, then one must generally forswear the principle of respect for autonomy as the basis of biomedical ethics. Such a position is argued by Bernard Gert, admittedly on the basis of the assumption that the concept of autonomy is contextually not definable precisely and the difference between moral duties and ideals are blurred (cf. Gert et al. 1997, p. 81 f.).



p. 158). As a consequence, he does not use an ideal standard in his model of autonomy of agency, but a concept of substantial autonomy, the fulfillment of which is realist (Faden and Beauchamp 1986, p. 241). The question now is why such a standard should not be applicable in the framework of personal autonomy. There is no principle reason not to likewise substantiate the degree and extent of reflection presumed for personal autonomy.

This leads directly to the second main objection to Beauchamp's critique. He imputes a strong conception of personal autonomy in which a person is required to identify herself in a reflexive way with each of her beliefs and wants.<sup>78</sup> However, his criticism is not invalidated therewith, since in one section of his book, Beauchamp discusses possible reformulations of the authenticity condition (Faden and Beauchamp 1986, p. 266 ff.). Therein, two basic strategies are proposed for dealing with the problem of excessive demands, which correspond largely to the further developments of the hierarchical approach and have also been adopted in this study. Getting one of the strategies, a "less ominous criterion" (ibid. p. 266) consists, according to Beauchamp, in requiring non-reflexive identification, but merely stability or coherence among the values underlying a decision. The other strategy, which Beauchamp considers more promising and which has been mainly adopted in this study, consists not in making reflexive identification a condition but the absence of resistance (ibid. p. 267).<sup>79</sup> Since Beauchamp also sustains his objection of excessive demands for this modified conception and submits counter-examples, it must now be examined whether these examples are really suitable to show that personal autonomy is in principle an unsuitable basis for the principle of respect for autonomy. Against the strong conception of personal autonomy, according to which reflexive identification with a want (or belief) is necessary and sufficient for the action in question in which these wants are realized to be regarded as autonomous, Beauchamp cites two kinds of examples. In the first case, the action of a politically engaged woman is described, who is explicitly concerned about her upcoming voting decision, but at the same time has never critically scrutinized the basic ideals and beliefs underlying her political decision (Faden and Beauchamp 1986, p. 265). Beauchamp's objection is that it would be unacceptable for this woman's voting decision not to qualify as autonomous simply because she has never identified herself reflexively with her ideals and beliefs. According to Beauchamp's supposed strong conception of personal autonomy, this would happen, since this woman has not critically acquired all wants and ideals relevant to this voting decision. Beauchamp's criticism is justified and shows that the strong conception does not adequately cover our practice of respect for autonomous actions. With the second kind of example Beauchamp then tries to show that even cases in which there has

---

<sup>78</sup> Alongside the objections discussed above, he also formulates the standard objections to hierarchical conceptions of personal autonomy, which have already been allowed for in my own conception (Beauchamp and Childress 1994, p. 122).

<sup>79</sup> In a footnote, Beauchamp refers to an as yet unpublished paper by Frankfurt, through which he became aware of this second option, but without mentioning its title (Faden and Beauchamp 1986, p. 273 fn. 39).

been no reflexive identification at all, can be legitimately understood as autonomous actions. One of this type of case is actions in which social conventions are observed or satisfied or someone behaves according to certain morals or customs (ibid. p. 264). As Beauchamp correctly records, such actions are often executed without the agent having first consciously dealt with and reflexively identified these socio-cultural practices. This criticism of Beauchamp's is also justified and shows that a conception of personal autonomy must not be set up in a way that renders autonomy and socialization incompatible (cf. above Sect. 5.3.1.3). The second type of case are actions by persons who derive their ideals and beliefs from particular world views. Here, Beauchamp mentions the decision of a Jehovah's Witness not to allow blood transfusions, whereby this person has never given thought to the values and beliefs which are the basis of this decision (ibid. p. 265). Beauchamp considers it inappropriate not to qualify this decision as autonomous. Even if one could overrule it in favor of other ethical principles, it would be a debasement of this action if one did not even recognize it as autonomous. Elsewhere, when discussing the relation of autonomy, authority and community, Beauchamp again refers to the Jehovah's witness and now describes him as someone who has decided to follow that world view (Beauchamp and Childress 1994, p. 124). This decision is the reason that there is no conflict between autonomy and authority and the decision to refuse a blood transfusion could count as autonomous. Therewith, Beauchamp seems to justify the autonomy of the decision not to allow a blood transfusion by the decision to align his own life to this religious orientation. Different from the case of following social conventions, morals and customs, the autonomy of the action is thus after all linked back to a reflexive decision.<sup>80</sup>

If these examples were meant as objections to a strong conception, in which reflexive identification with wants and beliefs is necessary and sufficient for autonomy, Beauchamp now also turns against a modified hierarchical analysis of personal autonomy in which not reflexive identification, but the weaker form of the absence of explicit repudiation is required („not reflexively repudiate“).<sup>81</sup> Beauchamp

---

<sup>80</sup>Cf. for a detailed discussion of this problem my analysis of the dispute between Berofsky and Dworkin (Sect. 5.3.2.1). My proposed conception of prereflexive authenticity enables the respect for the decision made by the Jehovah's Witness to be understood as respect for his personal integrity (cf. Sect. 5.3.2.2).

<sup>81</sup>The other option, which Beauchamp regards as a further development of the hierarchical approach, replaces the condition of reflexive identification by the criterion of stability or consistency of ideals underlying the decisions, which is considered necessary and adequate (Faden and Beauchamp 1986, p. 266). Such a conception is understood here not as personal autonomy but as prereflexive authenticity. Beauchamp's objection that such a modified conception would still exclude too many actions as not autonomous, is admittedly unconvincing. Thereby, he recurses to the possibility of executing an action "out of character". True, in his eyes it is also *prima facie* an indication that the action is not autonomous. But on closer scrutiny it can turn out that e.g. it has been executed due to the want to have "a new experience" (ibid.). But in the latter case the agent must refer reflexively to his ideals and beliefs. For one thing, he must be familiar with his ideals, must – at least for this action – mask some of them (reflexive negative identification) and above all consciously pursue the want to have a new experience. Herewith, this case can be described adequately in the framework of the model of personal autonomy proposed here.

himself considers this to be the “promising” strategy for a “developing narrower criterion of authenticity” (Faden and Beauchamp 1986, p. 266 f.). One advantage of this conception is *prima facie* that in this way actions which according to Beauchamp’s criteria must be viewed as autonomous while not being intuitively evaluated as such by us, can be classified as not being autonomous. He names as examples weak willed actions and pathologically compulsive acts which are experienced by the acting person as compulsion and are only executed with reluctance (e.g. *ablutomania*). Beauchamp counteracts this disadvantage of his own theory with the observation that the reaction to the model of personal autonomy would invert everyday forms of action in which possible opposition or reluctance of the agent did not really endanger the autonomy of the action. Here, the proposal to specify the absence of reluctance as necessary and sufficient for autonomy would again lead to actions which are autonomous according to our prior understanding, having to be regarded, unacceptably, as not being autonomous. Beauchamp’s example is a housewife who wants to free herself of the ideals tied to this role and thus does the necessary housework reluctantly. But her activities continue to be uncontrolled, intentional and invested with a high degree of understanding and therefore certainly autonomous. This example of everyday actions not taking on pathological forms, which are executed with some reluctance, is not a convincing objection. For one thing, it should be noted that actions are only executed intentionally, willingly or reluctantly under certain descriptions. In the case of the housewife this means that e.g. she evaluates making coffee positively under one description (“to be able to drink coffee with a piece of cake”), but negatively under another (“as an expression of the reduction of women to the role of the housewife and of the dominance of the patriarchic world”). Even if the actions are one and the same event<sup>82</sup>, they are *qua* intentional and evaluated still not independent of the (true) descriptions. Thus, ultimately this example does not contradict, but actually supports the concept of personal autonomy proposed here. With his criterion of relevance, Beauchamp had also established an indirect link to the agent’s personality, which now becomes important. If one considers that not only autonomy of agency but also personal autonomy can come in degrees, then such an action can be counted as autonomous as long as the reluctance does not lead to incapacity of action, i.e. „significantly effect” (Christman 1993, p. 287) it. The presence of resistance does to a certain extent harm the condition of autonomy which Frankfurt calls „wholeheartedness“ and leads to a reduction in the autonomy of this action. But such a gradual evaluation does correspond fully to our intuitive perception.<sup>83</sup> It is e.g. possible that, taking

---

<sup>82</sup>The problems of general theory of agency linked herewith are to remain unconsidered here.

<sup>83</sup>This is backed by the fact that we do not consider weak-willed actions to be properly autonomous. This phenomenon also rests on actions being intentional under certain descriptions. Under one description the agent aspires to eat an ice (“ice-cream is tasty”), but not under another (“ice-cream makes you fat”). The general valuation of giving preference to the value of being slim over the value of experiencing enjoyment does not become effective for the agent although he wants this value were his will. With this he understands his action of ordering and eating an ice (for this reason Beauchamp’s conception of autonomy of agency also comes into it here). But in a certain sense he does not understand himself (the level of personal autonomy). Thus the concept of personal autonomy provides the correct analysis and an apt reconstruction of our intuitive attitude.

all her ideals and beliefs into consideration, the housewife in our example comes to the conclusion that the action of making coffee is all in all the most appropriate option here. But then there is no reason not to interpret the resistance to this action as the expression of a certain incoherence in her personality, given the empirical circumstances.<sup>84</sup> If one bears in mind that the absence of resistance is not the only criterion and is open to gradual interpretation, then the objection raised by Beauchamp can not only be invalidated, but also basically shows once again that a usable concept of autonomous actions cannot ultimately do without an adequate model of personal autonomy.<sup>85</sup>

The advantages of relating the principle of respect for autonomy to personal autonomy (or prereflexive authenticity on personal integrity in borderline cases) are firstly that the recourse to the agent's capacities can be made explicit. Secondly, implicit or putative consent can be included as adequate forms of justification as long as they remain related to the patient's personality.<sup>86</sup> In all, both the autonomy of individual actions and the personality with the constitutive capacity of acting autonomously are indispensable for the adequate justification and implementation of the principle of respect for autonomy.<sup>87</sup> Moreover, the analysis of Beauchamp's arguments and theses has shown that a concrete implementation of this principle on the concept of personality developed here in the sense of the biographical identity of a person cannot be dispensed with.<sup>88</sup> The conception of personality not only

---

<sup>84</sup>If this action were to be executed quasi mechanically, as a reflex to practiced behavior patterns, or if she were to execute this action with such reluctance that she understands herself as being weak-willed, then one would no longer be able to assess the action as autonomous.

<sup>85</sup>Beauchamp concedes emphatically that at least for some problem cases (pathological compulsive acts, weakness of will) his model of autonomy of agency is just as unsatisfactory here as a hierarchical concept of personal autonomy. His reason is that we need "a theory of the self and of self-identity" which we do not have, in order to solve these cases (Faden and Beauchamp 1986, p. 268). In my view, the conception of personality and personal autonomy proposed here is suitable for delivering a useful basis for the treatment of such and less dramatic cases.

<sup>86</sup>Hereby it should be borne in mind that whilst personal integrity is fashioned by the individual within the framework of her social conditions and natural properties, this fashioning must not be interpreted according to the model of autonomous action, at least not in every case. For otherwise there would be firstly, infinite regress to the analysis of autonomy, since decisions would also have to be understood as autonomous actions. Secondly, the analysis of autonomy of agency in concepts of personal autonomy attempted here would open out into an explanatory cycle and thirdly, the argument of excessive demands could not then be really dispelled.

<sup>87</sup>This result implies explicitly that criteria of autonomy of agency as proposed by Beauchamp are also indispensable for an adequate interpretation of the concept of informed consent and for an acceptable implementation of a principle of respect for autonomy. But this autonomy of agency is dependent on a suitable model of personal autonomy resp. the integrity of a personality.

<sup>88</sup>In contrast, Beauchamp quite rightly criticizes that the identity of the person – in the sense of persistence – is not sustainable in the context of the ethical evaluation of advance directives (Beauchamp and Childress 1994, p. 132). I fully agree with his thesis that these decisions about one's own future are also autonomous (cf. Chap. 7). Due to the differentiation between personal identity as human persistence on the one hand and biographical identity on the other, I am, however, able to link this respectability back to personal autonomy. Beauchamp's rejection of the complex theories of human persistence seems for him to simultaneously coerce the rejection of personal autonomy as the basis of these autonomous decisions. In this study I will try to show that such a link can be severed for good reasons.

allows development of an adequate concept of autonomy. It also provides a plausible basis for the principle of respect for autonomy (cf. also Quante 2011a). Thus, it has now been shown in a general way that personal identity is a relevant principle for biomedical ethics. In the following chapters the model of personality developed up to now will be further defined contextually on the basis of the discussion of concrete problems in biomedical ethics.

# Chapter 6

## Dying Autonomously

### 6.1 Targets and Preliminary Clarifications

There is, as Tom L. Beauchamp (1996, p. 1) observes, no longer-lasting or stronger proscription in medicine than that which says that one must not kill patients. The trend towards the principle of respect for autonomy and medical-technological developments have, however, led increasingly to this ban coming under fire in recent decades, whereby these two tendencies go hand in hand: where medical possibilities, if anything, promote the quantity rather than quality of life, and patients can (and must) be kept alive under conditions which are not considered worth living, the question arises as to why a human should not be allowed to end these phases in his life (or have them ended). The more dominant the notion of an autonomous life style and the basic attitude that ultimately everyone is solely responsible for his valuations becomes, the more the claim suggests itself that doctors should, with their knowledge and skill, provide relief for the suffering of patients who are merely surviving.

At the same time, there have always been exceptions to the prohibition of killing, which are considered justified by many. One needs only think of cases of self-b, homicide in war, or the death sentence. Even in the narrower medical context there were and are always situations in which it seems *prima facie* appropriate to allow exceptions to the prohibition of killing. Philosophical theory has always tried to come to terms with these exceptions, i.e. to show that in this context different kinds of actions can be distinguished as regards their ethical quality. The differentiation between active and passive and respectively voluntary, nonvoluntary and involuntary euthanasia<sup>1</sup> today thereby plays just as central a role in the ethical debate as the

---

<sup>1</sup>In Germany, different from most other countries, the term euthanasia is weighted down above all through National Socialism. Not least because of this, in the German debate the term mercy killing has become established excepting the strongest opponents. Whilst I have dispensed with the term elsewhere so as to avoid emotional associative chains, which frequently take the place of arguments, in the following I will use the term euthanasia because it is the most suitable; cf. also my deliberations on this in Quante (2014b, p. 168 f.).

doctrine of the double effect according to which one can and must make an ethical difference between the intentional and merely anticipated aspects of an action.

The following is concerned with proof that the conception of biographical identity of persons unfolded here is to be seen as an argument for the permissibility of *voluntary* euthanasia. In other words: If the analysis of personality I propose here is plausible, then a categorical ban on voluntary euthanasia cannot be justified. This finding does not however rule out that there are possibly relevant arguments fit to tip the scales against the ethical permissibility of such homicidal actions in the framework of weighing alternatives. The perspective on the problem of euthanasia arising from the topic of my study implicates that only one section of the whole complex can be dealt with, so that the following will only be about the case of voluntary euthanasia. The case of involuntary euthanasia, which has in any case hardly ever been discussed seriously, in which a person is killed against her autonomous will, plays only an indirect role in the discussion of slippery slope arguments below.<sup>2</sup> But even the indisputably profoundly problematical case of nonvoluntary euthanasia can only be mentioned in passing in this context, although it is an urgent social problem. This is because nonvoluntary euthanasia applies to those patients in whom there is and can be no autonomous response, so that they are not included in my superordinate problem. Furthermore, I will only be able to deal with the other differentiations relevant to the discussion (active v. passive, act v. omission, intend v. accept) in the framework of the discussion of classical objections. In the following I assume that the differentiation between active and passive euthanasia cannot mark a categorical ethical difference, nor can the differentiation between intended and accepted aspects of actions or consequences of actions constitute a categorical ethical difference.<sup>3</sup> For this reason I am using the term voluntary euthanasia in its wide sense so that it includes assisted suicide as a special case.<sup>4</sup> Thereby the terms

---

<sup>2</sup>I also count the following case as involuntary euthanasia. If a patient capable of autonomous decisions does not respond to the question of whether he would like to be killed, or when he makes no effort on his part to ask for his life to be terminated, then his killing is a case of involuntary euthanasia. The missing expression of the wish to be killed is enough to attribute the person in question with a positive identification with his own existence, so that this evaluation must be accepted. This also applies when this patient does not comment on such a suggestion or does not resist a homicide attempt. In other words: As long as the capacity for autonomous decisions is present, the patient's request for an ethically permissible case of euthanasia is always needed.

<sup>3</sup>I have dealt in detail with these differences and their relation from both conceptual and ethical viewpoints elsewhere; cf. Quante (2014b, chapter VIII) and Siep and Quante (1999).

<sup>4</sup>Sissela Bok (1998, p. 87 ff.) differentiated between various types of theory as regards the ethical evaluation of assisted suicide and voluntary euthanasia (in the narrower sense of killing another person). My position can be classified as theory type A: there is no principle (or categorical) difference in value between assisted suicide and voluntary euthanasia (in the narrower sense; both forms are however subject to certain restrictive conditions). Bok herself defends position B for reasons which I shall discuss presently (cf. Sect. 6.4), in which assisted suicide is ethically permissible under certain restrictive conditions, whereas voluntary euthanasia (in the narrower sense) is to be rejected in principle. On the other hand, the third type of theory, which above all coincides with the position of the sanctity of life (cf. Sect. 6.3), categorically rejects both forms of killing.

“euthanasia” and “kill” must be understood in a purely descriptive sense. This condition of neutrality is indispensable, so as to be able to avoid a petition of the ethical evaluation of euthanasia treatments. In all, for my purposes the following definition which I am taking over from Beauchamp and Davidson (1979, p. 304) is adequate and based on the following argumentation. This definition is that:

The death of a human being, A, is an instance of euthanasia if and only if

1. A's death is intended by at least one other human being, B, where B is either the cause of death or a causally relevant feature of the event resulting in death (whether by action or by omission);
2. there is either sufficient current evidence for B to believe that A is acutely suffering or irreversibly comatose, or there is sufficient current evidence related to A's present condition such that one or more causal laws support B's belief that A will be in a condition of acute suffering or irreversible comatoseness;
3. (a) B's primary reason for intending A's death is cessation of A's (actual or predicted future) suffering or irreversible comatoseness, where B does not intend A's death for a different primary reason, though there may be other relevant reasons, and (b) there is sufficient current evidence for either A or B that causal means to A's death will not produce any more suffering than would be produced for A if B were not to intervene;
4. the causal means to the event of A's death are chosen by A or B to be as painless as possible, unless either A or B has an overriding reason for a more painful means, where the reason for choosing the latter causal means does not conflict with the evidence in (3.b);
5. A is a nonfetal organism.

(Definition of “Euthanasia” according to Beauchamp & Davidson)

Although, due to my limitation to the case of voluntary euthanasia, not all the conditions of this complex definition<sup>5</sup> will be relevant in the following, I would nevertheless like to make a few explanatory comments.

So, *firstly*, the fifth condition serves only to exclude abortions from euthanasia treatments, a decision which is acceptable by virtue of the differing ethical problem, but which is itself already also ethically laden. The fact that the above definition is thus not purely descriptive on this point has no negative consequences for the following question, since it remains ethically neutral as regards the specific aspects of the euthanasia problem. *Secondly*, it should be noted that A must be a human being, so that animals cannot be the object of euthanasia treatments. *Thirdly*, A and B must be two numerically different humans. This excludes pure cases of suicide without

---

<sup>5</sup>Beauchamp (1996, p. 4) later presented a less elaborate definition which fulfills the purpose there and largely agrees with the above: “In this introduction, a death will be considered euthanasia of any type if and only if the following conditions are satisfied: (1) The death is intended by at least one other person whose action is a contributing cause of death; (2) the person who dies is either acutely suffering or irreversibly comatose (or soon will be), and this condition alone is the primary reason for intending the person's death; and (3) the means chosen to produce the death must be as painless as possible, or a sufficient moral justification must exist for a more painful method”.



the intentional involvement of others from the area of euthanasia treatments, but includes cases of assisted suicide, whereby I understand the first condition thus, that B, because of his actions (or omissions), is a cause or causally relevant aspect of the sequence of events leading to A's death. If it is the case that B is a physician (which is not explicitly required by the above definition), then the result is a "physician assisted suicide".<sup>6</sup> Moreover, *fourthly*, the first condition identifies a factor that is independent of the participants' descriptions with the causal relation. An action is only then a case of euthanasia when the implied causal relation really does exist.<sup>7</sup> Alongside this external element, *fifthly*, a further aspect appears which also transcends the inner perspective of those concerned. Recourse to "sufficient current evidence" – in conditions (2) and (3.b) – guarantees that the beliefs of A and above all of B correspond to the epistemic standards, in the case of a physician i.e. the anticipated medical knowledge. With this, an evaluative aspect is built into the condition, which permits irrational or grossly negligent cases to be excluded. The condition of neutrality is, however not damaged herewith, as these evaluative aspects are general standards of rationality, knowledge and justification, which are not specific to the ethical permissibility of homicidal actions. And *sixthly*, it is also important that the characterization of an action as euthanasia must not be understood in the sense of a success verb. The second, third and fourth conditions are therefore consistent in that the preconditions made by A and B are false. Whereas the causal relation in the first condition must exist factually for an action to be counted as euthanasia, both A and B could, e.g. in respect to A's real state of health, be just as mistaken as to the question of whether the chosen method is really "as painless as possible". The truth of the beliefs named in the second, third and fourth conditions is irrelevant for the question of whether an action really is euthanasia. In contrast, they do of course continue to be relevant as justifying reasons for B's actions.

Thus, all in all, the above definition does, through the causal condition, at least keep open the option of differentiating between doing and omitting to provide a basis that is independent of our valuations.<sup>8</sup> Furthermore, it allows for the denotation of an action as euthanasia also having to draw on the perspective of those concerned, whereby at the same time an intersubjective yardstick is integrated which

---

<sup>6</sup>The above definition embraces all three kinds of causality which fall under the judicial concept of accountability: direct homicidal actions, indirect homicidal actions through the creation of suitable circumstances and also negative causation through omissions (on the problem of negative causation cf. Birnbacher 1995).

<sup>7</sup>The epistemic problem of under which conditions we can know that a causal relation is really present, must remain undiscussed here. However, it is important that with this condition an external element is being included in the definition: Whether an action is a case of euthanasia depends – according to the demand of the definition – not just on the perspectives and interpretations of the participants.

<sup>8</sup>In his discussion, Birnbacher (1995, p. 29) comes to the conclusion that the differentiation between action and omission as means to a normative differentiation remains uninteresting "as long as no *fundamentum in re* can be given for the differentiation between actions and omissions." But this theory, which I share, does not imply that the differentiation provides such a normative differentiation even if such a foundation can be found.

clearly excludes irrational assumptions or negligent actions. This makes it possible to free the discussions on the ethical permissibility of euthanasia of those drastic cases which are ethically unacceptable for reasons of their irrationality or the negligence of those concerned. Thus it is possible to understand the class of euthanasia acts in such a way that whether such actions can be justified or not is a substantial ethical question.

In the following I will support the thesis that on the basis of the conception of biographical identity developed here, an argument *for* the ethical permissibility of voluntary euthanasia can be developed (Sect. 6.2). Subsequently I will look into a fundamental criticism of the ethical permissibility of voluntary euthanasia, which tries to formulate *categorical* objections on the basis of a doctrine of the “sanctity of life” (Sect. 6.3). After that, *gradualistic* objections will be discussed, according to which the ethical permissibility of euthanasia granted *prima facie* is in the end untenable (Sect. 6.4). In conclusion, I will discuss the question of whether the entitlement to voluntary euthanasia which I support also embraces a corresponding “duty to kill” (Sect. 6.5).

## 6.2 The Value of Life, the Personality and the ‘Right’ to Die

In this section, the argument which arises for the ethical permissibility of voluntary euthanasia from the conception of personal identity unfolded in this study will first be amplified (Sect. 6.2.1). Then I will respond to John Harris’s theory, to which my approach shows similarities (Sect. 6.2.2). Remarks on the relation of personal autonomy and voluntary euthanasia bring this section to a close (Sect. 6.2.3).

### 6.2.1 The Argument

The biographical identity of a person which is manifested in her personality, though it rests on the persistence of the purely biologically understood human individual, is itself of an evaluative nature. Persons constitute their personality through – both positive and negative – *identification* with their experienced past and anticipated future.<sup>9</sup> From the evaluative constitution of personality, which has already been analyzed in this study, emerges a direct argument for the *prima facie* permissibility of voluntary euthanasia. Because a personality is constituted through the evaluative *identification with*, the evaluation of one’s own life is inscribed in one’s personal

---

<sup>9</sup>Stoffell (1998, p. 275) says that this biographical identity leads to an individual view of the world, to a “personal scheme of things”. This scheme can best be understood “in terms of the conscientious commitments that structure a life and allow a person to express an authentic way of life”. It is this personal view of things that we must make the starting bases of our evaluation of a person’s “end-of-life decisions”.

life. This evaluation occurs in a social space and must satisfy certain conditions of coherence.<sup>10</sup> It is at the same time the articulation of a person's individual ideals and in this sense the expression of her personal autonomy. As such it deserves not only the respect which autonomous decisions are meanwhile granted indisputably in our culture, but over and above this it is itself the realization of an ideal that is highly esteemed in our culture: living one's own life, developing one's own personality, expressing one's own individuality. One of the essential conditions for personality is knowledge of one's own existence. This is why persons are not only able to weigh individual aspects of their own future biography, but also to subject this future biography itself to an evaluation. This opens up the possibility of accepting or rejecting this future biography as a whole (with the open options it includes). So, from the evaluative constitution of persons' biographical identity follow not only the possibility of refusing certain aspects of a possible future development and giving preference to others. This happens when, for example, a person refuses a painful operation and chooses a gentler procedure. It follows also the possibility of rejecting the continuation of one's own entire existence and expressing the want to die, to commit suicide or to be killed. Such a want can be perfectly rational and the expression of an autonomous decision; and it can, from the stance of the interpreter, actually match this personality exactly. If such a wish is not irrational, but a comprehensible autonomous decision that is coherent with the personality, then this decision deserves our ethical respect.<sup>11</sup>

### 6.2.2 *The Value of Life According to John Harris*

Starting from the Lockean conception of personal identity, John Harris develops a theory of person which is intended to answer two questions simultaneously: one is the question of what differentiates persons from other entities, and the other is the question of how persons differ from other entities as regards their moral relevance (Harris 1985, p. 18). He suggests that the capacities which Locke considers necessary for personhood are at the same time those capacities for reasons of which an

---

<sup>10</sup>This does not mean that a personality must factually comprise a consistent set of beliefs and values. On the contrary, a person's personal life must be interpretable as the attempt to develop a coherent unity. Such a coherent unity, that includes not only the rational, but also the emotional aspects of a human being, need not be understood as an explicitly formulated 'life plan', on the basis of which a person leads her life. The conception of biographical identity demands no such measure of explicitness, but that a person's individual decisions can be understood in such a way that a personality is articulated therein (cf. also Schechtman 1996, Chap. 5).

<sup>11</sup>This applies at least *prima facie*. When it transpires after all that such a decision is irrational (is based on false assumptions or is the expression of emotional confusion), then it loses its claim of validity. Since the outline of life manifested in a person's biography need not be explicit or reflected, in borderline cases it can be appropriate to override a person's autonomous decision – at least for some time – in the name of her own personality (cf. Chap. 8 on this). Moreover, it can happen that such an autonomous decision must be overridden because of other ethically relevant aspects (cf. Sect. 6.5).

entity is able to evaluate her own existence (Harris 1999, p. 303 f.). Because persons behave judgmentally towards their own existence, this existence itself is of value to them (Harris 1985, p. 16). In contrast to merely sentient creatures that can only be harmed by the infliction of pain on them, persons can in addition be wronged if one either takes their life, which they value, or forces them to go on living their life which they wish to terminate (ibid. p. 79 f.). Harris defines a person as “*a creature capable of valuing its own existence*” (1999, p. 307), whereas this valuation does not actually have to be implemented (“capable”) and can be both positive and negative (Harris 1985, p. 17). So persons have a special moral status, because they value their life themselves and we must take these valuations into account. According to Harris, these valuations and decisions express the individual autonomy of a person who defines both who she wants to be and – through her actions – influences how the world looks: “The chooser is a world maker and a character builder at the same time” (Harris 1996b, p. 39). Such decisions, which according to Harris also provide the basis for the valuations of one’s own existence, are the expression of a person’s autonomy and therefore deserve respect. Harris considers it to be an advantage of his theory, that in it the value of a person’s life does not consist in which concrete values she realizes in her biography, or what things she decides on for whatever convictions or valuations (Harris 1985, p. 16). His criterion takes up the higher level of the capacity for such evaluations and can thus remain formal. Hence he also sums up his positions thus: “The value of someone’s life is primarily a value *to them*” (ibid. p. 80). Since this special moral value of personal life depends on the individual autonomous decisions of the respective individuals, a right to voluntary euthanasia also follows from the respect for this autonomy, as the valuation of a person which is to be respected is also expressed in this. Hence Harris (1996a, p. 10) supports, to quote him, a “liberal view of euthanasia” which rests on the respect of persons and their individual autonomous decisions respectively.

This theory of Harris’s shows extensive commonalities with the conception of biographical identity that I have developed.<sup>12</sup> The most important differences which I would like to mention briefly at this point have, as regards the ethical valuation of voluntary euthanasia, no immediate ramifications as far as I can see. In fact they pertain to general philosophical premises and to the significance of autonomy in the framework of biomedical ethics.

- (a) For instance, Harris interprets the loss of personality as the end of the existence of a person (Harris 1996b, p. 41) and not as loss of status, thus landing himself with Finnis’s (1996a, p. 31 f.) objection of dualism, which is justified as long as the human person and the human organism are understood as two persisting entities with differing spatiotemporal extensions. Harris’s rejoinders (1996b, p. 41 f.) to this objection show that he does not intend and would like to avoid such dualistic consequences. However, his own presentation, contrary to the

---

<sup>12</sup>Harris (1999, p. 305) himself even speaks once of the existence of a person in the “biographical” sense.

approach suggested in this study, does not allow him any satisfactory answer to the objection of dualism.

- (b) A second difference is that in my approach the evaluative nature of a person's biographical identity is not limited to the ability to value existence, but rather constitutes the identity of persons qua personality. Whilst Harris locates persons' features in the ability to evaluate their own existence and thus makes a contribution to the concept of personhood, I have tried in this study to establish the evaluative *identification* of a person *with* her own existence as a basic structure of the biographical identity and personality respectively. This leads to the concept of individuality and personality being defined more comprehensively than by Harris. And it opens up the possibility of developing a principle of autonomy that is not purely formal, which is closely tied to the concept of the biographical identity of persons.
- (c) The third difference concerns metaethical assumptions about values. Although Harris says little on this problem, elsewhere he does speak of the "intrinsic importance" of the value of human life (Harris 1996a, p. 8). As his dispute with John Finnes, an exponent of the doctrine of the sanctity of life, and his discussion of Ronald Dworkin's conception (1993) suggests, Harris assumes that values depend on subjective achievements. In contrast, my approach bases on a wide conception of evaluative relations (cf. Quante 2004a): but neither can they be reduced to consciousness or even self-consciousness.
- (d) The fourth difference between Harris's approach and my own deliberations is of an ethical nature. Admittedly, on this point his arguments are not completely clear. It is unarguable that for Harris the value of personal life is of greater dignity than, for example, the value of merely sentient life (cf. Harris 1985, p. 8; 1996a, p. 18 f. and 1996b, p. 44). It is, however, unclear whether this concerns a categorical difference in the sense that the higher value can on no account be weighed up against the 'lower' values, or whether Harris does all in all represent a continuous gradualistic conception which permits general consideration.<sup>13</sup> Because of this unclarity, there can be no decision as to whether there really is a difference between the assumptions of Harris and myself on this point. At all events, as already discussed in the discussion on an alleged 'right' to live (cf. Chap. 3), I dispute that there are ethical rights which are categorically divorced from one another.<sup>14</sup>
- (e) The fifth and final difference I would like to mention here is that, in contrast to Harris, it is in a twofold sense that I lay out a not purely subjective or individual conception of the value of personal life. This difference arises from the

---

<sup>13</sup>The subordination of "critical interests" in the case of a PVS-patient, which are documented in his advance directive, in favor of those actual persons who value their life, suggests the reading that Harris has a categorical difference in mind (cf. Harris 1996a, p. 18 f.). His talk of the 'right' to live also points in this direction.

<sup>14</sup>But as I neither assume that there is precisely on kind of ethical or non-ethical requirements or goods, nor support the thesis that weighing up with purely quantitative procedures can be carried out, the theory of ethical evaluation I propose is not a variant of utilitarianism; cf. for more detail on this Quante and Vieth (2002).

differences (b) and (c) which have already been mentioned. Because firstly, the personality is constituted in the context of social processes of recognition, intersubjective evaluations enter into the value of life of every person.<sup>15</sup> This value does not limit itself, as Harris argues (1996a, p. 11), to “the value we give to our lives”, but also always contains a dimension of intersubjective recognition and plausibility. Since, secondly, I do not assume that values can be reduced just to individual achievements of self-consciousness or qualities of perceptiveness, valuations implicitly or explicitly contain a realist assumption: not all valuations are meant only as “valuable for me”; some are, in the sense of Charles Taylor’s strong valuations, also meant as “valuable as such”. Since I assume the existence of evaluative relations that are not reducible to subjectivity achievements, I can take this realist dimension of our attitudes to values and experiences seriously (cf. Quante 2013c, Chapter VI and Vieth and Quante 2010). Thus my conception comes nearer to that of Ronald Dworkin (1993) who, via the recognition of values that are not limited to purely individual autonomy, likewise attempts to retain a few of those intuitions that stem from the doctrine of the sanctity of life beyond its theological origins.<sup>16</sup> Such an approximation does not mean that persons’ individual decisions may be automatically overridden in favor of absolute values, in particular when they satisfy the formal requirements of autonomy. But it opens up the possibility of extending a purely formal conception of individual autonomy by a conception of the good and a not purely subjectivistic concept of successful life. In this way, a not inconsiderable part of the ethical intuitions based in our culture on the profoundly anchored doctrine of the sanctity of human life can be retained. Furthermore, my arguments on personality and the connection between biographical identity and autonomy show that such an extension of the framework of ethical evaluation is not simply a ‘hostile’ factor confronting the respective person, but rather, the essential and constitutive frame in which valuations and the development of an own personality become possible in the first place.

### 6.2.3 *Voluntary Euthanasia and Personal Autonomy*

From the principle of biographical identity, the perspective on the problem of euthanasia includes only voluntary euthanasia and focuses on the autonomous decision of the patients who express the wish to be killed. Thus the wide area of nonvoluntary euthanasia must remain undebated. Nonvoluntary euthanasia affects most severely

---

<sup>15</sup>I have discussed this in Quante (2013d) using our ascriptions of responsibility as my starting point.

<sup>16</sup>Consistently, Harris (1996a, p. 14 ff.) has no room for this realist dimension of valuations and accordingly also tries to dismantle Dworkin’s middle position in the choice between respect for personal autonomy and sanctity of life, whereby he just understands the former in the sense of a formal and individualistic conception of autonomy.

deformed newborns and such human beings who are not capable of autonomous decisions during their entire existence. Inasmuch as no relevant advance directive is available, it also affects those human beings who, as erstwhile persons, have irreversibly lost their capacity for autonomous decisions. That this question is not being dealt with here is solely due to the inquiry of my study and not perhaps the unfounded optimism that one could limit the whole problem to the least controversial case of voluntary euthanasia. This is not the case. Not dealing with it is, above, all *not* a phrasing of the thesis that nonvoluntary euthanasia is ethically wrong in every case (cf. Quante 2014b, Chapter VIII). The explicit rejection of this assumption will become even more important below (cf. Sect. 6.4). Clarification of which and whose evaluation benchmarks are to be applied is crucial to an apposite discussion of both voluntary and nonvoluntary euthanasia. Since I have just argued for a concept of autonomy that considers an intersubjective and realist dimension to be indispensable, implicitly the first modules for dealing with nonvoluntary euthanasia are provided, even when in the following the primary concern will be the rejection of objections to the ethical permissibility of voluntary euthanasia. Before I grapple with the categorical objections in the next section, first allow me to discuss two other objections which aim at a putative contradiction between autonomy and the wish to be killed by another person. For instance, Anselm W. Müller, who considers this problem to be a “significant drawback” of the liberal conception, poses the following question: “Why should *self-determination* actually come into its own in that one gets oneself killed by *someone else*?” (Müller 1997, p. 150 f.)<sup>17</sup> In R. Harri Wettstein, who himself considers voluntary euthanasia to be ethically permissible, the following statement can be found:

It would be too simple if the dying person could demand active euthanasia as long as he can still discern assisted suicide as an alternative action. This statement might sound cynical, but it contains an eminently ethical concern. For with the most suppliant request of the doctor to be ‘released’ through a syringe, the decisive action is being shifted off to a different actor and the (possibly religious) conscience of the suppliant is being eased. (Footnote omitted, M.Q.) The danger of such a shifting of responsibility, with which the compliance is left entirely to the physician, must be vetoed categorically. Otherwise active euthanasia cannot become ethically ripe for decision. (Wettstein 1995, p. 84 f.)

First of all, it is important to differentiate between the direction of attack of these two objections. With his objection, Müller is trying to generally weaken the justification of voluntary euthanasia based on respect for autonomy. Against this strategy it must be said that the decision to ask to be killed is still the expression of an individual and autonomous decision which *prima facie* has to be respected. On the other hand, the correct insight must be recognized, that the more the active role remains with the party who expresses the wish to die, the better autonomy can be protected and thus deserves increased respect. This is also the starting point for Wettstein. For

---

<sup>17</sup> Müller considers this to be the less principle objection to the ethical permissibility of voluntary euthanasia, because even on the basis of the doctrine of the sanctity of human life he contradicts that autonomy includes the right to value his life negatively. However, with this second objection he claims to be formulating an objection with which every defense of voluntary euthanasia based on autonomy has to grapple. One has to agree with him on the latter.

him, voluntary euthanasia can only be ethically legitimate when the other choices (suicide, physician assisted suicide) are not to be had. The advantage of this concept is that relieving the patient in question of the decision and the associated ethical evaluation can thus be avoided. Another advantage is that the ethical evaluation of the physician involved in the assisted suicide possibly turns out differently toward his own action from in the case of a direct homicidal action (at least in the area of medical action). And a third advantage is that the ethical evaluation of third parties, insofar as they do not categorically reject homicidal actions (at least in the area of medical action), will be all the more consenting, the more active the patient is himself involved. This is because the patient's autonomy is then also brought to bear more and is respected by us.<sup>18</sup>

The strategy I have adopted in this chapter, of speaking of voluntary euthanasia in a broad sense which also includes assisted suicide, is certainly compatible with this concession, because it is based solely on the premise that there is no categorical ethical difference between voluntary euthanasia in the narrower sense and assisted suicide. However, the assumption of gradual ethical differences is nonetheless reconcilable herewith. Which option is the right one, and when a patient can be attributed with the choice of action broached by Wettstein, that of carrying out assisted suicide will depend on the respective situation. The intuitive ranking we undertake here must at all events count in general as an argument *for* an approach to the problem of voluntary euthanasia based on autonomy. This consequence of the conception of biographical identity presented here thus provides a *prima facie* argument for the permissibility of voluntary euthanasia. With this, my deliberations are in direct opposition to a position known as the doctrine of the sanctity of life, which opines the theory that voluntary euthanasia is *categorically* ethically wrong, whereby I will proceed by differentiating between categorical, intrinsic and extrinsic justifications. Categorical justifications claim to be able to show that a specific type of action is at all events ethically wrong (or demanded). I define as intrinsic justifications those non-categorical arguments which draw only on the features of the type of situation (type of action and its participants). I understand extrinsic justifications as those non-categorical arguments in which the effects on other individuals (e.g. relatives), other sections of the population (e.g. the disabled), social institutions (e.g. the medical sector), or consequences concerning all society are declared to be ethically relevant.

---

<sup>18</sup>Two other lines of argument must be differentiated from the justification of the thesis built on autonomy, that assisted suicide is ethically preferable to voluntary killing by another person. Firstly, Margaret P. Battin (1994, Chap. 12), considers that the question of what is ethically preferable can also depend on historical and cultural singularities; cf. on this also the detailed portrayal in Gordijn (2000). Sisella Bok (1998, p. 112 ff.) gives preference to physician-assisted suicide secondly on the grounds of the possible abuse of voluntary euthanasia in the narrower sense (cf. also Sect. 6.4 on this).



### 6.3 The Categorical Objection

In this section, first a – twofold limited – categorical objection will be presented which rests on the doctrine of the sanctity of life (Sect. 6.3.1). Following that there will be discussion on whether it is possible to develop an acceptable ethical stance on this basis (Sect. 6.3.2). Finally, the issue will be to weigh the conception I proposed in the last section against the position under discussion here (Sect. 6.3.3).

#### 6.3.1 *The Objection*

One of the prominent, if not always clearly articulated positions in current bioethical altercations is the doctrine of the sanctity of life (cf. Kuhse 1987, chap. 1 and Dworkin 1993). In its generally represented form, e.g. by Albert Schweitzer, it refers to all forms of life. In the context of biomedical ethics, however, a twofold limited conception is effective.

- (a) So firstly it is a matter of human life, whose inviolability, absolute value or dignity is asserted.<sup>19</sup> The basic assumption is that human life has a value dimension that is in principle detracted from any human value judgment. Every human being has to bear his own life, regardless of the physical or psychological condition in which he finds himself, as long as its abandonment is not requisite for others. Nobody has the right to regulate human life, be it because human life represents an immeasurable, absolute value; be it that it is interpreted as a duty imposed by God or a gift of God.<sup>20</sup> In the face of this absolute value any attempt to introduce ethical distinctions on the basis of human valuations must remain arbitrary and highhanded and inevitably lead to catastrophes (cf. for this argumentation e.g. Gormally 1995, also Finnis 1996a, b). Even if the theological origins of this position are highly visible and the attempts to make them the

---

<sup>19</sup> Thus the doctrine of the sanctity of human life is closely related to the - above all in the German discussion - central concept of human dignity. But even the latter has its inner tensions and ambiguities; cf. also Bayertz (1996) and Birnbacher (1996). A formal concept of human dignity based exclusively on individual autonomy can, in the area of euthanasia, certainly lead to different justifications from those of the doctrine of the sanctity of human life. A concept of human dignity that draws on the species or is theologically charged, which avoids this conflict, cannot be made generally binding to the same degree as an understanding of human dignity based on the autonomous individual; cf. also Gutmann (2008). I cannot address questions of animal ethics which arise from the first limitation and will be dealt with under the keyword speciesism objection, cf. my analysis in Quante (2001b).

<sup>20</sup> Dworkin (1993) suggested that one should not regard human life as an absolute value, though an intrinsic one. With this he means that a human life can also be attributed value independently of the individual value assessment of the person, whose life it concerns. But in contrast to the perception of sanctity, or – the secular philosophical equivalent – the thesis of the absolute value an intrinsic value does permit weighing. For reasons of this modification, Dworkin's position is not only compatible with the one I take, but also to a great extent congruent.

basis of generally binding regulations hence problematic in a plural society, the attitude of many people and the course of the current debate show that these assumptions are indeed firmly anchored, at least in West European culture. But between the supporters of the doctrine of the sanctity of human life and those defenders of an ethics of the quality of life based on personal autonomy, such as John Harris (1985) or Helga Kuhse (1987, Chap. 5) there is a fundamental difference in what is understood as a right to live. Whilst for the latter it represents an individual right to defense against others, which, however, at the same time can by all means be surrendered by the respective individual, the former understand it as an absolute and undisposable right which may not even be surrendered by the autonomous individual himself.<sup>21</sup>

- (b) The doctrine of the sanctity of life is not only modified by limitation to human life, this throwing up problems I shall not expound further here. Beyond that, it is also subject to a second limitation (cf. Kuhse 1987, Chap. 1). Not every form of voluntary euthanasia is rejected, only voluntary euthanasia in the narrower sense. The supporters of the doctrine of the sanctity of life understand this as the *deliberate* killing of a human being who has autonomously asked to be killed. With this the assumption is made that between the deliberate killing and allowing to die or the merely accepted death of a patient there is a *categorical* ethical difference (cf. e.g. Finnis 1995a, b).<sup>22</sup> In order to understand this second modification of the doctrine of the sanctity of human life and to be able to question its plausibility, it is necessary to deal with the distinctions active vs. passive, doing vs. omitting and intending vs. accepting.

### 6.3.2 *Subtle Distinctions?*

Alongside the two distinctions ‘doing vs. omitting’ and ‘intending vs. accepting’, the opposition of ‘active vs. passive’ is above all prominent in the debate over euthanasia. But despite the familiar ethical co-ordination of active, intending and doing on the one hand and passive, accepting and omitting on the other, the differentiation between active and passive euthanasia is still neither congruent with that of doing and omitting nor with that of intending and accepting. Since, furthermore, the differentiation between active and passive behavior which is useful in *lebensweltlichen* contexts cannot, despite the common prejudice to the contrary, be sufficiently

---

<sup>21</sup> Müller (1997, p. 144 ff.) unrestrainedly supports the thesis that the surrender of the right to live, as it occurs in the case of the plea for voluntary euthanasia in the narrow sense, constitutes “impermissible permission”. In general on the difference between the two concepts of the right to live cf. also Taylor (1992, p. 11 f.).

<sup>22</sup> So the following section (Sect. 6.3.2) concerns the attempt to justify a categorical ethical difference between deliberate killings and other forms of voluntary euthanasia. Justifications of the differentiations between active and passive or killing and allowing to die which want to render the *gradual* ethical differences plausible will be discussed in the fourth section. Such a strategy is pursued e.g., if in a different way, by Sissela Bok (1998) and Tom L. Beauchamp (1978).

selective for ethical purposes and is increasingly losing its anchorage within the context of a medicine that is increasingly based on action with the deployment of machines, it is sensible to rely primarily on the other two distinctions. It will thereby be shown that – at least under the auspices of the attempt to justify a categorical ethical difference – these two differentiations are not independent of one another either from the descriptive or the ethical standpoint. Moreover, the analysis of the differentiation between intending vs. accepting also opens up a possibility of maintaining an ethically significant sense of the opposition of active vs. passive.

The differentiation between doing and omitting, or rather, the philosophical analysis thereof, comes under the thematic area of the philosophy of action, a philosophical discipline that was above all in the 50s, 60s and 70s of the last century, a central research object in analytic philosophy. Two main strategies can be distinguished: *intentionalists* in the philosophy of action, mostly departing from Ludwig Wittgenstein, support the theory that one cannot explain actions like natural occurrences as the causal consequences of events, but should interpret action in terms of social practice. According to the intentionalists, concrete actions are made understandable through recourse to types of action and general codes of behavior. *Causalists* in the philosophy of action react to this conception by saying that explanations of actions are a subspecies of causal explanations. They claim that one explains the occurrence of an action causally as a concrete spatiotemporal event by means of specification of the primary reason for its occurrence.

The supporters of the doctrine of the sanctity of human life who support the thesis that the killing of a human being is categorically wrong, whereas that does not apply to allowing somebody to die, try to show that there is such a categorical ethical difference. If this target of proof is to be justified without a *petitio principii* with the assistance of the distinction between doing and omitting, then this distinction must itself have a “*fundamentum in re*” (Birnbacher 1995, p. 29). It must be able to be rendered plausible independently of ethical aspects. As in my elucidation of the first condition in the definition of the concept of euthanasia and also in the framework of my analysis of persistence, I assume that this sustainability is only present when a difference can be found on the level of causal relations that is independent of description and interpretation. Thus an intentionalist in the philosophy of action cannot help us along at this point, since in this conception no reference to causal relations is actually acknowledged. The only allocation undertaken by intentionalists to a given social practice is unsuitable for yielding a neutral basis for an ethical distinction. Without the sole recourse to causal relations independent of description and interpretation there is a risk that the description of actions by two observers will be adapted to their own preset theoretical views and ethical assumptions.<sup>23</sup>

---

<sup>23</sup>Thus, in the first step of his argumentation, Reichenbach (1987) developed a perspectivistic analysis, i.e. dependent of description, of the difference between doing and omitting and in the second step comes to the conclusion that this difference possesses no ethical sustainability, but is to be understood as an alibi or avoidance strategy. Conversely, John M. Freeman and Edmund D. Pellegrino, who vary considerably as regards their metaphysical and ethical premises, have themselves to their astonishment established that they mostly agreed as far as their concrete views on what should be done are concerned: “Despite these sharp differences in ethical justification, we

The obvious attempt to award doings but not omissions causal effectiveness is doomed to failure. The initiated absence of a consequence  $f$  due to an omission  $u$  can be a marginal causal condition for a temporally later event  $x$ . I didn't put the bathroom heater on because I thought it was warm enough ( $= u$ ). The unheated bathroom ( $= f$ ) is one of the causally relevant marginal conditions for my later catching a cold while bathing ( $= x$ ). This example shows: Omissions are also spatiotemporal events and causally relevant; differences arise at most as regards the probability with which the effect occurs (cf. Birnbacher 1995, Chap. 3).

If the difference between doing and omitting cannot be grasped through the distinction of causally relevant vs. causally irrelevant, then the question is how this difference can be grasped at all. How does an omission become one? If I drive past an unattended accident victim lying at the roadside, then that is not automatically an omission. Everything that follows depends on whether I have noticed the victim.<sup>24</sup> To avoid the dependence on interpretation of the classification of an event as doing or omitting, an actual feature of the event in question must be named, which makes the difference. The most plausible candidate for this roll is the following<sup>25</sup>: An action<sup>26</sup> is an omission if and only if it is either *caused* by the *intention* not to perform a certain act type  $A$  in this situation,<sup>27</sup> or it is performed in the *knowledge* that in this way the type of action  $A$  is not performed in this situation (and maybe later on can no longer be executed in this situation). According to this thesis, even the realist interpretation of the distinction between doing and omitting requires recourse to the reasons for action and thus implicitly refers back to the differentiation between intending and accepting. Analysis of this relation very soon shows that only the latter offers any chance at all of justifying the sought categorical ethical difference.

---

were struck by our agreement on what ought to be done in certain specific clinical instances" (Freeman and Pellegrino 1996, p. 187). This result ensues when the classification of an action as doing or omitting resp. active or passive is adapted to the own ethical intuitions. To avoid this - and the doctrine of the sanctity of life under discussion here must avoid this consequence at all costs - the distinctions require a base that is independent of interpretation.

<sup>24</sup> If I didn't see the accident victim, then my continuation of my journey was not an omission of "first aid". That does not preclude that one can accuse me of carelessness because I did not notice the victim, although one could expect that of me according to normal "human discretion". Not noticing a plainly visible accident victim e.g. because of juggling with the radio stereo or telephoning with a cell phone is careless, but is not omitting to notice the accident victim.

<sup>25</sup> As a terminological specification the following applies according to Davidson (1982, Chap. 1) The *primary reason for action* means the factually causally effective reason for an action. Such a reason for action always reveals two aspects. The evaluative-volitional aspect I call the *intention* and the cognitive aspect, the *conviction for action*: In reality the two aspects will always embrace several intentions and beliefs. In the following I assume the simplification that the prevailing intention and the relevant or explicit beliefs can be identified and depicted as conjunction. Further clarifications come under the task area of the theory of agency and cannot be dealt with here.

<sup>26</sup> Through the precondition that an action is present, I am not making the problem for a causalist theory of agency of wayward causal chains the subject of discussion, as I am not concerned here with the defense of a causal theory of agency.

<sup>27</sup> This part of the concept should be taken cum grano salis since strictly speaking the same situation does not occur twice. In the context of the discussion on voluntary euthanasia it means that one does not carry out certain measures on certain patients and cannot do so later, either.

So it is reasonable to look for categorical difference in the area of primary reasons for action. If we limit this area to the context of voluntary euthanasia, then it concerns the death of a patient P that is to be brought about deliberately through my action A (either doing or omitting), or is known to be the accepted feature or consequence of my action. So four relevant kinds of reasons for action arise, whereby for one thing it is assumed that the primary reason for action does not reveal any further ethically impermissible aspects, and for another (iii) should also include the cases in which A brings about P's death faster:

- (i) I intend with A to cause the death of P.
- (ii) I intend with A not to prevent the death of P.
- (iii) I know that A means that the death of P can no longer be prevented.
- (iv) I know that A *does not prevent* the death of P.

(Four types of reason for action)

On the basis of these four aspects of reasons for actions, both the distinctions doing vs. omitting and intending vs. accepting and – at least in a specific interpretation – the distinction active vs. passive can be rendered plausible as descriptive distinctions. This becomes obvious when one formulates three ethical positions that each with the help of one of these distinctions can express the categorical ethical difference sought by the supporters of the doctrine of the sanctity of human life. Here, position I uses the distinction doing vs. omitting, position II the distinction intending vs. accepting and position III the distinction active vs. passive (in a specific reading):

Primary reasons for action which contain the intentions of Type (i) are categorically ethically wrong, whilst primary reasons for action which contain intentions of Type (ii) or convictions for actions of Type (iii) or (iv) can be ethically acceptable where appropriate.

(Position I)

Primary reasons for action which contain intentions of Type (i) or (ii) are categorically ethically wrong, whilst primary reasons for action which contain convictions for action of Type (iii) or (iv) can be ethically acceptable where appropriate.

(Position II)

Primary reasons for action which contain intentions of Type (i) or convictions for action Type (iii), are categorically ethically wrong, whilst primary reasons for action which contain intentions of Type (ii) or beliefs of Type (iv) can be ethically acceptable where appropriate.

(Position III)

A diagrammatic view results in the following chart of the three ethical options<sup>28</sup>:

---

<sup>28</sup>If no sustainable distinctions can be found, then there remains either the option of assuming there is no categorical difference (the liberal position), or rejecting all four aspects of primary reasons for action mentioned as categorically ethically wrong. As far as I can see the other positions possible through combination will not enter into the discussion.

Options	Categorically ethically wrong	Not categorically ethically wrong	Corresponding intuitive distinction
Position I	(i)	(ii) & (iii) & (iv)	Doing versus omitting
Position II	(i) & (ii)	(iii) & (iv)	Intending versus accepting
Position III	(i) & (iii)	(ii) & (iv)	Active versus passive

(Three ethical options)

What should one make of these three positions? *Position I*, which attempts to base its categorical ethical difference on the distinction doing vs. omitting, proves on closer inspection to be untenable. If no intuitions come along which owe themselves to the other two differentiations, the intentional causation of death through doing will be assessed as categorically ethically wrong, whereas the intentional causation through omitting and the two forms of accepting will possibly be regarded as ethically justifiable kinds of agency. However, this is not plausible, at least as justification of a categorical distinction. For the specific difference rests either on the value of the intention, which, however, causes the first position to dissolve into one of the other two positions, or on the property through which doing and omitting are differentiated objectively as the causal contribution in the causation of death. But this is only the different degree of probability and this difference is at most suitable for justifying gradual ethical differences. In this way, categorical ethical differences remain out of reach.

If the defender of the doctrine of the sanctity of human life does not refer to the objective causal contribution but to the causal role the agent ascribes to himself, then *Position III* ensues. Whoever executes an action and attributes to himself the intention (i) or the belief (iii) understands his action as a decisive<sup>29</sup> causal contribution leading to the patient's death. In contrast, the agent regards the omission based on intention (ii) and the action accompanying the reason for action (iv) merely as omitting a possible intervention in a process of dying, without the possibility of such an intervention being thwarted in general. This reconstruction of the distinction active vs. passive, which is supported by the inner perspective of the agent, has some plausibility. Above all it permits the comprehension of possibly grave psychological differences of both the participants and – through the acceptance of perspective – bystanders, which e.g. are factually reflected in the various assessments of voluntary euthanasia and assisted suicide. Since the value of an action cannot, however, be reduced to the inner perspective of the agent, this difference is itself not sufficient to justify a categorical ethical difference. Moreover, a philosophical

<sup>29</sup> Since at this point it concerns the subjective perspective of the agent of his causal contribution, the explanatory pragmatic dimension contained in this formulation is compatible with my realist position as regards causal relations.

analysis and clarification of the ethical intuitions and principles on which these inner perspectives are based can also be conducive to changing this attitude. For it is obvious that the distinction active vs. passive rests on the following two assumptions: firstly, on the principle of common sense, that doing is generally graver than omitting, and secondly, on the assumption that the degree of causal involvement determines the degree of ethical accountability. As long as the first assumption is understood as rule of thumb, and as long as one only understands the second assumption as the naming of one of the factors for the ethical accountability for an action, but not the sole indicator, this position makes sense. However, it is even then only suitable for justifying gradual ethical differences. The aim of spotting a categorical ethical difference does not thusly converge on the partisans of the doctrine of the sanctity of human life.

The third relevant<sup>30</sup> option rests on the distinction intending vs. accepting. Representatives of *Position II* see the ethical boundary between intentions and convictions for action getting lost. The differentiation drawn on is plausibly describable due to the differences between evaluative-volitional and cognitive aspects of mental conditions. Insofar, this position has a descriptive base. Furthermore, there is also as regards the skeptical objection, that one cannot determine whether an aspect or a consequence of my action is intentional or merely accepted, in principle a test. If it is a matter of a merely accepted consequence, then the agent must prefer, or at least regard as on a par, a counterfactual choice of action, in which – under otherwise the same conditions – only this aspect or only this consequence respectively is not present. But a great disadvantage of Position II is firstly, that an ethical difference weighted in this way is tied solely to the primary reasons for action. Even if one is prepared to accept the accompanying epistemic uncertainty, this does represent an unsatisfactory constriction of the ethically relevant aspects of actions. But secondly, one must above all doubt that an ethically sustainable justification can even be given with this differentiation. For both evaluative-volitional and cognitive aspects are a part of every action and every primary reason. Every reason for action embraces wanting and knowing in equal measure. What is more: It is an ethically and judicially firmly established concept that is also intrinsic to our everyday practice of evaluation, that the ethical accountability for aspects and consequences of action cannot be limited to the evaluative-volitional area. Even merely anticipated and not only positively desired aspects and consequences must be accounted for, of possibly to a different degree. Freed of the two assumptions on which the first and third positions base, Position II can thus no longer invoke common sense. The following argument, which is frequently introduced to support this position, is also unsustainable: Surely not every unintended aspect and not every unintended consequence of an action is apportioned ethically or judicially, despite acceptance. This statement is trivial and true, but does not hit the nail on the head. For the counter thesis to Position II is not that: every accepted or anticipated consequence of an aspect and

---

<sup>30</sup>Further options (e.g. only (iv) not to be regarded as categorically ethically wrong) arise through various combinations of the three differences. However, they inherit the difficulties of the three above named positions and so do not have to be discussed specially.

every accepted or anticipated consequence of an action is to be assigned ethically or judicially. On the contrary, it is claimed that the boundary between ethically relevant and non-relevant aspects and consequences includes a subset of acceptance.

Thus the defender of the doctrine of the sanctity of human life must retreat to the thesis that it is not generally the difference between intending and accepting on which his argumentation is based, but is a matter of the specific nature of the intentions. But this consists in undertaking an evaluation of the life of a human being – in which sense has yet to be clarified (cf. Sect. 6.4). The attempt to justify the categorical ethical difference fails. For this reason, it is necessary for the supporters of the doctrine of the sanctity of human life to extend the categorical ban to all four cases. Herewith he gets into a contradiction of the widely accepted ethical practice that is much graver than the opposite conclusion, which consists in assuming merely gradual ethical differences in this area.

### 6.3.3 *Sanctity of Life and Plural Society: First Conclusion*

Thus, in the end everything boils down to the basic assumption that nobody has the right to evaluate human life. Even if this assumption can be made plausible within a theological framework, it cannot be defended as a generally binding ethical norm in a secularized, plural and pluralistic society. The individual subject retains the right to refuse for himself voluntary euthanasia in the narrower sense on the basis of such a religious or philosophical assumption. This has consequences for the question of what duties ensue from the right to voluntary euthanasia. But here it is only a matter of the permissibility of *voluntary* euthanasia and thus of such cases in which an autonomous wish to be killed is expressed. If such a case is not present, because the patient in question excludes such an option ethically, then this is certainly compatible with the line of argument in this chapter. A further disadvantage of the doctrine of the sanctity of human life is moreover that it is difficult to reconcile it with the composition of the biographical identity of persons, if these persons do not regard their own life as a task set by God or as an absolute value.<sup>31</sup> A liberal position which awards persons the right to their own evaluation of their own existence and allows the quality of the life of autonomous persons to depend decisively on this self-evaluation, corresponds much better to or plural and pluralistic culture. And if one thereby avoids sliding into a purely subjective theory of valuations and a purely formal conception of autonomy, one furthermore retains sufficient means to criticize and maybe even in an ethically justifiable way overrule extreme and incomprehensible subjective evaluations of one's own existence.

---

<sup>31</sup>The difficulty a person grasping herself as religious in this way has in at the same time viewing herself as autonomous must remain undiscussed. It is only important to note that such a religious self-interpretation is compatible with the analysis proposed here, since it does not in fact demand that a person takes an evaluative stance towards her existence as a whole.



Up to now, with the deliberations in this section only objections have been dispelled which wanted to show that voluntary euthanasia in the narrower sense is categorically wrong. With this the possibility still remains of showing through consideration of the appreciation of values or through slippery slope arguments that voluntary euthanasia, at least in the narrower sense, is in all ethically unacceptable. Supporters of the strategy of setting up and justifying tables of categorical bans object to this gradualist approach that thus an arbitrary basis of justification is being formed which cannot ultimately be safe. Herewith, “arbitrary” can have various meanings: (a) If, thereby, every justification not revealed by God or proven by final rationale is to be rejected, then this argument is surely unconvincing. The thesis that an ethical justification resting on intersubjective rational agreement cannot in principle suffice to justify and secure a human social practice can apparently only convince somebody who is already convinced. Questions of consideration might be controversial and sometimes not clearly decidable, or even in individual cases lead to unsatisfactory results. But it cannot be deduced from this that only categorical forms of justification provide adequate protection. Such a stance would in the end actually eliminate room for discretion, the ethical power of judgment and hence the autonomy of the participants. (b) “Arbitrary” can also mean that a highhanded differentiation that cannot be fixed plausibly to situations cannot be a stable basis for ethical practice. In this sense the objection is justified. So it all depends on which aspects are to count as ethically relevant and therefore as not arbitrary. Here the recourse to the tenet of double effect seems to me to be highhanded and irreconcilable with established ethical and judicial practice, at least, when with its help a categorical ethical difference is to be justified or claimed so as to cover all ethically relevant aspects with this tenet. The consequence of this objection in the end couches the demand that the area of voluntary euthanasia needs explicit societal explanation, so as to determine and develop a non-arbitrary evaluation basis in intersubjective discourses. For these problems are dealt with really arbitrarily when they are left to or imposed on the sole decision of the individual physician and the practice of voluntary euthanasia is conducted behind a veil of silence.

## 6.4 Gradualist Objections

The classical form of gradualist objections against the permissibility of voluntary euthanasia are slippery slope arguments. On the understanding that the first step brought up for discussion, in our case the permissibility of voluntary euthanasia in the narrower sense, is in itself ethically acceptable, the argumentation against this approval points out the consequences of such an approval which are bound to follow.<sup>32</sup> Since these consequences, in our case the occurrence of nonvoluntary or even

---

<sup>32</sup>This concession, that the first step is *prima facie* ethically acceptable, belongs to the logic of slippery slope arguments, but it can, as one sees in Müller’s case (cp Sect. 6.4.1.2) be conceded just for the sake of argument; cf. also Den Hartogh (1998).

involuntary euthanasia, are clearly viewed as ethically impermissible, it is now inferred that the first step up for discussion may ultimately also not be permitted, although it is *prima facie* ethically justified. Otherwise one would enter a “slippery slope” on which there would be no more stopping.

Such slippery slope arguments can be divided into two variants: logical-causal and purely causal.<sup>33</sup> In the former, the vagueness of terminology and the existence of grey areas lead to the conclusion that one inevitably slides from the ethically acceptable end of the slippery slope into a catastrophe due to psychological mechanisms.<sup>34</sup> In the latter, the same consequence is deduced without recourse to such logical elements. Thus Yale Kamisar (1995, p. 233), to name a prominent representative of this position, deduces from the continuous spectrum of cases the consequence that it will be “extremely difficult” to respect the boundary between assisted suicide and voluntary euthanasia in the narrower sense. And John Keown has several times tried on the current example of The Netherlands to support such slippery slope arguments with empirical evidence (cf. Keown 1995a, b, c).<sup>35</sup>

What do slippery slope arguments accomplish? If one leaves out the purely logical forms which do not basically belong to this type of argument, it should first be captured that they can only be gradualist arguments, since they rest on causal probabilities and assumptions of probability. Factually, in the discussion they frequently serve more as illustrations of fears and, so long as they can make the ingress of the feared consequences plausible, result in a reversal of the burden of proof. But it is questionable how well the named consequences can really be made plausible as probable consequences. The weakness of these arguments cannot be remedied as long as the imputation of causal consequences, whether they be individual-psychological or societal, is not empirically allocatable. And yet, despite this weakness one should not dismiss either the factual effectiveness of these considerations or the reservations expressed in them as just plain irrelevant. The bad consequences mainly brought forth are the following (cf. Brock 1993, p. 218 ff.):

- the ethos of the medical profession is destroyed,
- the willingness of society to provide optimal care of dying patients is weakened,

---

<sup>33</sup>This subdivision can still be refined considerably, cf. on this Walton (1992) and Guckes (1997, Chap. 1). For the purposes of this section, the above dichotomy is however adequate.

<sup>34</sup>Purely logical argumentations, i.e. those who are run without recourse to probable psychological consequences, are therefore not slippery slope arguments and therefore do not have to be of a gradualistic nature, either (cf. Sect. 6.4.1.1). They are, however, if they base on vagueness, also hardly convincing, as the analogy to Sorite’s paradox shows: “That there is a grey area, and that its boundaries are indeterminate themselves, doesn’t imply that we cannot identify points which are definitely at one or the other side of the grey area” (Den Hartogh 1998, p. 283).

<sup>35</sup>But even the availability of empirical data, which is really a welcome and necessary precondition for the controlled application of such slippery slope arguments, does not suffice to decide the question of whether The Netherlands really would slide down a slippery slope to an ethical disaster; for a different interpretation of the empirical findings cf. e.g. Kuhse (1996) and the contributions in Thomasma et al. (1998) on the overall situation in The Netherlands.

- the laboriously asserted right of patients to refuse life-prolonging measures is subverted and endangered,
- the offer of voluntary euthanasia in the narrower sense harms some patients,
- the general ban on killing innocent human beings is undermined.
- (Putative negative consequences)

These five objections are attributed diverging weight: It is for one thing doubtful whether voluntary euthanasia, even in the narrower sense, is really irreconcilable with medical ethos. Not only the medical profession's having for the most part few problems with conducting abortions belies this; above all, it is belied in that the avoidance of suffering belongs equally to the medical mandate as the preservation of human life (cf. the analysis by Gert et al. 1997, Chap. 12). The second and third points can hardly be verified empirically, so that they are of dubious value and it remains doubtful whether one should justify the prolongation of patients' suffering on the basis of such "unsecured checks". The fourth consequence cannot be overlooked: When suffering patients, who for themselves e.g. accept the doctrine of the sanctity of human life, are confronted with the possibility of voluntary euthanasia in the narrower sense, this can lead to additional conflicts of conscience. But from this it can only be deduced that one should not unfold such an offer individually to these patients.

The existence of a general option can, however, also be expected of e.g. religious human beings, if it may not be withheld as a legitimate right in a plural society. The fifth objection is the decisive one for the issue in this section. It is usually clothed in the argument that the admission of voluntary euthanasia in the narrower sense must lead to the introduction of nonvoluntary and even involuntary euthanasia into our social practice. I will now turn my attention to this and other objections directed against the equality of voluntary euthanasia in the narrower sense and assisted suicide.

### **6.4.1 Four Objections**

The following four objections are directed in various ways against theses which have been defended in this chapter. So they need consideration, whereby I would like to take the opportunity of going into more detail about certain aspects of the proposal submitted here.

#### **6.4.1.1 The Objection by Grisez and Boyle**

In their discussion of voluntary euthanasia in the narrower sense, Grisez and Boyle (1979, p. 171 ff.) try to show that one must of necessity proceed from there to non-voluntary euthanasia. With that, they do not believe their argument to be a slippery slope argument, because they wish to draw attention to a logical contiguity (ibid. p. 173). According to their perception, the defenders of voluntary euthanasia in the

narrower sense avail themselves of premises which “logically entail nonvoluntary euthanasia” (ibid.). Not that this were the fault of the clumsy choice of examples, but due to the fact that the defenders of voluntary euthanasia in the narrower sense are building on a principle of quality of life or a “meaningful life” (ibid. p. 172). If one were to follow the project of “beneficent killing” (ibid. p. 173), then in their eyes the following consequence would result:

To stop short of killing all those whom they sincerely believe would be better off dead would be a completely irrational and arbitrary limitation upon the unfolding – according to its inner dynamics – of their well-intentioned project of beneficent killing. (ibid.)

Whilst even as a rational being one might hope for the probable consequences of an action that these undesirable consequences do not occur after all, their alleged link is purely logical and therefore undeniable. For this reason, they assume their considerations also have a force which could not be reached by mere slippery slope arguments. The decisive premise of this argument is thereby:

If one holds that certain sorts of people would be better off dead and would be kindly treated by being killed, it matters little whether these people are competent to consent or not. (ibid.)

The decisive weakness of this argument is that it does not distinguish two kinds of evaluation of quality of life. Whereas in the case of autonomous persons one can comprehendingly reconstruct their self-assessment of their own life (*personality based* standard), for non-autonomous persons such an assessment of the quality of life must be undertaken on the basis of objective criteria (*objective* standard). In both cases it concerns the question of how the quality of life turns out for the respective patient, not an assessment based maybe on scientific or purely social standards (social benefit etc.).

It must be conceded to Grisez and Boyle that in the case of nonvoluntary euthanasia one must introduce such an objective standard. But since it thereby concerns a different standard from the personality based one, these two issues can be kept apart. So logical force is out of the question. Because in the case of voluntary and nonvoluntary euthanasia it is a matter of two different assessment standards, one can take up the position I represent, that the objective standard is only applicable where the personality based standard cannot in principle go through. In other words: When a patient decides autonomously that he wants to continue his life, then the objective standard must not be applied. But even when a patient is in principle<sup>36</sup> capable of deciding autonomously, but does not factually express himself, this objective standard must not be applied.<sup>37</sup> The reasons for this position are for one thing the

<sup>36</sup>Thus such cases in which an emergency intervention has to be undertaken to save life, for which no actual informed consent is available, are not covered. If one can assume that the patient recovers his decision competence after the intervention, then as a rule the anticipated quality of life will be so high that an intervention is justifiable. Besides, should this not be the case, the option of voluntary euthanasia would subsequently still be open. But the possibility of such tragic cases is above all an argument for basing the instrument more strongly on advance directives (cf. Chap. 7).

<sup>37</sup>At the same time, it is a possible reconstruction of Harris’s position, for persons to justify a ‘right’ to live that is not dependent on a categorical difference.

principle of respect for personal autonomy. For another, it is the assumption that every person who does not express himself to the effect that one should end his own life awards his own existence a value which must be acknowledged. This is the reason why I have qualified euthanasia acts without the request by autonomous patients generally as cases of involuntary euthanasia.

#### 6.4.1.2 Müller's Objection

Müller also tries – after he has already put forth categorical objections to the permissibility of voluntary euthanasia in the narrower sense – to prove that the permissibility of voluntary euthanasia must end with the introduction of nonvoluntary and even involuntary euthanasia:

I am *not* concerned here with evidence that the admittance of voluntary euthanasia can only too easily via misuse have the practice of involuntary euthanasia in its wake, but with the following conclusions: 1. If weighing up my own interests against one another can provide a reason for having me killed, then weighing up between them and the interests of *others* can also provide such a reason. 2. This reason can also be identified with by the killer. 3. His predisposition towards this reason can waive the consent of the one killed. (Müller 1997, p. 86 f.)

Because Müller only speaks of “can” here, it is a case of a slippery slope argument, whose individual steps are, admittedly, not a matter of course in the eyes of its originator, either, as one can see from his extensive annotations (cf. *ibid.* p. 87–91).

If one leaves undebated the question of what Müller means exactly by “can” and imputes its strongest possible plausible meaning, then the force of the whole argument depends on the relation of the three steps. What does this look like? To see more clearly here, two further arguments of Müller's must be accepted. *Firstly*, he gives four reasons with which one can justify typical euthanasia acts (*ibid.* p. 63): (i) to promote the wellbeing of the other person; (ii) to comply with the other person's request; (iii) to respect the right to self-determination and (iv) to do something good for someone. In his further analysis Müller then claims *secondly* that (i) is ultimately dominant in all defenses of euthanasia acts (*ibid.* p. 71). This is a decisive step in his deliberations, which is justified as follows: Alone the fulfillment of a request is not sufficient in such a grave case. Therefore (ii) refers either to (i) or the agent appropriates the suppliant's reason. (iii) refers again to (i), as the respect for autonomy is justified in that disregarding an autonomous decision runs contrary to the person's wellbeing.<sup>38</sup> – Müller's argument can be reconstructed on this basis:

1. A's expressed want to be killed can also base on altruistic motives i.e. the wellbeing of third persons.

---

<sup>38</sup>As already discussed, Müller claims that at least in the case of the want to be killed by someone else, this is not a genuine exercise of one's own autonomy. Thus only the aspect of wellbeing is left over of (iii), which consists in the autonomous decision not being overruled. I shall discuss the problematical aspects of this whole figure of argumentation shortly.

2. B, when summoned to the euthanasia act must (2.1.) either orient on A's wellbeing, or (2.2) appropriate A's motive. The autonomous decision itself has no independent ethical value (2.3).
3. If one is geared to wellbeing, then one has a yardstick that is independent of A's expressed wish and thus also allows cases of nonvoluntary euthanasia (follows directly from 2.1)
4. If one appropriates A's altruistic reason, then one accepts the wellbeing of third persons as a legitimate reason for euthanasia acts. (follows from 1 and 2.2)
5. Every defense of euthanasia acts challenges an absolute, non-available value of human life.
6. With this, one is on track to weigh up A's life against the wellbeing of others. (follows from 4 and 5)
7. Since the autonomous decision has no independent value, neither the consent nor the veto by A must be taken into consideration in the weighing up against the wellbeing of third persons. Hence involuntary euthanasia is in principle also permissible. (from 2.3 and 6)

(Müller's Argument)

- (a) Three things must be said about this derivative of nonvoluntary euthanasia (step 3 of the argument). If Müller, like Grisez and Boyle, was of the opinion that every justification of voluntary euthanasia in the narrower sense oriented on wellbeing would also have justified nonvoluntary euthanasia for logical reasons, then in contradiction one would *firstly* have to say that his crossover plainly rests on the lacking differentiation between wellbeing qua the objective standard of assessing quality of life and the value of life qua personality based standard. But since Müller tries to show that the autonomous decision has no intrinsic dignity in this context, *secondly* this assumption of reducibility must be challenged (cf. Sect. 6.2.3). The frustration of an autonomous decision may be harmful to the patient's wellbeing, but it represents an injustice primarily because it injures his personal autonomy. It is erroneous of Müller not to acknowledge this difference. Even when an autonomous decision is ethically wrong it still remains autonomous so that its disregard cannot be justified alone by recourse to the patient's wellbeing. *Thirdly*, one must take Müller's deliberations seriously if one understands them as an indication of a possible psychological consequence. If one considers nonvoluntary euthanasia to be ethically justifiable in some cases, as I do, then one must introduce an objective standard of assessment of quality of life alongside the personality based standard. If this happens, then it is still essential to explicitly point out the difference between the two and, respectively, bring the euthanasia practice into line, as well as specifying their relation to one another. Both of this I have already done in this section (cf. Sect. 6.4.1.1).
- (b) The crossover to *nonvoluntary* euthanasia in Müller's argument reveals two weaknesses and moreover rests on an implausible interpretation. The first weakness is here also the disregard or the lack of distinction between the patient's wellbeing and the value of his existence based on his personality.

Müller's argumentation by and large gives the impression that under the latter he could only recognize such conceptions that award the person an absolute, inalienable value. This is, however, as has already been shown, hardly compatible with the basic disposition of my conception of biographical identity and the moral ideals in our culture (cf. Sect. 6.3). The second weakness consists in Müller's argument appearing constructed, because it must contain the precondition that a patient wants his own death for altruistic motives (the "can" in the first step becomes an "is"). But since I do not challenge either the existence or the ethical permissibility of such altruistic motives in the context of voluntary euthanasia, as long as they are turned into postulated altruism, this construction shall be conceded for the sake of the continuing argument.<sup>39</sup> The real weak point in Müller's argument rests on two implausible aspects of his interpretation of the fact that euthanasia agent B must respect the patient's autonomous decision and make the assessment contained therein "his own".

The first weak point is of a general nature: It is not demanded that for the adoption of the perspective B takes over A's evaluation as factually his own. It is certainly possible for B to arrive at the result that A's decision, on the basis of A's evaluations, is understandable and acceptable, without B himself having to arrive at the disposition that in A's situation he himself would likewise have to express such a want. I can understand why someone else – relatively to his ideals – expresses certain wants, and I can also respect these as the expression of his personality, without having to appropriate his ideals. Furthermore, Müller's interpretation reveals a second implausibility. For B's perspective recognition of A's altruistic attitude contains the wellbeing of third parties, *in as far* as it is regarded by A as a constitutive aspect of his own value of life, but not the wellbeing of third parties as such. However, if B does not have to make the material content of A's evaluation "his own", then the wellbeing of third parties as such cannot be inferred therefrom.

#### 6.4.1.3 Bok's Objection

The consideration which lead Sisella Bok (1998, p. 112 ff.) to the conclusion that a suicide act supported by a physician is preferable to voluntary euthanasia in the narrower sense, represents slippery slope arguments in the classical sense. She also considers the continuous crossover of cases to be a problem (ibid. p. 114) and moreover points out unfavorable societal factors which will probably allow a negative development – she mentions „societal forces such as demographic pressures, scarcity of resources, religious zealotry, racism, or sheer greed“ (ibid. p. 115). Furthermore she forcefully warns against overrating the competence of medical decision-making under real practice conditions and demands that the discussion of

---

<sup>39</sup>The admissibility of such altruistic motives and the difficulty of integrating them into the model of proxy decisions is discussed in Battin (1994, Chap. 2).

the permissibility of voluntary euthanasia should take the perspective of doctors and their opinions into particular consideration.

These objections should be taken *very* seriously, but in my opinion they cannot justify the conclusion that one should ethically and judicially ban voluntary euthanasia in the narrower sense, but not medically assisted suicide. It is just Bok's realist view of medical practice that suggests two things: for one, that such decisions should not be taken by individual doctors, but in a team; and for another, that accessible and controlled practice are much more likely to detect the real grievances and both negative and positive tendencies in this area. Only in this way will it be possible to develop and establish suitable precautionary and prophylactic measures. But Bok's arguments do not suffice for a generally ethically disparate treatment of voluntary euthanasia in the narrower sense and assisted suicide, and hence for an ethically worse status of the weakest, who are no longer even capable of assisted suicide.

#### 6.4.1.4 Beauchamp's Objection

The defense of the distinction active vs. passive and the thesis that there is an ethically significant boundary between active euthanasia and passive allowing to die is undertaken by Beauchamp (1978, p. 251 ff.) through a combination of slippery slope arguments and rule utilitarian considerations.<sup>40</sup> He concedes "that the active/passive distinction is *sometimes* morally irrelevant" (ibid. p. 247), but he subsequently, depicts the possible or probable negative consequences that would ensue from an ethical and judicial equalization of active and passive euthanasia. These consequences then serve him to formulate a rule utilitarian justification for the ban on active euthanasia. This kind of justification does not start with the individual acts, but with the positive consequences the retention or introduction of a rule would have for a society. Beauchamp is of the opinion that the rule in question, which forbids active and permits passive euthanasia has more positive consequences overall than the alternative rule which allows both.

This argument is also in my view unconvincing. For one thing, rule utilitarian considerations can hardly be asserted. In fact, one might suspect that they are being introduced elsewhere as additional justifications for a rule that is considered ethically correct. For another, in this way the autonomy of the individual person is being unduly restricted. For this gradualist argumentation sure enough makes it possible for individual cases of voluntary euthanasia in the narrower sense to be justifiable, but then prohibits them on the grounds of rule utilitarian considerations, with the conjectured overall societal benefits in mind. This ethical approach not only pays too little respect to the ethical importance of individual autonomy, but thereby also disregards even the positive consequences of the admittance of voluntary euthanasia in the narrower sense (cf. Brock 1993, p. 213 ff.). This can be seen above all therein, that such a changed practice allows patients to preserve their autonomy even in

---

<sup>40</sup>The same argumentation can also be found in Beauchamp and Childress (1994, pp. 219–235).



dying and therewith have security and an ethically legitimized right. For another thing, Beauchamp's position tends to suggest reviewing the individual case and permitting ethically legitimate cases of voluntary euthanasia in the narrower sense. Here, pointing out possible negative consequences should also lead to dangers of abuse being recognized and combated and negative developments being identified and stopped. With this – and this much must also be conceded to this objection – securing and attendant measures are postulated which are suitable for making the *prima facie* legitimate right to a self-determined death possible.

#### ***6.4.2 The Social Context of Voluntary Euthanasia: Second Conclusion***

The gradualist objections show that the question of the ethical permissibility and – a *fortiori* – that of the legalization of voluntary euthanasia in the narrower sense cannot be decided without recourse to the concrete social context of a society. Empirical developments can here tip the scales in favor of, but also against such permissibility. Even when the categorical objections to the admittance of voluntary euthanasia can be disallowed, it does not automatically follow from the personality based justification that the legalization has to follow from the *prima facie* existing ethical claim. However, if one speaks out against such an arrangement without falling back on categorical arguments, then this must happen with full awareness that one is therewith overruling personal autonomy in favor of putative societal welfare. Accordingly, one must produce empirical evidence that the alleged negative consequences will really or at least probably set in. Since the social structures, to which the historical consciousness of a society also belongs, can be different from country to country, it is moreover conceivable that differing national regulations are ethically and judicially necessary. One can allow for the differing gravity attributed to the exceptions from the ban on killing through variously strict restrictive conditions to which euthanasia acts must be liable. The waiving of categorical justifications thereby demands reliance on an intersubjective and rational culture of justification. As already mentioned this also argues in favor of allowing voluntary euthanasia in all forms and subjecting this legalized practice to public control and social reflection on values.

This result cannot satisfy those who are searching for categorical answers and simple solutions. But in view of the complexity of the issue it seems to me to be the position most likely to be justifiable. It should not, however, be understood as euthanasia enthusiasm, since this result not only does not exclude the demand for alternative solutions (hospices, palliative care), but on the contrary and exactly thanks to the recognition of the ethical problem, represents much more of a challenge to put more effort into these alternative routes. But converse enthusiasm should be avoided. Not all patients who express the wish to be killed 'in actual fact' only want to go to

a hospice, and neither will all cases be able to be solved by palliative treatment (cf. on this also Gert et al. 1997, p. 280).<sup>41</sup>

## 6.5 Is There a Duty to Kill?

If the deliberations up to now have succeeded in making the right to voluntary euthanasia in the narrower sense plausible, then two questions arise: How far does this right go? And on the basis of which claims can such a wish to be killed be ethically overruled by legitimate means? The last section of this chapter deals with these two questions.

### 6.5.1 *A Duty to Kill?*

If one argues for a right to voluntary euthanasia in the narrower sense, the question arises whether such a right corresponds to a duty to kill. To anticipate the result, this is not the case, since such a duty would not respect the autonomy and integrity of those who would have to carry out such a euthanasia act. As discussed earlier (cf. Chap. 3), I understand by a ‘right’ an ethically justifiable claim which, though in principle remaining measurable against other legitimate claims, is still of singular importance. Calling it a right should hint at this special status, but not generally preclude weighing alternatives.

If one distinguishes negative rights in the sense of a demand that no other intervenes in its realization as long as he is not seriously harmed through the concrete realization, from positive rights in the sense that others are duty bound to actively participate in the realization of this right, then the right to a self-determined death exhibits both a strongly negative and a weakly positive right. This positive right can be curtailed only by intrinsic and massively extrinsic reasons (cf. below Sect. 6.5.2). The intrinsic reasons include the personalities of all persons involved in the situation, in the case of voluntary euthanasia in the narrower sense, i.e. the person to whom they want to be killed is directed. From the ethical viewpoint, such a person need not pursue this *prima facie* legitimate matter, if the execution of such an act collides with her own ethical beliefs. Since ethical beliefs are a central constituent of the personality, personal integrity at this point opposes personal autonomy.<sup>42</sup> If, e.g. a physician who for himself ethically rejects voluntary euthanasia in the narrower sense is confronted with such a want, is, however, obliged not to frustrate that want and if necessary to pass it on to somebody who can reconcile such an act with his ethical beliefs.

---

<sup>41</sup> Here, one must entirely agree with Stoffell (1998, p. 274) that: “The key issue is not suffering *per se*, but its place in the personal scheme of things”.

<sup>42</sup> General recourse to the professional ethos of the physician is in my view inadequate.

For society, the consequence of this is that it must overall provide a judicial and institutional framework which facilitates the implementation of this weak positive right. Though, due to the above discussed intrinsic counter-arguments, no general obligation may be introduced to guarantee freedom of conscience and personal autonomy, a practice that is regulated judicially and through control procedures should nevertheless make it possible for such a want for voluntary euthanasia in the narrower sense to be fulfilled in a humane way. If, e.g. physicians exclude themselves as a professional group from involvement in this practice with recourse to their professional ethics, then from the ethical viewpoint it would have to be demanded that the right to be involved be transferred to other members of the medical caregiver teams or to the relatives. In view of the spectrum of ethical attitudes to this issue observed among physicians, such a collective refusal is unlikely provided the practice is legalized and controlled.

### **6.5.2 *Who Has Veto Rights?***

Alongside the persons immediately involved in voluntary euthanasia in the narrower sense, two more possible bearers of rights come to mind, who might possibly stake ethically legitimate claims: relatives (in the social sense) and society as such.

- (a) Under extreme circumstances it is conceivable that the wish for voluntary euthanasia in the narrower sense should not be considered because relatives of the patient are negatively affected in a massive way by such a euthanasia act. Not the occasional case turning up in the literature and discussion, in which relatives insist on life prolongation so as to continue to have the pension or warranty protection at their disposal is meant here. Such economic reasons can certainly not be regarded as adequate to overrule an autonomous decision. But rather, I am thinking of cases in which the personal integrity of the relatives would be severely affected by the performance of the euthanasia act. This could be the case e.g. because of religious beliefs, or because the relatives severely reproach themselves for the patient's wish to be killed. Since here the rights of personal integrity and personal autonomy could also be in opposition, non-fulfillment of the want for voluntary euthanasia in the narrower sense can be ethically acceptable, if it is not accompanied by greater suffering for the patient. But ultimately, here one will have to pass judgment in the concrete situation.
- (b) If massive and irreversible effects on society as such are probable, this can ethically justify a general ban on voluntary euthanasia in the narrower sense (cf. Sect. 6.4.1.3). For example, the threat for minorities, the socially weak or marginal groups and also general loss of trust in the medical institution would count as such negative consequences. Whether such a danger exists cannot be determined for all nations and not without reference to the concrete social situation. But if it is probable, and if the admittance of voluntary euthanasia in the narrower sense has the consequence that the right to life of others is endangered,

then there is a respectable reason for not allowing this practice or for putting an end to it. For even then rights based on personal autonomy conflict with one another. Admittedly, it is anything but easy to provide the reality or the probability of such dangerous circumstances with the necessary measure of security which must be demanded so that the ethically *prima facie* legitimate right to voluntary euthanasia in the narrower sense may be overruled.

### 6.5.3 *Dying Autonomously: A Closing Conclusion*

Both the right to live and the right to dying autonomously are strongly negative and weakly positive rights. Although they can generally be weighed up against other massive ethically legitimate rights, they deserve both ethical and judicial protection and an appropriate social and institutional framework in which they can be realized. From the strongly negative right to live it follows that persons may not be regarded as objects of nonvoluntary euthanasia as long as they are capable of autonomous decisions. As long as an autonomous person does not express for herself that she wants to die or to be killed, one must assume that she values her life which must be respected. This value must not – as the discussion of the objections of Grisez and Boyle and of Müller have shown – be equated with wellbeing. Persons are capable of an evaluative *identification* with their own existence. Wellbeing is included as an aspect of this identification with one's own life; however, the value of wellbeing remains divorced because of the internal evaluative response. A suffering or painful existence can nonetheless be regarded as valuable. Whether this is the case depends on a person's attitude to her own existence. At any rate, the principle of respect for personal autonomy demands that this identification with the own existence be respected when it is the expression of just this autonomy. Ignoring it in favor of the objective standards for the assessment of quality of life is not justifiable, either in the case of positive or that of negative identification.

# Chapter 7

## Extended Autonomy

### 7.1 Targets

Modern ethical consciousness tends even in the medical field towards an ever increasing recognition of personal autonomy as its central ethical value. This is reflected above all through medical actions which can be supported by the patient's autonomy being considered ethically less problematical than those based on alternative decision bases (proxy decisions, the objective assessment of the quality of life). Advance directives, in which a person defines how she wants to be treated in situations in which she is currently no longer able to exercise her personal autonomy, are gaining ever more weight because they offer two attractive aspects. On the one hand, they allow a person to extend her personal autonomy into life periods in which she is no longer able to presently exercise her autonomy. Viewed in this way, advance directives promise a possibility of greater steering and control for one's own life. And on the other hand, the availability of an advance directive can relieve those who are involved in a medical decision if the option chosen can be declared the realization of the documented want of the patient. It is therefore no wonder that Stephen Post (1995, p. 307) pronounces that:

It is now a common bioethical assumption that the extension of patient autonomy through advance directives is in principle correct, even if its implementation may be complex.

I concur unreservedly with this general assessment.<sup>1</sup> But things are already less clear regarding the ethical assessment when it concerns the question of whether

---

<sup>1</sup>Admittedly, taking up a formulation by Nancy Neveloff Dubler (1995, p. 289), I do not consider advance directives to be the “‘magic bullets’ of bioethics”. I don't think that they will generally be able “to vanquish the anguish and angst that always accompany the intractable dilemmas of decision making for incapacitated and incompetent patients” (ibid.). For even advance directives still demand the active, interpretive involvement of those who have to implement them. But above all, advance directives are not the only means of choice for every person in every case. It can certainly be the case that the provision of a proxy decision maker is more suitable for patients with specific moral concepts and within the appropriate social structures; cf. Lynn (1991, p. 104) for such argumentation.

advance directives go as far as justifying voluntary euthanasia in the narrower sense.<sup>2</sup> Even more controversial is the issue of whether such advance directives should be respected when they concern patients with Alzheimer's or other progressive and irreversible dementia diseases. And whether they are a suitable means and can be considered in the psychiatric context is also heavily disputed. One reason for this is, as shall be shown below, that, in the debate, the two aspects of personal identity (persistence and personality) distinguished in this study are identified with one another. The other reason is that in these contexts our conceptions of biographical identity and autonomy come into conflict with one another.

For more clarity here, in a first step it is necessary to carry out preliminary elucidation of terminology and case differentiation (Sect. 7.2). On this basis, three objections to the attempt to award advance directives an ethically binding force will then be discussed (Sect. 7.3). Because of the limitation stemming from the superordinate problem in this study, not all the problems of advance directives can be dealt with exhaustively here. However, it will be shown that the concept of personal identity unfolded here is suitable for invalidating the only categorical objection (Sect. 7.3.3). The remaining arguments should conversely be regarded as gradual, since they are unable to annul the ethically binding character of advance directives (Sects. 7.3.1 and 7.3.2). Whilst the third section is concerned with the question of whether there are limits to the extension of personal autonomy through advance directives, following on that the so-called Ulysses contracts, a kind of advance directive in which conflicts arise within personal autonomy, will be discussed (Sect. 7.4). According to my analysis, these conflicts come about through tension between our concepts of biographical identity on the one hand and personal autonomy on the other. For this reason the last section in this chapter is devoted to determining the causes of this tension and thereby attesting the ethical relevance of biographical identity in this context (Sect. 7.5).

## 7.2 Personal Identity, Autonomy and the Passage of Time

To justify the ethical respectability of advance directives even in controversial areas, it is first of all important to point out the perspectivity contained in the chosen terminology, since this for one thing reveals the relation to personal identity and for another possibly transports hidden ethical connotations (Sect. 7.2.1). Secondly, for a clarifying discussion it is essential to disentangle the various issues through case differentiation (Sect. 7.2.2).

---

<sup>2</sup>This case is discussed e.g. by Francis (1993), who develops a cumulative argument out of the legitimate qualms against advance directives and voluntary euthanasia as a whole, which show the indubitable difficulties. However, his negative result only results from the addition of the objection of non-identity (cf. *ibid.* p. 317). This argument is, as it will be shown, not sound (cf. Sect. 7.3.3). But one must agree with Francis (*ibid.* p. 318) that advance directives must not be implemented when the patient is capable of expressing a present autonomous want, unless the conditions of the so-called 'Ulysses contract' have been fulfilled (cf. Sect. 7.4).

### 7.2.1 *The Perspectivity of the Terminology*

Descriptions of the exercise of personal autonomy through advance directives differ in respect of their temporal orientation. The initial situation is as follows: X filled out an advance directive at a point in time ( $= t_0$ ) at which he was able to make autonomous decisions, which expresses what X wants for his own future. This advance directive is now ( $= t_1$ ) relevant and states that X, who in the meantime is no longer competent and in the condition he described in the advance directive wants to receive no life-prolonging measures or to be killed.

In the literature, the autonomy that X exercises through an advance directive is characterized in different ways: Cantor (1992) calls it “prospective autonomy”, whereas Dresser (1994a, p. 629) speaks of “future-oriented autonomy” and elsewhere, adopting the terminological suggestion made by Dworkin (1993, p. 226), speaks of “precedent autonomy” (Dresser 1994b, p. S3). Moreover, Post (1995) uses “extended autonomy” and “precedent autonomy” as equivalent denotations. Even if the general nomenclature of a future-oriented autonomy is unspecific, because every decision is oriented to the future, Dresser’s terminology is from one specific respect revealing. If one describes advance directives as cases of “extended” or “visionary” autonomy, one takes the moment  $t_0$ , in which the advance directive was composed, as the starting point and assumes the perspective that X assumed while he was capable of autonomous decisions. If, in contrast, one uses the denotation “preceding” (or even anticipated autonomy, then one takes X and his present situation to  $t_1$  as the starting point and interprets the advance directive as a decision which has foreseen the present situation and has also defined it through the assessment in the advance directive. In the first form of description, the advance directive, starting from the perspective assuming X to  $t_0$ , is viewed as an instrument for the extension of his autonomy. In contrast, the second form of description interprets the advance directive, seen from the view of X to  $t_1$ , as a stipulation through which the present situation of X is determined in a certain way or at least decisively restricted as regards the agency options.

It is obvious that these two forms of description are not ethically neutral, at least, not when one understands the situation as a conflict between the interests of X to  $t_0$  and those of X to  $t_1$ .<sup>3</sup> With this I do not want to impute that the choice of one of these

---

<sup>3</sup>Savulescu and Dickenson (1998, p. 234 ff.) solve this conflict by developing a dispositional conception of interests and defend the thesis that advance directives are only respectable when the interests expressed in them are at least of “enduring present preference” as disposition (*ibid.* p. 240). In contrast to Savulescu and Dickenson I am, however, firstly skeptical whether one can award a sufficiently precise sense to the talk of dispositional interests in this sense to make them serviceable as a basis of such relevant ethical decisions. But above all I am, secondly, not in agreement with their thesis that only present interests, whether dispositional or not, deserve our respect. If one like Dworkin (1993) differentiates between experiential and critical interests and one takes the structure of biographical identity as a valuing attitude to one’s own future into account, then one can reconstruct the conflict between the values documented in the advance directive and the present interests. But above all, one can show on this basis that autonomy consists not only of the present exercising of specific capabilities, but also as personal autonomy makes, as Dworkin writes

forms of description is always the expression of an explicit ethical assessment (but in my case the discourse on extended autonomy should be understood as such). However, I suspect that ethical intuitions or presuppositions which lead to one perspective being favored over the other are operative here. Since these intuitions or presuppositions are also a source of misunderstandings or differences of opinion, it is important to point out this dimension of the terminology used and to thereby render the leading ethical assessments explicit.

### 7.2.2 *Types of Case*

Four types of case can be differentiated, in which biographical identity and advance directives are related to each other in different ways. Some of these types yield further subtypes as soon as one makes allowances for the fact that there are changes in personality or the loss of the ability to make autonomous decisions. To simplify the discussion, I shall in the following assume that having a personality is necessary and sufficient for the competence to make autonomous decisions. Only during a life span in which X has a personality is X accorded “critical interests” (cf. Dworkin 1993, p. 210 ff.). In the life episodes in which X no longer has a personality, but nevertheless is subject of conscious episodes such as pain sensibility, X has “experiential interests” (cf. Dworkin 1993, p. 201 ff.). If X loses consciousness completely and permanently while still alive, he is no longer the subject of any interests. This determination is a simplification because there are transitional areas between these three stages; but there are also clear-cut cases. Hence the underlying distinction can be met unequivocally. In all, because of the general ethical problem, I am pursuing a conservative maxim: The transition from the interval with critical interests to the interval with only experiential interests is only completed when there is no further flickering of personality or competence (analog to the other transition); flickering can frequently be observed in patients with Alzheimer’s disease. Thus, with this ethically motivated maxim, low standards for competence and autonomous decisions are being applied.

#### **Case A**

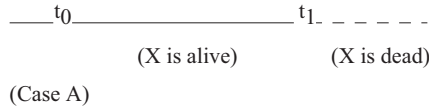
The borderline case of an advance directive represents the consent of a person (to  $t_0$ ) to be an organ donor after death (to  $t_1$ ).<sup>4</sup>

---

(1986, p. 8), each one of us responsible for “for shaping his own life according to some coherent and distinctive sense of character, conviction, and interest”. Advance directives retain their validity in the biography of a person because the integrity of that person is manifested therein, which transcends the single and momentary autonomous decisions: see also Brock (1998, p. 251 f) in his response to Savulescu and Dickenson (1998). This conception of autonomy which I have developed here (Chap. 5) is not reduced to single decisions, but emphasizes, as Dworkin (1986, p. 8) formulates, “the place of each decision in a more general program or picture of life the agent is creating and constructing”.

<sup>4</sup>The following applies to the illustrations in this section: a drawn through line marks the time period during which a human organism X persists; the dotted line marks the time period in which X does not exist. The direction of time runs from left to right and  $t_0$ ,  $t_1$ ,  $t_2$  stand for points in time

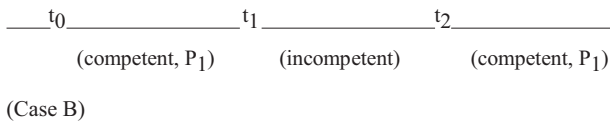




If one understands the event of death as the moment at which the relevant human organism X ceases to exist, then this presents a widely accepted case of extended autonomy, in which persistence is not a necessary condition.<sup>5</sup> But this case is significantly different from such cases in which a person makes decisions about the future of a person who is numerically different from herself. Thus, e.g. X cannot decide that another person Y is available as an organ donor after Y’s death. I am introducing Case A only because it shows that persistence is not a necessary condition for the ethical respectability of advance directives in every case. As will be shown later on, there are also cases in which it is not a sufficient condition.

**Case B**

In this case, X formulates an advance directive in the interval  $[t_0-t_1]$ , which excludes life-prolonging or other medical procedures for certain conditions. This advance directive shall be valid should X no longer be capable of making autonomous decisions at the applicable point in time. Now X finds herself currently (in the interval  $[t_1-t_2]$ ) in such a situation, but it is certain that appropriate medical procedures will lead to X regaining her competence for autonomous decisions, whereby it is beyond question that the personality of X before and after her period of incompetence will be identical.



This case of a *reversible* loss of competence and personality is occasionally submitted as an objection to the *absolutely* binding and *overruling* power of advance directives.<sup>6</sup> There is clearly no good reason for pursuing an advance directive in

---

which limit the relevant intervals in the life of X for the individual cases. P<sub>1</sub>, P<sub>2</sub>, P<sub>3</sub> denote different personalities of X during his life.

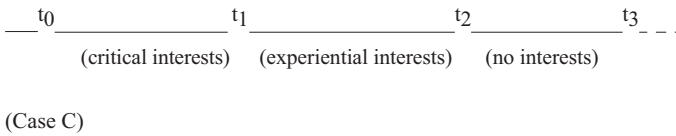
<sup>5</sup>This legitimacy can be contested either under the assumption that human corpses are public property and therefore nobody has the right to decide about “his” corpse alone, or under the assumption that organ transplantations are ethically inadmissible.

<sup>6</sup>Here, my line of action contains two further simplifications; firstly I am not discussing questions regarding the probability of the prognosis, whereby I assume that a decreasing degree at the same time allows conformity to the advance directive to become less acceptable. Secondly, I am not making allowances for the problem which arises as regards the identity criteria for personalities, in that I am assuming that we are dealing either with the same or with a different personality. This can also lead to uncertainty and vagueness. But for one thing, my above case differentiation is only meant as a classificatory framework against which the various ethical aspects can be identified. And for another, here the conformity of the advance directive also becomes the more in need of justification, the less the conservation of the personality can be plausibly assumed.

Case B. Quite generally, advance directives should be so composed or certainly understood to contain the following clause: This directive is valid only in situations in which my current personality cannot be restored, or restored only under unacceptable stress for me. Using Case B as an argument against the validity of advance directives as regards Alzheimer’s disease or other forms of dementia in the psychiatric context (states Dresser 1995, p. 35) misses the real core of this problem, which will become visible in the next case.

**Case C**

In this case a person loses her personality and competence *irreversibly*. To describe this case adequately, it is helpful to differentiate between critical interests (values and convictions), which constitute an essential part of a person’s biographical identity, and experiential interests (experience of feelings or pain).



(Case C)

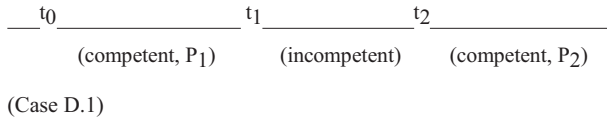
Three stages can thus be differentiated in Alzheimer’s disease and other forms of dementia: a stage in which X is still competent and has critical interests, a stage in which he only has experiential interests and a final stage in which X is no longer a subject of interests. In this case the disease process is irreversible and the decisive question is: May or must one follow an advance directive filled in by X in the interval  $[t_0-t_1]$  if it forbids medical procedures during the period in which X still has experiential interests? Should one respect the extended patient’s autonomy as documented in his advance directive and dispense with medical interventions although this patient still experiences feelings and pain?<sup>7</sup>

**Case D**

The fourth type of case is confusing. Whilst Case C can be described as an inexorable loss of personality and competence, the fourth kind is characterized by *personality changes*. Sometimes competence is lost in phases during illness, sometimes it is present continuously. It is therefore helpful to distinguish three subtypes.

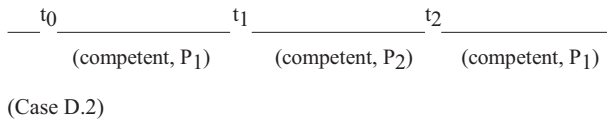
In the first, Variant D.1, the loss of competence of X is reversible, but the medical intervention causes such personality changes that X must be described as having a different personality after regaining his competence.

<sup>7</sup>For a detailed analysis of this case cf. Kuhse (1999). Admittedly, in her discussion she assumes two premises which I do not share. Firstly, she accepts the psychological criterion for personal identity (*ibid.* p. 355 f.) and therewith presumes that an analysis of persistence can be carried out with the concept of person. In addition, her deliberations are secondly based on the premise that non-personal life has no right to live and therefore painless killing is ethically unobjectionable as long as the interests of others do not come into play (*cf. ibid.* p. 361).



Here we have a patient X with a personality P<sub>1</sub> in the interval [t<sub>0</sub>–t<sub>1</sub>] and a different personality P<sub>2</sub> in the interval [from t<sub>2</sub> on], who lives through an interim stage (interval [t<sub>1</sub>–t<sub>2</sub>]) in which he is incompetent. If patient X has signed an advance directive during the interval [t<sub>0</sub>–t<sub>1</sub>], in which medical interventions are forbidden for the conditions present in [t<sub>1</sub>–t<sub>2</sub>], should one respect this autonomous decision? Or should one overrule the advance directive because X will later (after t<sub>2</sub>) regain his competence? In other words: Should cases D.1 and B basically be dealt with identically, or should the personality change be attributed an ethical meaning that is adequate to tip the scales in favor of the advance directive?

D.2 has two features which distinguish it from D.1: For one thing, during the entire period X remains competent to make autonomous decisions, but is subject to a personality change in the interval [t<sub>1</sub>–t<sub>2</sub>]. For another, the former personality of X before t<sub>1</sub> can be restored through a medical procedure.

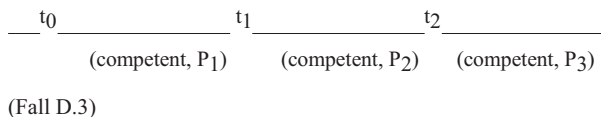


Now, during the interval [t<sub>0</sub>–t<sub>1</sub>] X has filled out an advance directive which stipulates that in case of a personality change (as postulated in interval [t<sub>1</sub>–t<sub>2</sub>]) he wants his former personality P<sub>1</sub> to be restored. In addition he states clearly that one should ignore his refusal of medical procedures should he announce them in the interval [t<sub>1</sub>–t<sub>2</sub>], and regard the advance directive as binding. This case is not merely a thought experiment. It happens in reality and is discussed under the headword “Ulysses contracts” (Dresser 1984). In a Ulysses contract the previous consent to medical treatment documented in the advance directive stands in opposition to the present refusal of this treatment. So the question is which competent decision expresses the ‘true’ (or ‘deep’ or ‘real’) personality of X and should therefore be respected. It is important to distinguish between B and D.2, since in the latter case the ethical intuitions oscillate, because the value of the restoration of the former personality and the value of the respect for the present autonomous decision by X to refuse treatment conflict with one another.<sup>8</sup>

To identify more closely the various aspects that enter into the ethical considerations, a third subtype of D must be presented.

---

<sup>8</sup>Further features that may influence our ethical intuitions in this context are, first, the level of competence X has in the interval [t<sub>1</sub>–t<sub>2</sub>], second, the distinctness of the personality P<sub>2</sub> and, third, which personality, P<sub>1</sub> or P<sub>2</sub>, is socially more required.



D.3 is different from D.2 in the following way: A massive personality change is caused by the medical procedures, so that the previous personality  $P_1$  is not restored, but a new and third personality  $P_3$  is evoked. Should a Ulysses contract have binding power when a new personality is installed? It is important to distinguish between D.2 and D.3 because this illustrates that the difference between the restoration of the former and the evocation of a new personality makes a difference as regards the ethical valuation. In this way it can be seen, for one thing, that the personality is an ethically relevant aspect in biomedical ethics. For another, it will transpire that the justification for the overruling of a current refusal of treatment in favor of the advance directive in case D.3 requires a different kind of justification from the overruling in case D.2, in which one assumes that the original personality can be restored.<sup>9</sup>

Whilst the various variants of case D constitute the background of the dispute about the validity of Ulysses contracts, the arguments against the respectability of advance directives in the context of the care of dementia patients draw primarily on case C. Objections of this latter kind against the ethical admissibility of extended autonomy via advance directives shall be discussed next (Sect. 7.3).

Even if Ulysses contracts are likewise to be regarded as a kind of advance directive, I shall nevertheless discuss them separately (Sect. 7.4), since the ethical problems raised by the latter can be reconstructed as conflicts *within* personal autonomy, whereas the discussion of the admissibility of advance directives for patients with Alzheimer's or other dementia diseases can be understood as the question of what *limits* should be set for personal autonomy.

### 7.3 Limits to Extended Autonomy?

Three types of argument can be distinguished which are called upon to limit the range of extended autonomy and to prove the ethical invalidity of advance directives in the domain of patients with Alzheimer's or other dementia diseases. The first two are of a gradualist nature and themselves plausible, but not strong enough to annul the validity of advance directives. The third argument would scrape by if it were conclusive.

---

<sup>9</sup>Independently of the question of the validity of advance directives, the same structure can occur in the field of rehabilitation medicine. (cf. Chap. 8 on this).

### 7.3.1 *The Problem of Implementation*

It is undeniable that advance directives exhibit unrescindable vagueness: Life and death are complex events and no set form will ever be able to get all their details under control (cf. Dresser 1989). Hence one must acquiesce to Dresser that proxy decisions – whether by family members, courts of law or physicians – must bridge this gap (cf. Dresser 1994a, p. 621–624). The proxy decision will therefore only lead to refinements and specifications appropriate to the situation and not be allowed to change anything of the basic orientation of an advance directive. Moreover, the necessity of such proxy decisions which are acceptable elsewhere too cannot justify advance directives being inadmissible in principle. However, the allusion to the problem of implementation can, firstly, justify better forms having to be developed, possibly especially for specific kinds of illness and situations. Secondly, it reveals that patients who fill out an advance directive must know that an unrescindable risk of vagueness and necessity for interpretation remains. It is therefore wise to name proxy deciders in an advance directive, wherever possible.<sup>10</sup> Thirdly, physicians who sometimes have to close this gap by recourse to their own moral concepts should know what they are doing and what is more know that this is by all means ethically acceptable, even though in need of justification. Such decisions should not be hidden behind a veil of seemingly purely descriptive medical necessity, but explicitly disclosed to the patient as long as he is still competent and to the relatives, and in addition should be documented if at all possible. After all, the problem of implementation shows, fourthly, that because of this unrescindable vagueness, advance directives cannot serve to make extended autonomy a non-overrullable value in every case. One must also agree unreservedly with Dresser (1994a, p. 629) that there are constraints to the adherence to such advance directives. One of these constraints is that patients themselves should document in their directives under what circumstances they want to have them suspended. This would clearly be the case when adherence would mean considerable suffering and injury to the patient. Nothing can be said against life-prolonging procedures being forborne and palliative treatment being carried out. However, such constraints are for this reason, contrary to what Dresser would have one believe (1994a, p. 631), still not an argument against the ethical relevance of advance directives. Rather, they reflect their limits and constitute an attempt to deal with them in a sensible way (cf. Post 1995, p. 312). In addition, there is conjecture that the implementation of advance directives is not possible because many persons cannot comprehend the complex nature of this instrument, or because they would not share its underlying value of autonomy. (Dresser 1994b, p. S3) Neither can be regarded an adequate reason for not developing advance directives suitable for those patients who want to extend their autonomy in this way. For, even when the respectability of such advance directives is

---

<sup>10</sup>On account of this relation alone it would make no sense to disqualify advance directives as a mere expression of “fear” (Churchill 1989) and instead award them general “confidence”. Whoever fills out an advance directive wants to extend his personal autonomy, but cannot do this without having general confidence in the reasonableness of institutions and the social environment.

ethically and judicially accepted, every patient is free to desist from making use of this possibility. So even if Dresser does not succeed in formulating a convincing principle objection to the respectability of advance directives, their arguments are, as regards a tendency which can be observed e.g. in Ronald Dworkin's theory (1993), nevertheless significant. Since my own approach corresponds to that of Dworkin in many respects, I would like to go into this 'subliminal' tendency.

As Dresser shows (1995), some of the arguments in Dworkin's approach seem to justify overruling the experiential interests of patient X even when no advance directive is available for patient X. This consequence seems to arise out of two theses which Dworkin and I accept:

- (P 1) Advance directives must be respected if they lead to the critical interests of a patient dominating his experiential interests.
- (P 2) An advance directive cannot be respected without the addition of some details which are not stipulated in the directive. This can come about either through a proxy decision or via the application of an objective standard for nontreatment.

- 
- (C 1) If advance directives have to be respected in this situation, then the critical interests of a patient should also be respected in such cases in which no advance directive exists. In the latter cases one can directly invoke proxy decisions or objective standards for nontreatment.

#### (Dresser's Objection)

If proxy decisions or the recourse to an "objective nontreatment standard" (Dresser 1995, p. 36) are necessarily involved in respecting an advance directive, then (C 1) appears to follow, but this is not the case: When the elements (i) proxy decision and (ii) recourse to an objective nontreatment standard are embedded in the respect for an advance directive, one cannot deduce therefrom that these elements are to be attributed the same ethical weight as such cases in which an advance directive is available. For one thing, aspects of a decision state are not attributed definable and then summable ethical dimensions independently of the concrete situation. But rather, these elements only attain respect in the situation in question when they contribute to the implementation of the advance directive. For another, an argument submitted by Dresser himself is pertinent here. Since there are patients who do not want their experiential interests to dominate at all events, or assess the value of personal autonomy lower than others, one cannot infer such a generalization.

For this reason, my submissions hitherto must not be understood to mean that critical interests must take precedence over experiential interests in every case. In some cases in which no advance directive is at hand, the proxy decision through persons who are very familiar with the patient and his values can be adequate for drawing such a conclusion. The burden of proof is then on them. No objective nontreatment standard can specify that critical interests generally overrule experiential interests. For that, the ways in which persons realize their own biographies are simply too flexible and rich in variation to allow this result to become part of an *objective*

nontreatment standard.<sup>11</sup> The critical interests of a patient should only be allowed to dominate the experiential interests when he has expressed his values clearly in an advance directive, or when the proxy decider can produce adequately strong evidence that such a result best matches the patient's values. Both cases are distinguished through a shift in the burden of proof: those who want to overrule an advance directive must produce arguments that the patient's experiential interests should tip the scales, whereas the proxy deciders have to offer reasons in favor of respecting the patient's critical interests.

### 7.3.2 *The "Do no harm!" Objection*

Without doubt, an incompetent patient who can sense pain and feelings cannot be regarded as an intrinsically ethically neutral object. Therefore an advance directive must in such a case also be considered against the patient's experiential interests. Acknowledging advance directives in every situation and at all events as binding would be a clear case of "obsession with autonomy" (Dresser 1994a, p. 611). However, as Post (1995) convincingly describes, there are ways of satisfying both *prima facie* justified claims: the one justified by the advance directive and the one justified by experiential interests. For one thing, one can carry out palliative treatment to alleviate or remedy the patient's suffering without disregarding the want for no life-prolonging procedures to be carried out. For another, patients should include an 'exit clause' in their advance directive, so that those cases in which the patient's experiential interests and the advance directive are in irresolvable conflict with one another can be dealt with.<sup>12</sup> A further important point brought to notice by Dresser and Whitehouse (1994) is that the orientation on the principle of respect for autonomy on which the acceptance of advance directive bases must not lead to the life of non-autonomous patients being regarded as intrinsically ethically neutral. To avoid this, one must for one thing postulate a wide theory of evaluative relations such as I have already drawn on several times in this study. For another, one needs a 'thick' concept of autonomy which is not reduced to the capacity for rational decisions. During the course of this investigation I have also developed such a concept.<sup>13</sup> Appropriate treatment of non-competent or handicapped patients is difficult to guarantee without such a comprehensive concept of autonomy. Thus the warnings by Dresser and Whitehouse are important. But they cannot on their own show that the

---

<sup>11</sup> *A fortiori*, it cannot be part of a purely scientific standard, since critical interests cannot even be included in such a truncated concept of the quality of life. But thanks to this plurality, neither can it be – pace Gormally (1992, p. 62) – that well-being is automatically preordained.

<sup>12</sup> Otherwise the proxy decider, who then has the burden of proof, must produce reasons that make it plausible that X would have changed his mind under these circumstances; the mere possibility of such a change isn't enough (cf. Quante 2014b, chapter IX). Different from how Gormally (1992, p. 61) describes this, here the recourse to the personality of X also remains constitutive.

<sup>13</sup> George J. Agich (1993) shows that one cannot waive such a comprehensive concept of personal autonomy in the context of long-term care, either.

increasing acceptance of advance directives is ethically wrong and a sign of a dangerous development. For one thing, they reveal the necessity for a more complex conception of autonomy. For another, as the deliberations in this investigation show, the worth of advance directives already rests on such a comprehensive conception of personal autonomy. Further, the deliberations of Dresser and Whitehouse show that an ethics of health care must be developed (cf. Post 1995, p. 314–319).<sup>14</sup> And finally, the “First do no harm.!” objection can reveal those ethical differences which exist between cases in which a patient still has experiential interests and such cases in which he is no longer a subject of interests as e.g. in PVS cases. In the latter, the advance directive gains stronger weight because no experiential interests exist to balance against it.

### 7.3.3 *The Assumption of Non-identity*

The problems of implementation and the limits set to implementation by the principle of non-maleficence are not in themselves enough to annul *in principle* the respectability of advance directives in the context of Alzheimer’s and other dementia diseases even if they give rise to modifications, specifications and procedural decisions. It can therefore be presumed that the critics who draw such a strong conclusion either base it on a categorical rejection of voluntary euthanasia<sup>15</sup>, or else a further ethical intuition is involved. This additional intuition exists and it rests on the assumption of non-identity. Thus Brock (1994, p. S 10) states that a patient with Alzheimer’s who can no longer remember and identify with his former self should be treated as if he had never drawn up an advance directive in his past. Dresser (1994a, p. 612) comments that the implementation of an advance directive in such cases “distracts us from the *real* people before us” (my emphasis, M.Q.).

What is going on here? In my view, here certain intuitions and philosophical assumptions about the identity of person over time are playing a decisive role. The clearest case in which the autonomous decision of a person need not be respected is that in which this decision does not apply to this person but to another one. Dresser

---

<sup>14</sup>Cf. also Post (1997, Chap. 4) and Martin and Post (1992), where the ethical bases of care by relatives are discussed. Under the current conditions of sparse resources, the increasing number of dementia diseased will of necessity lead to a reorientation of medical care instituted through reorientations as regards values and accompanying philosophical reflection: „That kind of balance will only be likely if we can develop a more coherent, rounded philosophy of health care for the elderly. We require a philosophy that recognizes that the elderly need not an aggressive search for a longer life as an age group, but the avoidance of a premature death and the living of a decent quality of life within that boundary. Our present implicit philosophy – as evidenced by our practices, insurance coverage, and entitlement programs – is heavily biased toward curative medicine“ (Callahan 1992, p. 144).

<sup>15</sup>So e.g. in Gormally (1992, p. 61). Whilst one can agree with his general thesis that advance directives may not contain any ethically impermissible demands, where the limits of the permissible lie is quite a another question (cf. Chap. 6).



speaks of the present patient as the real person and suggests therewith that the person who has drawn up the advance directive no longer exists. Brock suggests handling the case as if the former self had never spoken – because it has no right to decide for (or over) a different self?

Dresser and Brock thereby assume a complex theory of personal identity in which the use of the concept of person is based on a psychological criterion.<sup>16</sup> In doing so, the degree of psychological connectedness is used as a criterion to decide some questions in respect of the validity of advance directives. In this theoretical framework it is possible that we are dealing with two different entities, the former person and the present ‘real’ patient, when a specific threshold value of psychological connectedness has not been attained (cf. Buchanan 1988; Buchanan and Brock 1990). If this were really the case and could the metaphysics of “successive selves” (Radden 1996) be established really convincingly, then there would be a categorical principle and not merely a gradualist argument against the respectability of advance directives in cases in which the identity of the person over time is not available. If one respects it notwithstanding, then the reason can at least no longer be seen in the self-determination of the person (Kuhse 1999, p. 361). However, this argumentation collapses for reasons which have been discussed before (cf. Chap. 2). The analysis of persistence is dependent on the concept of the human organism and can therefore only provide information as to whether we are dealing with one persisting human being or with two different ones. Since a human organism does not cease to persist merely because he suffers a personality change or the loss of personality, persistence can at most be a necessary condition for the respect for advance directives.<sup>17</sup> The concepts of personhood and personality cannot be analyzed without evaluative concepts. Hence a purely naturalistic analysis on the basis of causal relations must underdetermine the biographical identity of persons. The assumption of non-identity fails because it mingles biographical identity with persistence and views personality changes or the loss of personhood as the end of persistence. As long as the patient is alive, case C cannot be described as the decision of a formerly existing person A over the life of a patient B who is not identical with her. The naturalistic dimension of persistence does not include the dimension of biographical identity; conditions sufficient for the range of respect for personal autonomy are only revealed when one describes case C under recourse to the concept of personality.<sup>18</sup>

In these cases there are actual experiential interests of the patient which conflict with his critical interests as documented in the advance directive. If it is not possible to reconcile them with one another, then the question is: which interests should be respected? This question cannot be decided through recourse to ontology, but needs

---

<sup>16</sup>Cf. for this metaphysical precondition Dresser (1986, p. 379 ff.; 1989, p. 156 ff. and 1994a, p. 612) and Buchanan and Brock (1990, p. 152–189); for the effectiveness of this theory of personal identity see also Buchanan (1988), Kuhse (1999) or Meran (1996, p. 384–395).

<sup>17</sup>This applies if one does not regard the organ donor ID card as an advance directive (case A); for a widely concordant criticism of the assumption of non-identity which also draws on a biological approach see DeGrazia (1999).

<sup>18</sup>Kuczewski (1994, p. 39 ff.) also reaches this conclusion.

ethical consideration. In his advance directive the patient has signified that he (at  $t_0$ ) values negatively the life situation in which he now (at  $t_1$ ) finds himself and desires voluntary euthanasia. With this, the patient has himself expressed clearly at a previous point in time that in this situation he does not want his experiential interests to be made the basis of medical decisions. If one adds to this the above proposed restrictive conditions, which should be included in an advance directive, then a strong argument arises out of this explicitly negative identification in favor of generally respecting the advance directive, even if therewith the experiential interests cannot be overruled in every conceivable case. Leading her life as a person means identifying herself judgmentally with her own future (positively or negatively), including her own future experiential interests. As soon as in a culture the concept of leading one's life *as a person* is central, there is a strong reason for recognizing the evaluations of a person which show up in the advance directive as the exercise of her personal autonomy, which is to be respected. Freed from the misguided intuitions concerning the identity of persons over time, advance directives in case C thus deliver a strong means of extending one's own autonomy, though not automatically one that overrules all other aspects.

## 7.4 Conflicts Within Personal Autonomy

In Type C cases, the patient is not able to express any competent refusal of treatment at the point in time at which the advance directive becomes relevant. This is different in Type D cases. The basal conflict in Type C cases can be understood as strife about the extent and *limits of personal autonomy*: Should the value of respect for personal autonomy be allowed to dominate the value expressed in the principle of non-maleficence? On the other hand, in the different variants of Type D cases there is a *conflict within personal autonomy*: What should be respected – the past autonomous decision documented in the Ulysses contract, or the present competent refusal fulfilling at least the minimum standards?

### 7.4.1 Ulysses Contracts

Ulysses contracts, which are valid in the field of mental illnesses and apply to the problem of compulsory hospitalization and involuntary medical treatment, are confusing and raise difficult ethical questions.<sup>19</sup> In accordance with the subject of my inquiry I will limit myself here to those questions that arise regarding personal

---

<sup>19</sup>The legal situation concerning involuntary commitment in Germany is described in Scherr (2015); for a general discussion of involuntary hospitalization cf. Culver and Gert (1981). For the historical development of medical, ethical and legal standards in Great Britain see Fennell (1996, Chap. 15) and for a general overview of the situation in Europe see the contributions to Koch et al. (1996).

identity. Other ethically significant problems such as the basic question of a plausible concept of mental illness, the vagueness of diagnoses and prognoses in the field of psychiatry, or even questions of implementation or interpretation which are also raised in Ulysses contracts just as in other advance directives, will not be discussed below. My general thesis is that the overall situation concerning the validity of Ulysses contracts does not differ from that which has arisen regarding the validity of advance directives in patients with Alzheimer's or other dementia diseases. It is not possible to show that Ulysses contracts need not be respected ethically for reasons of principle (but see Sect. 7.5.2 for a modification of this statement). Thus Ulysses contracts should also be respected and every decision to ignore or overrule them must be justified.

The concept of "psychiatric will", whose 'invention' can arguably be attributed to Thomas Szasz (1982), was conceived as a variant of advance directives which was to be used in the field of psychiatry and generally in the context of mental illnesses. Szasz' main concern was to protect individuals through the formulation of a psychiatric will which prohibits compulsory hospitalization in certain cases of mental illness. In this context, Szasz also mentions a conversely oriented declaration in which an individual states the wish to be medically treated under certain conditions even when the individual refuses this treatment in the acute situation.<sup>20</sup> Whereas Szasz is thinking of the former variant, the latter became the subject of discussion (cf. e.g. Rosenson and Kasten 1991), whereby the term "Ulysses contract" was adopted – probably by Dresser (1984). Although I entirely agree with the declaration that this discussion is one-sided and does not take the possibility of total or partial refusals of treatment adequately enough into account (Rogers and Centifanti 1991), I am, however, pursuing the main line of the discussion, since here the questions of personal identity come into play. Ulysses contracts can be useful for such patients who suffer from recurring but treatable mental breakdowns. This kind of advance directive raises confusing questions in such cases in which the patient also remains (minimally) rational and competent during such a breakdown and does not consent to psychiatric hospitalization, because the outbreak of the illness is accompanied by a massive personality change.<sup>21</sup> With this the problem arises, whether the consent captured in the Ulysses contract or else the current refusal of treatment is to be respected. In such a case in which the individual remains competent throughout, a conflict arises within personal autonomy and the central question is: What should be counted as the "most authentic manifestation" (Appelbaum 1982, p. 27) of this individual's will.

---

<sup>20</sup>Accordingly, one can think of two different ways of handling this: "In the weak version present public policy would prevail unless the individual has executed a psychiatric will stating his or her refusal to be subjected to involuntary psychiatric interventions should these ever be considered. In the strong version, these policy assumptions would be reversed. Involuntary psychiatric hospitalizations could be carried out only on people who had executed a will authorizing such a procedure" (Chodoff and Peele 1983, p. 11).

<sup>21</sup>If competence is not retained, then it is a variant of case B, in which the living will requires exactly that which would have been indicated medically even independent of it. Then there is no specific problem for this variant.

### 7.4.2 *Paternalism or Self-Paternalism?*

In the characterization of these cases the term “self-paternalism” often crops up.<sup>22</sup> This description assumes a complex position of personal identity, in which the former and the current self’ is regarded as two different persons (cf. Radden 1996, p. 161–179). Self-paternalism means that the former self has acted paternalistically, and not that one acts paternalistically towards oneself. In contrast, if one rejects the metaphysics of this complex position and instead assumes the human organism to be the entity which persists, then Ulysses contracts should not be counted as paternalistic acts at all. (cf. Chap. 8). For it is extremely plausible to assume that one can only act paternalistically towards another human being. Using the term “self-paternalism” must therefore be assessed as a further symptom of a misguided conception of personal identity.<sup>23</sup>

Philosophers such as Feinberg (1986), who think of personal identity as evaluative self-relation, describe these cases as self-commitment and future planning, both essential parts of persons’ conduct of life. But the question remains whether Ulysses contracts are ethically respectable forms of self-commitment. So even if the ‘patient’ himself cannot act self-paternalistically, the psychiatrist sees himself confronted with a dilemma in which every choice of action seems *prima facie* paternalistic.<sup>24</sup> If he abides by the Ulysses contract, he is acting contrary to the current will of the patient. If he abides by the current will of the patient, he is disregarding the autonomous decision documented in the Ulysses contract, which was made when the patient was healthy and making provision for his own future. The pertinent rule for normal cases signifies that the current decision is to be respected, which revokes and thereby rescinds the former decision. Type D cases are, however, in many respects not ‘normal’. Firstly, the individual finds himself currently in a situation which he has foreseen and assessed negatively. Secondly, although he is certainly competent as measured by a minimal standard of rationality, he is in a condition of mental dysfunction which has provoked a personality change. Thirdly, it is known that medical treatment would lead to the restoration of the former personality. So one can also be certain that the individual in question would agree retrospectively to the involuntary treatment after this treatment had been carried out successfully. Moreover, fourthly, it applies that the individual frequently displays an ‘impoverished’ personality during the mental breakdown. These reasons speak in favor of

---

<sup>22</sup>Cf. Dresser (1984); cf. also the discussion by Macklin (1987, p. 48 f.) which refers to Jon Elster’s theory of imperfect rationality, see Elster (1987, Chap. 2).

<sup>23</sup>Another reason – a reason that is independent of the premises of a complex theory of personal identity, that works with the psychological criterion – for the assumption that we are dealing with two selves goes back to the following: Since the rationality of the subject is constitutive for its having propositional attitudes, gross cases of inconsistency such as self-deception or weakness of will are hard to explain. A strategy for solving this problem is to presuppose a division of the self, i.e. more than one self; cf. the contributions to Elster (1986).

<sup>24</sup>Whether this is actually the case crucially depends on the basic definition of paternalist actions; cf. the second section of the next chapter (Sect. 8.2).

respecting the Ulysses contract and carrying out the involuntary treatment, since in this way the personal integrity of this individual is esteemed to a greater extent. So if one describes the conflict in respect of the concept of personality there are good reasons for obeying the Ulysses contract. The description as “self-paternalism”, on the other hand, rests on implausible premises. Furthermore, from the psychiatrist’s viewpoint, both choices of action can be described equally as paternalism, so that anyhow, in this way the decisive ethical aspects do not come into it. But if one frees oneself from the fangs of an implausible conception of personal identity, the truly relevant aspects of human personhood in such cases become discernible.

## **7.5 The Ethical Relevance of Biographical Identity**

In this last section the aspects of human personhood relevant for Type D cases are to be portrayed (Sect. 7.5.1) and the ethical relevance of personality are identified through differentiation of various cases of this type (Sect. 7.5.2).

### ***7.5.1 Self-Commitment and Coherence***

If one omits the endangering of others, the problems of interpretation and implementation, and concentrates solely on the dimension of personal identity, then one must agree with Feinberg that this question must be approached under recourse to the concept of biographical identity, by understanding Ulysses contracts as a form of self-commitment. Since we are aware of the fallibility and seducibility of our rationality, we use self-commitments and the help of others to preserve our personal autonomy in situations in which it is endangered. The biographical identity is manifested in this provision for one’s own future. But even when the problem is thus brought to the right place, it is not thereby being solved. As Jennifer Radden (1994) emphasizes, self-commitment and provision for one’s own future are not the only central elements of our conception of personality. Changing one’s views and developing or adopting new moral concepts so as to react to changed circumstances are also important parts of a person’s conduct of life. The decisive question is therefore: What is the person’s authentic will? Even if the concepts of personality and biographical identity are admittedly not entirely precise and are open to interpretation, there are indeed clear-cut cases and distinctive features. Especially in cases of recurring, treatable mental breakdowns it is evident that the change of attitude is not the result of a rational or at least understandable process of decision-making, but the effect of purely causal, arational causes. The coherence of personality manifested in the life of this person is interrupted and the personality pattern underlying the current refusal of treatment is less stable, less coherent and in all ‘poorer’ as that on

which the Ulysses contract is based.<sup>25</sup> If one can restore the former ‘richer’ personality through medical treatment (Case D 2), then one has on hand two reasons in favor of the Ulysses contract.<sup>26</sup> Overall, there are thus criteria on the basis of which it can be decided that one is not dealing with a patient’s change of opinion but with a personality change caused by illness, which the patient has himself foreseen and for which he himself has requested medical treatment. Since advance directives must be respected as *prima facie* binding means of extending one’s own autonomy, the burden of proof in these cases lies with those who want to invalidate a Ulysses contract in favor of a currently expressed minimally rational relinquishment of treatment.

The arguments in this chapter certainly do not suffice to solve all the ethical problems connected with Ulysses contracts. The main target was to free the discussion of misleading intuitions and implausible theoretical assumptions about the identity of persons. Beyond this, the analysis of the merits of Ulysses contracts reveals several central aspects of biographical identity. In addition, it shows that we possess criteria for utilizing these features and that these aspects are factually effective in our ethical perception. The considerations up to now can thus also serve to make them explicit and also to justify this practice. A further relevant point for the conception of personality developed in this study is that in this context the constitutive function of social ideals also becomes apparent.<sup>27</sup> The questions “Which decision is more rational?” and “Which personality exhibits greater richness and greater coherence?” cannot be answered without bringing into play the social and cultural background as normative reference parameters. It is only possible to build a personality and formulate one’s own life as a biography within a network of personal relationships (appreciation and criticism) within a shared *Lebenswelt*. For this reason, purely individualistic conceptions of biographical identity or even personal autonomy are insufficient instruments for dealing appropriately with the problems discussed here (thus also Kuczewski 1994). However, the question remains: How should the individual and social aspects and moral concepts be weighed up against each other in this context?

---

<sup>25</sup> Thus Ruth Macklin reconstructs the portrayal of such cases by a psychiatrist as follows: “It is not simply that they act differently from other people – that they are ‘deviant’, in the usual sense. They are deviant in the additional sense that they depart from their *own* established character. They are not their *true selves*, they lack continuity with their typical or normal or characteristic personality” (Macklin 1982, p. 129). Macklin also arrives at the result (*ibid.*, p. 130) that the respect for the personal autonomy of these patients requires that the actual refusal of therapy be overruled in favor of the Ulysses contract.

<sup>26</sup> In order to detect the ethical relevance of reconstructing the personality without the ‘counterbalance’ of an actual refusal of therapy, one would have to weigh case B against a variant of case D.1 in which a refusal of therapy is fixed in the Ulysses contract.

<sup>27</sup> The social constitution of biographical identity is also emphasized by Kuczewski (1994).

### 7.5.2 *Restoration Versus Installation*

Let us take Case D3, in which a person has filled out a Ulysses contract, but the mental breakdown proceeds in such a way that the medical intervention is not carried out to restore the former personality  $P_1$ , but will lead to the installation of a new personality  $P_3$ . In order to exclude the problem of interpretation at this point, let us assume that the Ulysses contract contains a clause that needs to be observed in this case because the patient, when he remembers his personality  $P_2$  after the breakdown, deeply abhors it and does not want to remain in such a condition. But the patient is now in precisely such an acute phase of illness, manifests personality  $P_2$  and demands not to be treated.

This case D.3 is relevant because herewith, in contrast to D.2, the former personality is not restored, but a new personality installed. If the two cases provoke different reactions regarding whether the refusal of treatment may be overruled, then the ethical relevance and the weight of this difference is thereby pinpointed.<sup>28</sup> If, as my own ethical intuitions signify, D.2 and D.3 should not be treated in the same way, then this must rest on the difference between restoration of the former personality and the installation of a new one being ethically significant.

In my estimation, this difference is not sufficient for an overruling of the current refusal of treatment in Case D.3 to be generally ethically unacceptable. There may indeed be justified cases in which the Ulysses contract should also be respected on the conditions formulated in D.3. If, e.g. the personality  $P_2$  of the patient is so ‘poor’ and so limited rationally, or if the patient is a danger to himself or others in this condition, then there are good reasons for starting medical treatment. The decisive difference is, however, that in these cases one cannot rely primarily on the valuations expressed in the Ulysses contract. In D.3 one must rely much more strongly on general social ideals and the interests of others than in Case D.2, in which the patient’s former patterns of values and convictions can be restored. This difference should be considered in such a way that it leads to a reversal of the burden of proof. Whereas in D.2 one has to justify why one does not respect the Ulysses contract, in D.3 one must justify why one does not accept the current refusal of treatment. In the former case, one can draw upon the patient’s former values and his extended autonomy. When persons document such values then they wish to define who they ‘really’ are. In addition, one can draw upon the refusal of the former person to accept personality change. In contrast, in the latter case one must either weigh up the patient’s current demands against the interests of others, or one has to rely on general, socially shared moral ideals, according to which both the former and the new personality to be installed are more differentiated, multi-faceted and coherent than the actual one. One cannot simply draw upon the former valuations, since the personality – *ex hypothesi* – will be a different one. And since recourse to the anticipated consent of

---

<sup>28</sup>The same holds for the comparison of cases B and D.1 (in the modified form mentioned above). However, the above situation is different, for it can now be seen whether the difference between restoring the previous and installing a new personality is actually so important that it leads to a different result in the weighting against the individual’s actual refusal of therapy.

the subsequent personality does not, in my opinion, constitute an acceptable model of justification in this case either (cf. the fourth section of the next chapter on this), one cannot here rest upon the respect for the personal autonomy of this individual.

This ethical difference between D.2 and D.3 is, on its own, neither sufficient to justify the thesis that overruling the current refusal of treatment is always ethically unacceptable in the latter case; nor is it sufficient to justify the thesis that overruling is ethically always required in the former case. But it is important to differentiate between the two cases for two reasons. Firstly, one can see that the reasons for overruling are of a different nature in the two cases. Secondly, it can be helpful to track down some 'bad' arguments which are submitted against the respectability of Ulysses contracts. Social ideals and the normative background supplied by a society have a different purpose in these two cases. In D.3 they are introduced instead of the patient's individual values whilst in D.2 they are immediate constitutive parts of an individual's personal autonomy, without which nobody could exert their autonomy. The ethical problems accompanying its function in Type D.3 cases should not, therefore, be submitted as objections to its role in Type D.2 cases. Advance directives are thus all in all ethically respectable means of extending personal autonomy. As shown in this chapter, they rest not only on the ethical relevance of the personality, but elsewhere also raise the question of under which conditions paternalistic interventions can be justified; cf. Kleinig (2009). The next chapter is devoted to this problem area.



# Chapter 8

## Medical Paternalism

### 8.1 Targets of Proof

The previous two chapters were primarily concerned with justifying the demands of persons for autonomy as being ethically respectable in the interplay of personal autonomy and biographical identity.<sup>1</sup> Now the question is whether arguments for a justification of paternalistic medical actions can be developed on the basis of the conception of biographical identity unfolded here. Thus the ethical line of attack of the previous two chapters will be reversed; at the same time, the general strategy of the previous chapters will be retained and the theme of medical paternalism limited to the question of the suitability of personal identity as a justifying principle in biomedical ethics. For this reason, in the following not all the aspects of paternalism can be taken up: only the area of medical action will be taken into consideration. Nor shall all justification strategies which have been developed in the almost unwieldy literature be discussed and subjected to critical examination in the course of this chapter. But rather, it will be shown that personality is a relevant aspect in the justification of paternalistic medical action.

In doing so, my considerations are *firstly* not based on an atomistic conception of personality, because general social conditions are recognized as being constitutive for the development of a biographical identity (cf. Chap. 5). Furthermore, I *secondly* assume a ‘thick’ conception of personal autonomy which embraces both realist and objective aspects and is therefore not reduced to a formalistic theory of rational decision or autonomy of agency. Both the conception of personality and the presumed understanding of personal autonomy point implicitly to concepts of the good and successful life. On this basis, which first and foremost establishes the reversed

---

<sup>1</sup>In this chapter I also start from the simplifying assumption that personality and the capacity for competent, autonomous decisions are coextensive determinations. Furthermore, I understand competence such as to include voluntariness. And I assume that it is unquestionable in the discussion of the evaluation of A’s paternalist intervention that B’s actual and documented decision is competent in the sense that *prima facie* it deserves to be respected.

viewing direction of this chapter, the weak, oversimplified oppositeness of autonomy and paternalism can be overcome in favor of a differentiated conception of personal autonomy which can be fairer to both the realities of medical decision-making *and* widely held ethical intuitions equally.

After a suitable definition of medical paternalism has been developed and discussed in the second section, I will discuss Beauchamp's criticism of Feinberg's differentiation of two kinds of paternalistic action (Sect. 8.3). This will show that Beauchamp's argumentation can be consolidated through such a conception of personal autonomy and biographical identity such as was developed here.<sup>2</sup> Following that, in the fourth section *one* kind of paternalistic action will be discussed which can be justified through recourse to personality and biographical identity. Herewith, the difference between conserving the given and installing a new personality, which was already established as being ethically relevant in the previous chapter, will again prove to be ethically relevant.<sup>3</sup> Whilst the liberalist tradition tends to regard personal autonomy and medical decisions primarily as being in tension that can only be resolved in an ethically respectable way through the informed consent of the patient, the recourse to a thick conception of personality is overall more suitable for grasping and conserving our ethical intuitions. Even if I agree with the basic liberalist concept of conceding central importance to the personality of the patient, the conception of autonomy proposed here opens up the possibility of understanding the relationship between physician and patient as trustful cooperation (Sect. 8.5).

## 8.2 A Definition of Medical Paternalism

The concept of paternalistic medical actions to be developed (Sect. 8.2.1) and discussed (Sect. 8.2.2) below must satisfy two conditions: First, aspects of justification must remain factored out, i.e. it should not contain conditions which have only been taken up because of specific strategies of justification. Second, it should not contain any terms which already imply an ethical value, so as not to steer the question of the ethical permissibility of paternalistic actions *ex ante* into specific channels.<sup>4</sup> Every adequate concept must leave room for the possibility of both justified and unjustified paternalistic action (cf. Gert et al. 1997, p. 196). Over and beyond these conditions of adequacy, which are relevant to every definition of paternalism, the following

---

<sup>2</sup>This finding is bit surprising, as Beauchamp tries to manage without such a conception of personality. The deliberations presented in this chapter can thus be counted as additional arguments against his attempt (cf. Chap. 5).

<sup>3</sup>Those readers who have already been convinced by the argumentation in the previous chapter of the ethical relevance of this difference can understand the deliberations presented here as an extension of their scope. The other readers might like to view it as a cumulative argument for my thesis that the personality and biographical identity represent a significant ethical principle in biomedical ethics.

<sup>4</sup>The pejorative connotation of the term "paternalism" must therefore be stringently ignored in our context.

concept contains two constraints which arise out of my superordinate question. So thirdly, I am limiting the class of paternalistic actions taken into account to the area of medical actions. Since no general theory of paternalism is to be put forward, this procedure is justified so as to avoid e.g. cases from the area of state paternalism playing a role as arguments for or against the admissibility of state paternalism in our context.<sup>5</sup> Further, I will fourthly formulate the concept of medical paternalism in such a way that only such cases are included for which whether they can be justified or not is a substantial ethical question. Neither clear-cut cases of unjustified medical paternalism nor clear-cut and indisputably justified cases shall count in the extension of my concept of paternalism. In this way, for one thing the attempt to decide controversial cases through recourse to ethically unequivocal cases can be prevented. For another, I am proceeding thus because I cannot develop a comprehensive theory of paternalistic actions here, but aim directly at those problem cases which can be understood as a conflict between the orientation toward autonomy and the wellbeing of the patient.<sup>6</sup>

### 8.2.1 *The Definition*

The action (doing or omitting) *x* counts in following as a case of medical paternalistic action from *A* towards *B* iff the following conditions have been fulfilled:

- (1) *A* and *B* are two numerically different human beings.
- (2) *A* is medically qualified and capable of doing (or omitting) *x*.
- (3) *x* is an autonomous action by *A* which intervenes in the actual competent decision of *B* or one documented in an advance directive.
- (4) Actions of type *x* are medical actions.
- (5) At the time *A* carries out *x*, *B* is (minimally) competent (i) or has documented his competent decision in an advance directive pertinent to action *x* (ii), which is known to *A* (iii).
- (6) Action *x* is not justified due to general demands of justice or because damage to third parties can be avoided by doing *x* (justification on the basis of the principle of harm-to others).
- (7) *A* carries out *x* for the primary reason of promoting the wellbeing of *B*.
- (8) *A* considers *x* to be justified because *x* promotes the wellbeing of *B*.
- (9) *A* believes that the implementation of *x* is contrary to the actual competent decision by *B* or one documented by *B* in an advance directive.
- (10) *A* does not believe that *x* can be justified by recourse to the harm-to-others principle or reference to general demands of justice.

(Definition of Paternalism)

<sup>5</sup>For a recent analysis of this context cf. Düber (2016).

<sup>6</sup>This strategy is also preferred by Beauchamp (1995); for the purpose of a comprehensive analysis of the concept of paternalism this procedure may, however, appear unsuitable; cf. Kleinig (1983, p. 7 ff.), VanDeVeer (1986, p. 28 ff.) and Häyry (1991, Chap. 3).

Just as the definition of euthanasia actions adopted in the last chapter but one, this concept of medical paternalism also contains *internal* and *external* conditions. Since the individual clauses will be explained shortly, I would here just like to go briefly into their general structure. The first seven clauses formulate external conditions geared to causal relations (the first and seventh conditions) or to justifiable normative intersubjective standards (the fourth and sixth conditions). Conditions (2), (3) and (5) incorporate justifiable causal and intersubjective normative standards.<sup>7</sup> The last three clauses are *internal* conditions in which the primary reason for action and the beliefs in action relevant for the problem of paternalism are named.<sup>8</sup> With actions I also include omissions, whereby an omission can be counted as a medical action if in its definition the recourse to a medical action that is to be omitted is meant in essence.<sup>9</sup> In so doing, I use “medical action” in a wide sense which e.g. should also include care, informed consultation, the advice about the way a medication works, or the indication of a therapeutic option.<sup>10</sup>

### 8.2.2 *Elucidation of the Individual Conditions*

I would now like to elucidate the individual conditions and justify briefly why I consider them essential. I will thereby if necessary also entertain alternative suggestions.

- (a) *The first condition:* With the requirement that A and B must be two different human beings, a condition is being named which usually remains unmentioned, presumably because it is assumed to go without saying. However, as the discussion about self-paternalism in the context of Ulysses contracts has shown, it is necessary to mention it explicitly. From this first condition it follows that the personality change of a human being is not sufficient to regard his decisions qua personality P1 as paternalistic as opposed to his competent decisions qua personality P2.<sup>11</sup>

---

<sup>7</sup>At this point I cannot deal with the epistemic problem of determining whether the external conditions are met.

<sup>8</sup>To be exact, the tenth condition formulates that A shall not have a specific belief in action. I presume the simplifying assumption that one can understand not having a belief p as a logical conjunction of other beliefs. Those readers who do not accept this assumption can alternatively interpret condition (10) as external.

<sup>9</sup>This characterization is adequate for the purposes of this chapter, even if it is not satisfactory from the viewpoint of the theory of action. Incidentally, the above definition does not commit me to a causal theory of action, since no ethical difference is to be justified by the distinction between doing and omitting.

<sup>10</sup>However, medical experiments are not included, no matter whether serving the future use of B or entirely for someone else’s use.

<sup>11</sup>This also applies in the context of psychopathies such as the clinical picture of so-called multiple personalities; the question of how the principle of personal identity can be brought to bear in the context of psychopathies is beyond the scope of this study; cf. Gunnarsson (2010).

- (b) *The second condition:* This condition also expresses an implicitness whose explicit denotation serves primarily to accommodate specific objections against the ethical permissibility of actions, without having to regard them as specific objections against paternalistic action. This condition also makes allowances for the fact that it concerns medical actions. If A, who has absolutely no medical training, performs a complicated operation on B, then it is not an ethically impermissible action only if it is performed against B's will.<sup>12</sup>
- (c) *The third condition:* The third condition says three things. First, it should be an action by A, who performs it on his own resolve (not by force) in an attributable way (not in a momentary condition of e.g. irrationality, physical impairment or misinformation that impairs his capacity for judgment and action in a relevant way). This first condition should again preclude the ethical valuation of the action being influenced by other factors than the fact that it is a matter of paternalistic action. Second, (3) demands that x intervenes in B's *competent* decision. This implies that B is either actually a human person (this follows from the interplay of (1), (3) and the assumption that competence and personality are coextensive concepts), or has composed an advance directive that is pertinent to the given situation. So B must be an actual or former human person. Herewith, not only creatures belonging to the species human being, no matter whether persons or not, are just as excluded as all human beings who are not yet or no longer capable of competent decisions and have devised no pertinent advance directive. The first limitation to *human* persons arises from the limitation to medical action, whereas the second limitation is again a consequence of my strategy of including only those cases of paternalism in which the question of the ethical permissibility constitutes a substantial one.<sup>13</sup> It is not disputed that our action towards small children or incompetent patients should be aligned to their wellbeing. Thirdly, x must represent an *intervention* into B's competent decision. Here it is important that, for one thing, the term "*intervention*" is used purely descriptively and does not implicate any ethically negative connotations in the sense of improper interference (cf. VanDeVeer 1986, p. 17 ff.).<sup>14</sup> However, not every exertion of influence on B should be classed as an intervention, for otherwise even the doctor's consultation would also be an intervention. To make an intervention out of a mere exertion of influence, x must either go against B's actual or documented competent decision, or aim at influencing B's beliefs and

---

<sup>12</sup>The situation is different when A performs a medical act (e.g. an amputation) without training in an emergency with the emphatic consent by B. But this case is certainly not paternalistic action.

<sup>13</sup>This formulation is at least misleading, since one can very easily understand the qualification of the excluded cases as "not substantial" such that they are ethically without severe problems or even justified. This surely wouldn't be plausible since nudging etc. surely can be ethically problematic or wrong; cf. Fateh-Moghadam and Gutmann (2014), Gutmann (2014) and Schöne-Seifert (2009) for such critique. Therefore, I have to make explicit that "substantial" here means only that these cases do cause deep ethical problems on a conceptual level.

<sup>14</sup>Through the inclusion of this intervention criterion I am allowing for objections which show that the criteria of coercion towards B or the limitation of B's freedom are too narrow; cf. on this the discussion in Kleinig (1983, p. 5 ff.).

ideals in such a way as to circumvent B's cognitive abilities and which B would not have accepted had his attention been drawn to this exertion of influence.<sup>15</sup>

Thus it is possible that certain ways of formulating a diagnosis, conducting a consultation or describing risks can be classed as paternalistic actions. And the revelation of information against B's will is at all events an intervention and not just the exertion of influence, since it collides with a competent decision by B.<sup>16</sup>

- (d) *The fourth condition:* Here the range is limited to the question of this study, whereby medical actions should be understood in a broad sense so that they include e.g. the care of B. Thus, not only physicians, but also nurses or male nurses can take medical action. In addition, adequately informed or trained relatives who e.g. administer medicine to (or withhold it from) B count as persons capable of executing medical actions. The relaying to or depriving B of relevant medical information should count as medical action.<sup>17</sup> But medical experiments which do not represent an actual therapy are excluded. Thus e.g. such research on B conducted with the intention of serving B's wellbeing in future is excluded from the topic area I am surveying.<sup>18</sup>
- (e) *The fifth condition:* Two important points are named in this condition. Firstly, paternalistic actions in the area of medical action are restricted to actual or former persons, since the command over competence adequate for making auton-

---

<sup>15</sup> Cf. VanDeVeer (1986, p. 19); my formulation diverges from VanDeVeer's analysis, because he limits paternalistic actions on the one hand to the interference in actual values ("operative preferences, intention, or disposition") whilst on the other hand not limiting it to competently and autonomously held values. In contrast, I firstly also include intentions, then secondly, extend the relevant area to the competent decisions documented in advance directives and thirdly, limit the area to just such competent decisions made by B. In the previous chapter I already developed my reasons for taking into account not only actual competent decisions, but also those manifested in advance directives. The other divergence from VanderVeer's tactic arises from his endeavor to develop as comprehensive a conception of paternalism as possible whereas I am trying to isolate the most contentious cases for ethical theory.

<sup>16</sup> Since the availment of a right not to know can also be a legitimate expression of personal autonomy; vgl. Quante (2014b, chapter VII).

<sup>17</sup> The example frequently mentioned in the literature, in which a physician, on being asked by a dying mother whether her son is OK conceals that he was killed in a traffic accident the previous day, is, however, not a case of medical paternalism. Physicians can also perform paternalistic actions which are not medical actions. This fact, and the fact that non-physicians can also perform some medical actions, does not allow the area of paternalistic actions relevant to my questioning to be defined through the proviso that A must be a physician. I have therefore chosen the alternative strategy, even though I thereby have to take into account the less attractive consequence that there are possibly cases of justified paternalistic action by physicians which I cannot appropriate. However, I fail to see how the charitable lie can constitute a specific problem of medical paternalism, whereas some paternalistic actions by relatives are definitely applicable. For this reason I consider my strategy to be more fitting.

<sup>18</sup> The reason for this is first and foremost that the permissibility of human experiments in the medical field is not primarily a question of paternalism and least of all one of paternalism towards competent patients (as in the sense of paternalism defined here). Since I am reconstructing the problem of paternalism as a conflict between personal autonomy and patient wellbeing, it is expedient to ignore the non-congruent problem of the permissibility of human experiments; cf. McNeill (1993).

mous decisions and having a personality are by definition coextensive conditions. Through the additional qualification “*minimal competence*”, on the one hand, the area of paternalistic action should be kept wide, so as to expand the need for justification as far as possible. On the other, this condition bases on the ethical strategy of excluding as few patients as possible from the principle of respect for autonomy through an overly demanding concept of personal autonomy.<sup>19</sup> Secondly, according to the fifth condition, such competent decisions as those documented in advance directives are also relevant. This is the consequence of my deliberations in the previous chapter: If advance directives are ethically respectable means of prolonging personal autonomy, then overriding such decisions is also a case of paternalistic action. The consequence thereof is that one can act paternalistically not only towards actual persons but also towards former persons (not: personalities!), insofar as they have formulated an advance directive.<sup>20</sup>

- (f) *The sixth condition*: With this condition, those actions in which an intervention into B’s competent decision is justified for reasons of justice or possible danger to third parties are excluded from the sphere of paternalistic actions. Such actions as the following do not count as paternalistic action even if A performs them with the primary reason of promoting B’s wellbeing, because they can be justified by principles that are independent of the concept of paternalism.<sup>21</sup> Thus e.g. John Stuart Mill counts paternalistic interventions which are justified by the Harm-to-Others-Principle as clearly justified paternalistic actions (Mill 1998, p. 62).<sup>22</sup> If, as I do here, one wants to count as paternalistic only such actions in

---

<sup>19</sup>Here, the condition of competence must be understood as a threshold concept. It does not mean that a greater measure of competence on B’s side requires a higher level of paternalism on A’s side, even if it is true that the concurrence of B’s decisions that are the expression of his personal autonomy to a greater degree does demand stronger bases of paternalistic justification on A’s side than the concurrence of less well anchored decisions by B; cf. Kleinig (1983, p. 9). However, it does mean here that only from a certain measure of competence onward can one even speak of paternalistic actions towards B (cf. also Faden and Beauchamp 1986). Kleinig (1983, p.69), who interprets this condition as the voluntariness of B’s decision, comes to the same conclusion. Please note that here I understand a competent decision to be one which implies voluntariness in the sense of the absence of external coercion, disruptive psychological factors or massive misinformation.

<sup>20</sup>Arneson (1980) also takes documented competent decisions into account. However, his definition (ibid. p. 471) does not count the intervention into an actual competent decision by B as paternalism if B’s previous decision explicitly overrules the actual one. A conception of paternalism without any acknowledgment of documented decisions is developed e.g. by VanDeVeer (1986, p. 19 and p. 22 ff.); it should also be noted that clauses (i) and (ii) of my condition (5) can both be fulfilled at the same time (it is not a matter of an exclusive “or”). This was relevant to my analysis of Ulysses contracts (cf. Sect. 7.4.2 of the previous chapter).

<sup>21</sup>So it is possible that A gives recurring paternalistic justification of condition (7) for x even though x isn’t a paternalistic action at all.

<sup>22</sup>A classical formulation of the Harm-to-Others-Principle as regards state paternalism can be found in Feinberg (1986, p. xvi): “It is always a good reason in support of penal legislation that it would be effective in preventing (eliminating, reducing) harm to persons other than the other (the one prohibited from acting) *and* there is no means that is equally effective at no greater cost to other values”.

which an intervention into B's competent decision is undertaken and justified in favor of B's wellbeing, then it is reasonable to exclude the other cases which have a justification that is independent of paternalism and are by and large accepted. For otherwise, there is a danger that the question of a justification of paternalistic actions will be answered positively with recourse to such cases which are justified independently of paternalism. To subsume this condition somewhat more precisely, it is necessary to give a brief explanation of the Harm-to-Others-Principle. Someone can be injured (i) by another person's action, (ii) by an event such as a natural disaster, or even (iii) by his own action.<sup>23</sup> The Harm-to-Others-Principle can be applied as the reason for an action x with which A intervenes in B's competent decision only in the first two cases. If B inflicts injury on a third person in the realization of his competent decision (first case), then the Harm-to-Others-Principle can be drawn on as a justification for x.<sup>24</sup> However, this principle cannot be drawn on if one intervenes in competent decisions by B if B only injures himself due to a competent decision.<sup>25</sup> If in such a case one intervenes in favor of B's wellbeing, one must assume that paternalistic action can justify the intervention into B's autonomy. One can call this assumption *the principle of paternalism*; I accept this, as Beauchamp also does, as an independent principle for the justification of the limitation of autonomy. Paternalistic action shall be limited to those actions that the agent attempts to justify with this principle of paternalism and which fulfill the other required conditions.

At this point, a strategy of argumentation must also be addressed briefly, which, with recourse to a complex theory of personal identity, attempts to replace the principle of paternalism as an independent justification for such cases by the Harm-to-Others-Principle in which B's competent decision has repercussions on a period of life in which B has a different personality. Here, the present and future personality are treated as two distinct selves, so that one can intervene in B's actual decision with the wellbeing of this future successive self in mind and rest this on the Harm-to-Others-Principle.<sup>26</sup> In contrast, inter-

---

<sup>23</sup>It is not possible for me to go into detail about the concept of harm here. On the one hand, it includes minimization of the probably achievable wellbeing of the harmed individual, whereby wellbeing is not limited to experiential interests. On the other, the death of an individual can itself also be harmful for this individual if further existence had represented wellbeing for this individual; cf. on this traditionally controversial question of whether death can be an evil (Feldman 1992, Chap. 8 and 9).

<sup>24</sup>For the application of the Harm-to-Others-Principle in case (ii) cf. the third section of this chapter.

<sup>25</sup>Competence here includes B knowingly precipitating this limitation of his wellbeing through his decision. An intervention aimed at determining whether B's decision is competent in this sense is interpreted as weak paternalism by Feinberg, cf. Feinberg (1986, p. 12 – there Feinberg used the distinction of “soft” versus “hard” paternalism). However, it is more sensible not to regard these interventions as forms of paternalistic action (cf. also the discussion in the third section of this chapter).

<sup>26</sup>Such argumentation can be found in e.g. Regan (1983, p. 122 ff.); cf. also the discussion of this strategy in Kleinig (1983, p. 45 ff.) and VanDeVeer (1986, p. 155–163). Ludwig Siep (1987,



ventions with B's future personality in mind continue to be excluded from the application of the Harm-to-Others-Principle and therewith candidates for paternalistic action (also in VanDeVeer 1986, p. 163).

- (g) *The seventh condition*: The requirement that the primary reason for A's intervention must be a question of B's wellbeing should be interpreted liberally. Thus it is permissible for A inter alia to have additionally something else in view (e.g. the wellbeing of third persons, one's own emotional needs, ethical ideals or even one's own reputation). Here, the simplifying assumption that this primary reason is ascertainable is presumed, which may indeed be possible in the majority of cases. The important thing is that B's wellbeing is meant as either the actual wellbeing or one adequately clearly defined that is to ensue in the near future. In contrast, actions which intervene in B's competent decision so as to serve the wellbeing of A or third persons, or to respect other ethically relevant aspects, do not count as paternalistic action. If they can be justified, it is not by virtue of the principle of paternalism.<sup>27</sup> Furthermore, the wellbeing of B aspired through the paternalistic action must benefit B without B succumbing to a personality change in the time period between the paternalistic intervention and the realization of wellbeing (for more detail cf. Sect. 8.4).
- (h) *The eighth condition*: Here the primary reason for A, which must contain conviction of agency that x is justified because it promotes B's wellbeing, will be defined in more detail. For one thing, it is important that this conviction must be valid for the concrete action and does not belong to A's beliefs as a general principle. For another, it is significant that this concerns an internal condition. It does not follow from the eighth condition either that x is really justified or even that x really promotes B's wellbeing.<sup>28</sup>
- (i) *The ninth condition*: Here the requirement is that A understands his action as an intervention in B's competent decision. This includes the knowledge that actions of this type require ethical justification and A's conviction that in this concrete situation x can be justified through recourse to the principle of paternalism (follows in conjunction with the eighth condition). This is also an internal condition, so that it is possible that no competent decision by B exists

---

p. 253 f.) also sees in this option – above all of Derek Parfit's version of personal identity – positive opportunities for ethics, but is more skeptical than Parfit as regards the trade off with possible costs. Since I myself consider even the underlying theory of personal identity to be mistaken, this strategy is excluded here through condition (1).

<sup>27</sup>The case described by Fairbairn (1995, p. 168), in which a woman adapts her husband's nutrition for reasons of animal ethics without his noticing, therefore does not count as paternalistic action. Fairbairn formulates a definition of paternalism which forgoes my seventh condition (ibid. p. 167 f.).

<sup>28</sup>The expression "promote wellbeing" can be rendered even more precisely; cf. VanDeVeer (1986, p. 22): "A does (or omits) x with the primary or sole aim of promoting a benefit for B [a benefit, which, A believes, would not accrue to B in the absence of A's doing (or omitting) x] or preventing a harm to B [a harm which, A believes, would accrue to B in the absence of A's doing (or omitting) x]" (The abbreviations have been adjusted to mine; M.Q).

and A's action is thus not a case of paternalistic action, although A asserts paternalistic justification for x.

- (j) *The tenth condition*: This last condition, which I also interpret as an internal condition, states that A does not assume that one can justify x through recourse to the Harm-to-Others-Principle. This can either be fulfilled through A explicitly believing that x cannot be justified in this way, or through his accepting the Harm-to-Others-Principle and believing certain assumptions about B's competent decision to be true, which together preclude the Harm-to-Others-Principle from being applied to x (A himself does not come to this conclusion explicitly and does not even consider the Harm-to-Others-Principle in this concrete situation).

### 8.3 Strong and Weak Paternalism? Beauchamp's Criticism of Feinberg

In this section, first the distinction between strong and weak paternalism adopted by Joel Feinberg will be commented on (Sect. 8.3.1) and Tom Beauchamp's criticism presented (Sect. 8.3.2). After that, I will discuss various objections which have been raised against Beauchamp's criticism (Sect. 8.3.3). After the rejection of these objections and the associated clarification of Beauchamp's criticism it will be shown that the conception of personality I have developed provides a basis for Beauchamp's argumentation strategy (Sect. 8.3.4).

#### 8.3.1 Feinberg's Distinction

In his analysis of state paternalism, Joel Feinberg distinguished between weak and strong paternalism (Feinberg 1973, p. 50–54 and 1986, p. 16).<sup>29</sup> This distinction is also relevant for the area of medical action. If one formulates one's definitions appropriately and thereby does not make the subject of discussion that Feinberg makes his distinction with the justification of paternalistic action already in mind, then the following transpires: In a case of *weak* paternalism, B's decisions, actions or conduct are intervened in with the aim "to prevent self-regarding harmful conduct (so far it *looks* 'paternalistic') *when but only when* that conduct is substantially nonvoluntary, or when temporary intervention is necessary to establish whether it is

---

<sup>29</sup>In his earlier publications, Feinberg uses the characterizations "weak" and "strong", whereas later he uses "soft" and "hard". Both characterizations are according to his own information "perfectly interchangeable" (Feinberg 1986, p. 377 fn. 16). In the following I use the distinction between strong and weak paternalism that is currently more in use in the literature of biomedical ethics; cf. Beauchamp (1995, p. 1915).

voluntary or not”(Feinberg 1986, p. 12).<sup>30</sup> In a case of strong paternalism there is intervention in a competent decision by B with the aim “to protect competent adults, against their will, from the harmful consequences even of their fully voluntary choices and undertakings” (ibid. p. 12).

Feinberg himself has in the meanwhile withdrawn his distinction because of Beauchamp's criticism, and now no longer claims that there are cases of both weak and strong paternalistic action. But since he brought in the distinction with the aspect of justification also in mind, he still considers it useful in a transformed form. He suggests characterizing a position as “*soft anti-paternalism*” if it considers cases of strong paternalism (in the above sense) to be ethically impermissible and, in contrast, cases of weak paternalism (in the above sense) to be ethically justified (ibid. p. 15). In contrast to this, the stance of “*hard anti-paternalism*” (ibid.) would reject both kinds of intervention as ethically unjustified. I also consider this distinction rather unhelpful: If the cases of weak paternalism do not represent paternalistic action, then soft and hard paternalism are no different from one another as regards paternalistic actions. Feinberg's new distinction comes to nothing at best; presumably it is more likely to lead to confusion in the discussion about the justification of paternalistic action. But rather than pursuing this question, I would now like to present the criticism by Beauchamp which induced Feinberg to rescind his original suggestion.

### 8.3.2 *Beauchamp's Criticism*

Tom Beauchamp rejected Feinberg's suggestion and advanced the thesis that cases of weak paternalism should not even be counted as paternalistic action (Beauchamp 1995, p. 1917; cf. also Beauchamp 2009). His justification for this thesis consists in two arguments.

- (a) On the one hand, it ensues from the intellection that one should not conduct the debate on the ethical permissibility of paternalistic actions with recourse to such cases which are undisputed between supporters and opponents of paternalism. But as regards the cases subsumed under weak paternalism, there is simply no radical ethical dissent: “Weak paternalism has no clear substantive moral disagreement with antipaternalism” (ibid.).
- (b) On the other hand, Beauchamp's thesis rests on a consideration that arises directly from the first argument. If there could be a *substantial* ethical difference of opinion regarding the ethical permissibility of paternalistic action, then the

---

<sup>30</sup>It is clear that according to my own definition of paternalism this is *not* a case of paternalistic action. Feinberg's remark that this action would appear paternalistic under a certain aspect expresses the fact that he himself has accepted Beauchamp's criticism and meanwhile argues the thesis that “soft paternalism is really no kind of paternalism at all” (1986, p. 16). My discussion in this section does not serve to rehabilitate this distinction, but rather, that I wish to show that Beauchamp's criticism can be grounded through my conception of personality.

justification of such actions must base on an other separate principle. In other words: The dispute over the ethical permissibility of paternalistic action must not be interpreted as an application of other already respected principles of justification (justice, Harm-to-Others), but must be reconstructed as the dispute over the validity of the *principle of paternalism*, as an independent justification for the intervention in personal autonomy.

On this basis, Beauchamp's criticism of Feinberg's proposal to introduce cases of weak paternalism can be understood thus: In these cases there is no substantive ethical dissent because such an intervention can be justified under the Harm-to-Others-Principle: "Weak paternalism does not seem to rest on a liberty- or autonomy-limiting principle independent of some moral principle of beneficence that supports prevention of harm to others" (ibid.). In order to justify this objection, Beauchamp must therefore show that the cases characterized as weak paternalism by Feinberg can be justified with recourse to the Harm-to-Others-Principle.<sup>31</sup> VanDeVeer (1986, p. 32) distinguishes three ways in which someone can be harmed:

In general, harm may occur to S in three different types of circumstances: (1) the situation in which it primarily results from S's action, inadvertent or not; (2) the situation in which it primarily results from the action of others, inadvertent or not; and (3) the situation in which it primarily results from an act of nature, for example, an earthquake.

This segmentation must be modified if one wants to make Beauchamp's criticism plausible. Case (2) is unproblematic: If one intervenes in B's competent decision because harm to third persons can thereby be prevented, then the Harm-to-Others-Principle is relevant. Case (3) can also be interpreted unproblematically thus: Someone intervenes in the causal course of natural events to deflect harm from B or third persons and justifies this intervention with reference to this aspect. Since a pure "act of nature" (e.g. an earthquake or fire after lightening has struck) is not the expression of a competent decision by someone, such an intervention does not constitute a paternalistic action.<sup>32</sup> The problematic and for Beauchamp decisive case is the first one. Here, VanDeVeer's formulation cannot be accepted. VanDeVeer also counts such cases as self-induced harm that, though caused by B's action, were neither intended nor anticipated by him. This is of no use to Beauchamp's strategy of argumentation, because then the cases classified by Feinberg as weak paternalism would also belong to the category of self-induced harm. An intervention in favor of

---

<sup>31</sup> In a somewhat different context Beauchamp uses the term "injury" instead of "harm" to circumvent problems in Mill's conception of "harm"; cf. Beauchamp (1977, p. 71). I will not discuss these complications in the following and use both terms interchangeably. This means that "harm" contains the connotation of requiring ethical justification.

<sup>32</sup> In contrast to mere behavior, actions are not *purely* acts of nature, even when they for this reason do not need to have anything supernatural about them. However, they are intended by the agent. If one intervenes in a purely natural process initiated and intended by an action by e.g. diverting a boulder that is rolling towards a group of people, then one can assess that as an indirect intervention in the competent decision by the assassin. But the intervention can be justified through recourse to the Harm-to-Others-Principle and is thus likewise no paternalistic action.

B's wellbeing could then no longer be justified by the Harm-to-Others-Principle, but would have to be counted as paternalistic action – just a consequence that Beauchamp wants to avoid. So, if one would like to defend Beauchamp's strategy of argumentation, one must limit the first case to actions in which B intends self-induced harm, or at least knowingly takes it into account.<sup>33</sup> The other cases belonging to the category of weak paternalism would then have to be interpreted according to model of cases (2) or (3) instead. If one wants to avoid the extravagant hypothesis of interpreting unintentional self-induced harm as harm to a different self, as this is occasionally considered in the context of Ulysses contracts or in the explanation of the clinical picture of the multiple personality disorder, the only way left is to analyze the case belonging to the category of weak paternalism according to model (3). And this is what Beauchamp suggests.<sup>34</sup> At a point dedicated to the discussion of Mill's antipaternalistic conception, Beauchamp argues that Mill permits paternalistic actions as being ethically legitimate. Mill himself considers an intervention in persons' decisions, conduct or actions permissible if these persons would unintentionally or unwittingly harm themselves should one not intervene. As Beauchamp tries to show, this concession is, however, not an acknowledgment of paternalism, but can be justified by the properly aligned Harm-to-Others-Principle. According to Beauchamp, an intervention in B's decision, conduct or action is justified when:

There exist supportable grounds for believing that an individual or group of individuals has been or will be physically or mentally harmed by some cause or condition which is to that party not known or not within its control or both (Beauchamp 1977, p. 71).

If a person does harm to herself through her unintentional conduct or through unintentional or unknown consequences of her action, then one should not regard the cause of this conduct or this action as her own decision or her own action. So one should view it as a kind of harmful natural event and thereby obtain a variant of case (3). If somebody does not know that she is harming herself through an action *x*, then one should not regard the decision, through which *x* is in fact intended under one description, as an expression of her autonomy.<sup>35</sup> In the case of conduct, one should regard the underlying mental factors as just as much causal factors not belonging to the person as in the case of involuntary decisions which are made either due to mental disturbances or massive wrong convictions. At this point, Beauchamp (*ibid.* p. 75) quotes Feinberg approvingly: "Insofar as the choices are not voluntary they

---

<sup>33</sup>Here I am not discussing the additional complications of assumptions of probability and B's risk-taking.

<sup>34</sup>To be exact, herewith cases are being captured in which one intervenes in B's decision, action or behavior because one regards the probably resulting harm as unintentional. The other kind of cases, in which one intervenes sporadically in order to find out if B intends this consequence or knowingly takes it into account, are not directly embraced in this way. Such temporary interventions will not be dealt with in the following. I view them as occasional interventions with the aim of primarily clarifying the ethical situation rather than avilment of paternalism or the Harm-to-Others-Principle.

<sup>35</sup>The example given by John Stuart that has become a topos is that of the man who intends to cross a bridge and does not know that it will collapse when he sets foot on it.

are just as alien to him as the choices of someone else” (Feinberg 1973, p. 45). Just as we are allowed to protect an inebriated B from his harmful, fully unintentional conduct, we may also protect an agent B from the unintentional or unwitting consequences of his action (cf. Beauchamp 1977, p. 75). In both cases, the Harm-to-Others-Principle is applied as if one had to protect B from the effects of an act of nature. The noteworthiness consists in this case merely in that B’s conduct or B’s action is this natural process.<sup>36</sup> Before I show that these interpretations of cases which Feinberg has grouped as weak paternalism can be made plausible on the basis of the conception of personality that I have developed, I would first like to invalidate some of the objections that have been raised against Beauchamp.

### 8.3.3 Three Objections

- (a) John Kleinig’s objection against Beauchamp is that we cannot justify our intervention in these cases on the basis of the Harm-to-Others-Principle because no second person or party is involved:

The harm principle operates only where two parties are involved, and in this case [the man and the bridge; M.Q.] (as in many other cases) there is only one party involved (Kleinig 1983, p. 8 f.).

With reference to this case, Joel Feinberg, who calls it a “single-party case of nonvoluntary self-harming conduct” elaborates:

In the single-party case of nonvoluntary self-harming conduct it is *as if* the actor needs protection from another person, but of course it is not literally true that there is some other person in a comparable state of ignorance, retardation, or intoxication who must be restrained from ‘harming others’ (Feinberg 1986, p. 13).

One must absolutely agree with Feinberg that in these cases no other person exists who would have to be restrained from harming B. Such an extravagant conception of multiple selves is, however, quite unnecessary. Rather, one can interpret as natural processes the causal process in which one intervenes in such cases. It is not necessary to pretend in such cases that another person is involved. If Feinberg and Kleinig think that the Harm-to-Others-Principle cannot be applied because of the absence of a second person (or “party”) then they are overlooking the fact that not only persons but also natural events can cause harm. If one accepts this, then one can intervene in these natural processes so as to avert harm to third persons. That these interventions can be of the kind that is at the same time conduct or even action, is not a contradiction in terms.

---

<sup>36</sup>In the case of mere behavior, it is a matter of a purely natural process; in the case of an action one must decide between those that can be described as intended and those that are the expression of personal autonomy. There are deliberate actions which are too isolated from the agent’s personality to be able to be counted as the expression of personal autonomy (cf. Sect. 8.3.4 of this chapter).

- (b) VanDeVeer (1986, p. 29 f.) offers two reasons by Beauchamp for not counting interventions in cases of unintentional or unwitting self-harm as paternalistic action. For one:

It is not a question of protecting a man *against himself* or of interfering with his liberty of action. He is not acting at all in regard to this danger [the man and the bridge; M.Q.]. He needs protection from something which is precisely not himself, not his intended action, not in any remote sense of his own making (Beauchamp 1979, p. 68).

VanDeVeer (1986, p. 30) concedes that

there is *a sense* in which a person who did not know the bridge was unsafe was not voluntarily performing the act of crossing an unsafe bridge.

The target of his criticism is that an intervention in the attempt of this man to cross the bridge nevertheless represents an intervention in the freedom of agency of our bridge-crosser. It is therefore implausible for Mill not to recognize such an intervention as constraint of freedom of agency and therefore not justified paternalism.

This objection is justified, but it does not capture the essential point of Beauchamp's position. In the above quotation, though Beauchamp does in fact follow Mill in respect of the latter's statements on freedom of agency, the target of Beauchamp's own decisive point is that these cases are not an expression of personal autonomy ("against himself", "his own making"). Beauchamp presumably imputes a demanding concept of freedom of agency, in which it is a requirement that the action is an expression of the agent's personal autonomy and in this sense "of his own making". Beauchamp's general argument can be upheld if one does not understand paternalistic action as an intervention in freedom of agency but as an intervention in the competent decision or action and therefore in personal autonomy. If one emphasizes this aspect of Beauchamp's argumentation and avoids making the intervention in freedom of agency a necessary or sufficient condition of paternalistic action, then VanDeVeer's first argument runs aground.

- (c) A little later, VanDeVeer (1986, p. 34) formulates a second objection in his analysis of Beauchamp's position by showing that the application of the Harm-to-Others-Principle to cases of unintentional or unwitting self-harm represents mingling this principle with the principle of paternalism. So if Beauchamp were to stretch the Harm-to-Others-Principle in the necessary way, he would have to interpret this principle so that it contains "elements of paternalism" (*ibid.*). A disassociation of the two principles would thus be impossible and the point of Beauchamp's whole strategy therefore missed.

This objection goes astray for two reasons. For one thing, according to Beauchamp, in the cases concerned there is intervention in nature which on no account exhibits paternalistic action, even if this natural process is at the same time conduct or an action by B. But above all, Beauchamp's strategy rests on counting only cases that can be justified *solely* with recourse to the principle of paternalism as paternalistic action. Thus it is quite consistent for an action directed towards the

Harm-to-Others-Principle also to serve the wellbeing of B. For his argumentation, Beauchamp does not need such a strict independence of the two principles as is assumed by VanDeVeer's second objection. So such an intervention is not paternalism because it does not result from a competent decision or action by B. VanDeVeer's objection is therefore either petitious, because it presumes his own broad definition of paternalism, or it misses its target because VanDeVeer disregards the fact that though the intervention in question is directed at B's wellbeing, it is not a case of paternalistic action. Not every intervention in B's conduct or action that is directed at B's wellbeing is paternalistic.

### 8.3.4 *A Possible Basis for Beauchamp's Argumentation*

Both Beauchamp and Feinberg emphasize that in cases of the Mill type, in cases of weakness of will or in conditions of extreme cognitive and emotional confusion, the decisions, actions and conduct of persons should not be regarded as expressions of their "real self" or "real will" (Feinberg 1986, p. 14). It is not a matter of their "real choice" (ibid.), but in fact a threat to their "autonomous self" (ibid.). But what does it mean when Beauchamp and Feinberg say that the decisions are not "truly" those of the person, but can be treated as factors that are "external to his will" (ibid.)?

Clearly, these events, as long they are mental events, also belong to the unity of this mental subject in a causal-functional sense.<sup>37</sup> If this were not the case, then these events could not steer the person's conduct. The description by Beauchamp and Feinberg is thus not directed at this level of the subject's unity, but at the evaluative dimension of the self-consciousness. Such terms as "truly" or "real" are to be understood as valuing expressions which refer to two things. For one thing, these mental events are about first order wants not accorded with an "*identification with*" through the person. For another, they display little or no coherence with the ideals and beliefs constituting the personality of this person.<sup>38</sup> Thus, even if these events might be internal in a causal-functional sense interpreted from the observer perspective, from the participant perspective they are interpreted as external moments in the psyche of this person, which do not match her personality. If one understands Beauchamp's analysis in such a way that the harming events of the personality

---

<sup>37</sup>For the attempt to compass the unity of a self in a frame of naturalistic theory, i.e. exclusively in the observer perspective, cf. for example Brook (1994). A comprehensive theory in which both aspects of consciousness (the naturalistic unity of self and the evaluative level of personality) are tied to each other is developed in Hurley (1998). Hurley also differentiates between various forms of the distinction internal versus external, since this distinction has a different meaning in the personal perspective tied to the participant perspective and the sub-personal naturalistic perspective coupled to the observer perspective respectively.

<sup>38</sup>This distinction of two forms of 'mineness' (Jemeinigkeit) is not identical with the differentiation between the consciousness of the demarcations of self and the subject's action knowledge as proposed by Graham and Stephens (1994). The two differentiations are independent of one another and have their own right for various diagnostic or theoretical purposes.



interpreted as natural events are external, then his whole interpretation becomes plausible. This applies above all to such cases in which a person just conducts himself or, without further reflection, cultivates a primary reason that does not match her 'real' ideals or other convictions. In this way, actions in a description of a person which are intended can be treated: The inherent primary reason for these actions remains on the level of first order wants and in these cases contains no moment of valuing self-to-self conduct (the identification with) that is characteristic for persons and constitutive for personal autonomy. However, one must differentiate the Ulysses contract cases I described in the previous chapter from those cases which can be interpreted according to Beauchamp's model and thus do not fall into the domain of paternalistic action. When one speaks in such contexts of the minimal competent decision as not being "truly" that of the person or expression of her "real will", then these characterizations should be interpreted otherwise (for this other use cf. Kleinig 1983, p. 67 ff.). It is therefore necessary to keep the two meanings of this expression apart. In the way in which Beauchamp uses e.g. "not truly his own", it means that there is no "identification with" and that there is no coherence with the personality as such. In contrast, in Kleinig the same characterization means that though there is identification with the decision, this rests on ideals and beliefs which are barely anchored in the respective personality (little coherence, no stability etc.; cf. Kleinig 1983, p. 67).

The outcome of this analysis of Beauchamp's argumentation thus leads to the following result: Events which fall under his description can be interpreted with the Harm-to-Others-Principle and are therefore not cases of paternalistic action. But at that moment when the decision or action of the respective person is viewed as minimally competent, they belong to paternalistic actions and must be justified with the principle of paternalism when the intervention is primarily in the interest of that person's wellbeing (I will deal in more detail with the cases envisaged by Kleinig in the fourth section of this chapter). Beauchamp (1995, p. 1916) himself points out that B's decision must be adequately competent for the intervention by A to count as paternalistic action. Since he also concedes that the "degrees of control and voluntariness rest on a multi-level continuum" (Beauchamp 1977, p. 76), there is, in his opinion, a continuous passage between such interventions which must be counted as paternalistic action and such which are not to be regarded as paternalism.<sup>39</sup> This extremely plausible admission is, however, not an objection against the decisive

---

<sup>39</sup>This interpretation is contradictory to the analysis by Häyry (1991, p. 65 f.), who deduces from the fact that competence accumulates continuously and that therefore the cases meant by Beauchamp and Kleinig merge that "there is no clear-cut separation between cases of hard and soft paternalism" (ibid. p. 66). Admittedly, I share their opinion that in the classification it depends at all events on the concrete situation: "So much depends on the particular circumstances and relationships between people that no precise and concise overall measuring methods can be spelled out anyway" (ibid. p. 73). This basic particularistic insight does not however preclude that there are also clear cases. But above all, the trained perception of such a situation does not depend on such measuring systems, even if such objectivizations, whether through measuring processes or the specification of principles, are indispensable for the level of justification; cf. Quante and Vieth (2002).

consequence of Beauchamp's argumentation: to waive the distinction between weak and strong paternalism.

To summarize: There is a satisfactory background theory to the way in which Beauchamp deals with the Mill cases. It is vital to distinguish between the evaluative self-to-self conduct of persons and a purely causal-functional model of the unity of self. On the basis of the theory of personality I have provided, Beauchamp's argumentation strategy can be kept up plausibly. Since this has the additional advantage of limiting the question of the permissibility of paternalistic action directly to such cases for which substantial ethical differences of opinion are on hand, for one thing, a further reason ensues for adopting the conception of personality developed here (cf. Chap. 5). For another, in this way one obtains good reasons for accepting Beauchamp's strategy for the definition of paternalistic action.

## 8.4 Personality as Reason for Justification

After the problem of the definition of paternalism has been at the forefront of the previous sections, the aspect of the justification of paternalistic interventions shall now be discussed.<sup>40</sup> In the literature, the main distinction is made between a consequential strategy of justification geared to various models of consent and one to the principle of doing good (cf. Beauchamp and Childress 1994, p. 279 ff. and VanDeVeer 1986, chap. 2 and 3). Kleinig (1983, chap. 3) in addition introduces three further strategies of justification which he calls "the argument of the future selves" (ibid. p. 45), "the argument of interconnectedness" (ibid. p. 39 ff.) and "the argument of personal integrity" (ibid. p. 67 ff.). In accordance with the superordinate questioning of this study, I will limit myself below to a strategy of justification which is launched on the level of biographical identity and personality and thereby brings together Kleinig's argument of personal integrity with a version of the argument of reciprocal solidarity (Sect. 8.4.1).<sup>41</sup> After that I will discuss why anticipated consent is relevant for the justification of paternalistic action. In this context, the difference between conservation of the original personality and the generation of a new one will prove to be an ethically relevant distinction (Sect. 8.4.2).

---

<sup>40</sup>I am thereby limiting myself to the weaker variant, whereby paternalistic interventions are ethically admissible. However, I will not discuss the question of whether they are moreover possibly even ethically imperative; cf. Fulford (1989, p. 189 f.). From the ethical viewpoint, this involves a continuous spectrum of aspects needing to be weighed up. If, in a case, the wellbeing and integrity of the patient bear great weight, then a paternalistic intervention in his minimally competent decision or action can be ethically imperative.

<sup>41</sup>I have already rejected the "argument of the future selves", since it rests on an implausible conception of personal identity.

### 8.4.1 *Personality-Based Justifications of Paternalistic Action*

Mostly, only two strategies of justification are distinguished: a consequentialist one that is geared to the wellbeing of the person in whose competent decision or action there is intervention and a strategy which, through the application of various models of consent, tries to justify such interventions. Supporters of both camps justify their own strategy by exposing the weaknesses of the other. Thus e.g. Beauchamp (1995, p. 1916) rejects consent-based justifications because the target of paternalistic action is not respect for the patient's autonomy but his wellbeing. In his opinion, such a justification only causes confusion and does not clarify the questions. Conversely, VanDeVeer (1986, chap. 3) criticizes certain consequentialist strategies of justification, but at the same time also shows that whether a strict contradiction to consent-based justifications actually arises depends on the exact interpretation of wellbeing. One must agree with two points made by critics of consequentialist justifications. If such a justification is directed not at individual wellbeing but at the sum total of wellbeing, the autonomy of the person is too little respected. The same applies when only experiential interests are encompassed by wellbeing. However, it is a different matter when one avoids such errors. Thus Beauchamp e.g. sketches the following consequentialist justification:

Interventions are justified only if: no acceptable alternative to the paternalistic action exists; a person is at risk of serious harm; risks to the person that are introduced by the paternalistic action itself are not substantial; projected benefits to the person outweigh risks to the person; and any infringement of the principle of respect for autonomy is minimal (Beauchamp 1995, p. 1917).

It should be clear without further elucidation that such a consequentialist justification is on the one hand indebted to the principle of paternalism (orientation toward and justification through the patient's wellbeing) but on the other hand allows implicitly for personal autonomy. This is also confirmed by the following observation:

The problem of medical paternalism is the problem of putting just the right specification and balance of physician beneficence and patient autonomy in the patient-physician relationship (Beauchamp and Childress 1994, p. 284).

Now, it might be the case that "beneficence alone justifies truly paternalistic actions" (ibid. p. 281). But that does not mean that consent-based justifications should not play a role in the trade-off and that it is better "to keep autonomy-based justifications at arm's length from paternalism" (ibid.). This only applies to the list of conditions, but not to the various reasons which should be taken into consideration in the justification of a concrete paternalistic action. Even if I generally agree with Beauchamp that the primary intention of the paternalistic intervener must be the wellbeing of whoever is being treated paternalistically, I do not therefore believe that the recourse to a conception of consent is wholly useless in the question of justification (cf. Sect. 8.4.2). In the end, the sensible variants of the two models of

justification will converge.<sup>42</sup> For this reason I will now concentrate on sketching a personality-based strategy of justification which cuts across this argument, but lies directly in the line of questioning of my investigation.

With the “arguments from interconnectedness” (Kleinig 1983, p. 39) and the “argument from personal integrity” (ibid. p. 67) Kleinig distinguishes two further justifications, of which he regards the latter as the “most promising (,) argument for paternalism” (ibid.). I agree wholeheartedly with this view, since the personality-based strategy of justification is in a position to integrate both the consequentialist and the consent-based conceptions. However, in contrast to Kleinig, I do not think it is necessary to distinguish between the arguments for interconnectedness and that for personal integrity, as my conception of personality embraces the essential aspects of the former strategy. Supporters of liberalism usually stand out through their individualistic conception of personal autonomy in which the individual is regarded as an essentially asocial being, defined independently of social life, and entering into social relations only through choice (Kleinig 1983, p. 39).

In contrast, in justifications of paternalism which make use of “arguments from interconnectedness” (ibid.) it is emphasized that a human person is constituted essentially through social contexts. In the strongest form of this strategy, the evaluative self-relation that is manifested in the personality is to be understood as a social structure of recognition. In the weakest form, it is only pointed out that a person cannot survive without a suitable social environment. In a third variant distinguished by Kleinig, a social context is needed to make possible the “flourishing” (ibid. p. 41) of a person, the actualization and realization of her potential to become a personality. In this investigation I have proposed a conception of personality and personal autonomy which implies the first, strongest form. However, my considerations are not affected by the criticism Kleinig applies to the “arguments from interconnectedness” (ibid. p. 45). This is because I do not deduce from the social constitution of the personality (ontological holism) that there is also an ethical primacy of social i.e. of general wellbeing above the wellbeing and autonomy of the person (ethical holism).<sup>43</sup> Without further arguments, the discernment of a holistic conception of personality that is justified at the ontological level cannot be harnessed for the justification of paternalistic interventions. Since I have not introduced such an additional argument and neither do I believe that one can develop a plausible argument here, the liberalist criticism does not apply to me; rather, I can agree with the basic tendency. But even if ontological holism does not necessarily lead to ethical holism, it can be that during the weighing up in concrete cases of paternalistic interventions, the social constitution of the personality takes over a justifying function (thus also

---

<sup>42</sup>The remaining differences will either be traced back to deviant metaethical and ethical premises or have to be understood as the expression of diverse emphasis.

<sup>43</sup>I differentiate these two levels, which one must keep apart in the development of a holistic conception of personality, in Quante (1997b). Liberalists tend always to attribute to a holistic conception in the ontological sense the thesis of ethical holism, according to which it is justified to override the wellbeing of an individual in favor of general social wellbeing. However, this argumentation rests on a “non sequitur”; cf. Quante (2001c).

Kleinig 1983, p. 45). In this function one can, however, integrate the social constitution straight into the personality-based strategy of justification, so that one does not have to distinguish two different forms of justification of paternalistic action.

If one justifies a paternalistic intervention with recourse to the person's integrity, then one wants to promote her wellbeing in the sense of the "restoration or preservation of her integrity" (Kleinig 1983, p. 60). Different from the 'Beauchamp' cases, the wants and convictions underlying B's competent decision belong to B's personality:

They may belong to them in the way that moments of weakness belong – something for which they can be held responsible, but not something they desire to own further (ibid. p. 60).

Without mentioning the conception of second-order volitions, this quotation clearly confirms that Kleinig has in mind such decisions as are both competent and also the subject of a negative *identification with* on the part of B. With this, these cases fulfill a necessary precondition for a paternalistic intervention. But even if, in contrast to Beauchamp's envisaged cases, these wants can be counted as part of B's personality, they are less well anchored in it, less stable, and jeopardize possible intentions regarded by B to be seen as central elements in his biography:

Where our conduct or choices place our more permanent, stable, and central projects in jeopardy, and where what comes to expression in this conduct or these choices manifests aspects of our personality that do not rank highly in our constellation of desires, dispositions, etc., benevolent interference will constitute no violation of integrity (ibid. p. 68).

Here it is important that, for one thing, "highly" refers to the position within B's personality i.e. his self-evaluation of the aspects in question and not to an intersubjective ranking (this is relevant e.g. when "biographical" and "experiential interests" clash). It is also important that stability or the long term do not alone suffice to define the core of personality. As Macklin (1983) establishes, a person with a possibly life-long history of mental illness can display a structure of stable values and beliefs. If such a patient then begins to tackle his illness it is inappropriate to prevent these attempts in favor of his previous values and convictions. It would be presumptuous to want to specify a general rule. I do indeed believe that it is as a rule quite possible to distinguish between a developed personality and a merely stable and persistent pattern that is the result of a mental illness. My conception of personality and personal autonomy contains two elements – the condition of rationality and the evaluative self-relation – which prevent merely persistent and stable patterns from having to be valued up front as an expression of personality. However, this does not rule out that there can be borderline cases which will be difficult to decide. This is, however, not decisive for my present argumentation, for, on the one hand, the relevance of the conception of *personal* autonomy, called "authenticity" in Macklin's nomenclature, is not challenged in this weighing up, but on the contrary actually enlisted (cf. Radden 1996, p. 150 f.). On the other hand, it becomes clear that in this context, intersubjective standards (the concept of rationality or illness) must also be applied unavoidably. Since I view the social dimension as a constitutive element of personality, this is also consistent with my general conception.

In contrast to Jennifer Radden, I consider a justification of paternalism directed at personal integrity to be neither the result of a purely metaphorical expression of an “authentic self”, nor a theory defined by the questionable metaphysics of numerically different selves. If one accepts the analysis of personal identity presented here, then one can identify the personality as the relevant factor on which our ethical considerations are based in this context, without lapsing into misleading metaphysics of personal identity or into mere imagery. For this reason I agree entirely with Jennifer Radden (1996, p. 154) not wanting to justify paternalistic interventions either on the basis of such a “metaphysical theory of successive selves”, or on the basis of purely “figurative usage” (ibid.) of the parlance of an authentic self. But it is decisive that these two possibilities are not entirely exhaustive. The option of justifying paternalistic interventions for the conservation or restoration of personal integrity on the basis of the concept of biographical identity proposed here remains untouched by her criticism.<sup>44</sup> We have at our disposal a concept and an understanding of personality which factually guides our ethical intuitions regarding the permissibility of paternalistic interventions and can also be developed explicitly as a basis of justification. If one does not reduce the patient’s wellbeing to hedonism or the avoidance of pain, but recognizes the realization of an own personality as the central element of a successful personal life, then such a personality-based justification of paternalism is also suitable for integrating the appropriate aspects of the consequentialist strategy of justification. As will now be shown, the same also applies to the consent-oriented strategy of justification, since there is a close bond between the concept of personal integrity and the model of anticipated consent (cf. Kleinig 1983, p. 60).

### 8.4.2 *Anticipated Consent*

In the following, I would first like to distinguish briefly between two different models of consent and examine them for their compatibility with the above proposed personality-based strategy of justification. After that I will demonstrate by means of a differentiation of cases the ethical relevance of the difference between conserving what is there and installing a new personality.

---

<sup>44</sup>The limits of such a justification have, however, been reached when the marginality of wants and convictions which are to be overruled paternalistically cannot be made plausible (think of Faust’s ‘two souls’). Different from such cases of massive psychopathies, which due to absence of competence do not count as paternalistic actions in my theoretical framework, an intervention in such a torn personality, whilst being paternalistic, would not be justified in my view. The coherence of personalities permits tensions, so that one cannot intervene through recourse to integrity. The cases of actions arriving at installing autonomy or to help in the development of an own personality (e.g. upbringing) or to restore the capacity of making competent decisions discussed by Kleinig (1983, p. 67 f.), Fairbairn (1995, p. 185 ff.) or Harris (1985, p. 203) do not count as paternalistic action according to my understanding.

### 8.4.2.1 Forms of Consent

Various forms of consent-based justification of paternalism can be distinguished according to the model of consent on which the justification is based. Kleinig (1983, p. 55 ff.) distinguishes between the ‘real-will’-argument, the foregone, the anticipated, the subsequent and hypothetical rational consent. Foregone consent has already been discussed in detail in the context of the Ulysses contracts and can therefore be left aside at this point. Representatives who work with the ‘real will’ model start from the assumption that a person’s current wants and convictions are not the authentic expression of her ‘real’ personality. Alongside ideology-critical models, such an argumentation can also be understood as resting on the difference between a purely positivistic, empirical and, divorced therefrom, transcendental I.<sup>45</sup> In both cases the concrete personality is distorted. But in the cases.

where the real will is determined by reference to a person’s well-established life-plans, these tendencies are considerably diminished (Kleinig 1983, p. 59).

One then supports a strategy of justification which is built up on the concept of personality. For, in this approach, an evaluative concept of biographical identity which is manifested in a person’s concrete lifestyle is taken as a basis.

The philosophical construction of hypothetically rational consent is also out of the question for a personality-based strategy of justification, as here the concrete personality also passes from view. This can happen in two ways. If the model of hypothetical rational consent is understood in the sense of logical-ideal rationality, then one subjects the factual ideals of a person to a hypothetical process of correction in which consistency is the benchmark. In this way, the empirical personality is indeed conserved in the sense of a set of ideals, but, due to the unrealistic demands of logical-ideal rationality, the real composition of human persons and their personalities is masked.<sup>46</sup> Furthermore, if one integrates an element of criticism of ‘unreal’ or estranged ideals and wants into the model of hypothetical rational consent, then one removes oneself even further from the concrete personality and applies a moral ideal. That such a justification of paternalistic interventions cannot be rendered plausible surely needs no further discussion. On the contrary, care should always be taken that the principle of respect for autonomy is based on an evaluative, but at the same time also realist conception of personality (cf. Childress 1997, p. 60).<sup>47</sup> Thus

---

<sup>45</sup>If one takes the empirical I as merely the sum total of the factually psychological elements of a subject, then this lacks the evaluative and critical-normative dimension of autonomy; cf. Wolf (1990, Chap. 2). In contrast, if one conceives the real self on the basis of e.g. fundamental ontology, an essentialist philosophy of history or a normative theory of the noumenal self, there is a risk of ignoring a person’s real wants in favor of a philosophical construction; cf. for an analysis in which aspects of an idealist and a social conception of personality are combined, Bosanquet (1910).

<sup>46</sup>This kind of usage of the hypothetical rational consensus also contains the implausible premise that one can segregate ideals and convictions, or questions of value and fact cleanly (i.e. analytically); cf. on the effects of these models on the self-image of the biomedical ethician, critically Caplan (1992, p. 33 ff.) and Brock (1993, Chap. 2).

<sup>47</sup>I have suggested to take this into account via choosing a default-and-challenge strategy in analyzing personal autonomy; cf. Quante (2013b).

there are two good reasons for being cautious in the use of justifications of paternalistic interventions which refer to models of consent. Firstly, such a reference should only be understood as the naming of relevant aspects which must be taken into consideration in the concrete case. The recourse to consent should not go as far as the 'wit' of paternalistic action being the confection of consent. In this respect the criticism by Beauchamp and Childress should be conceded. (1994, p. 280 f.). Secondly, it is obvious that not all established models of consent are appropriate for personality-based justification. However, it does not follow from either that the recourse to consent would be totally inappropriate for the justification of individual paternalistic interventions. If one views the two models of anticipated and subsequent consent still remaining, then for one thing they have in common that there is an empirical possibility of review, and for another, that the paternalistic agent exposes himself to the risk of falsification. Indeed, it seems to me important to introduce a definitional difference between anticipated and subsequent consent. In the following it shall apply that anticipated consent requires that the personality of the paternalistically treated patient B be the same at the time of the intervention and the time of consent. If one supports the thesis that the assumption of future consent by B is adequate in every case, then one cannot preclude a case of a massive manipulation of B which results in a new personality. But this, as the discussion in Kleinig (1983, p. 60 ff.) and Carter (1977, p. 136) shows, is exactly what should not be included (an exception is such cases in which a massive personality change can be expected even without the paternalistic intervention). For otherwise, the notion of subsequent consent is simply 'question begging': Then, in addition to the paternalistic intervention, there is still brainwashing, which lets the victim of the intervention consent afterwards. It is clear that this cannot be an acceptable basis for the justification of paternalistic action. To take this into account, I therefore understand anticipated consent as the plausible assumption that B, under retention of his personality, will later consent to the intervention. It is thereby not the anticipated consent that is the aim of the paternalistic action, but rather, B's wellbeing. Anticipated consent serves as a justifying aspect because it signifies that B's integrity will be conserved and thus his personal autonomy only restricted or temporarily subordinated by the paternalistic action.

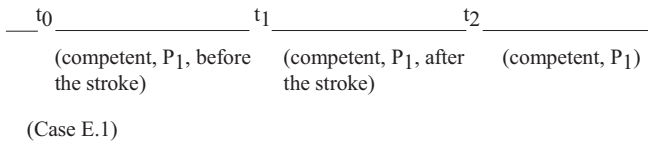
#### 8.4.2.2 Types of Cases

So as to show that the difference between the conservation of the former personality and the generation of a new one is ethically relevant, I would like to distinguish three cases:

- (a) In the first case, patient B has suffered a stroke and is subsequently so impaired in his subjective quality of life that he does not want to have any life-prolonging measures carried out. B is continuously capable of making competent decisions and also understands that his decision (in the interval [t1-t2]) will lead to him dying soon. But since he cannot come to terms with his new life situation (nega-



tive identification with), he consciously and willfully refuses life-prolonging measures.<sup>48</sup> Now in these cases experience teaches that patients in B's situation identify positively with their new life situation normally within a certain time without this being accompanied by such a massive change in their values and convictions that one would have to speak of a new personality. Furthermore, patients in B's situation subsequently consent to a paternalistic intervention in most cases (after point in time t<sub>2</sub>). Shown diagrammatically, the first case looks like this:



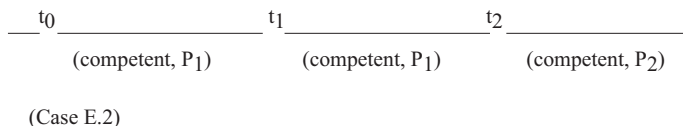
Now, if one does not accept the refusal of treatment, then this is a case of paternalistic action. In a different variant of the same case it is not a matter of life-conserving or life-prolonging measure being refused by B. Now B is refusing the implementation of rehabilitation measures of which one is empirically certain that they can considerably improve B's future quality of life. But B cannot be motivated to cooperate as he views his new life situation as overall not worth living. Here there is also reliable evidence that patients who have been 'pressured' into such measures for a while against their will, later consent to this paternalistic intervention (cf. Caplan 1992, p. 247 ff.).

In both variants of E.1 there is a lot going for a paternalistic intervention being justified. At least for a certain period of time, in which normally the re-orientation of the patient takes place, it is ethically acceptable, in view of the remaining quality of life or that being regained, to overrule B's competent decision paternalistically. The idea of facilitating an adapted personality is thereby directive; this case is therefore not identical with the situation in upbringing in which one wants to spawn or improve the capacity for autonomous decisions. However, if B's refusal remains stable beyond the normal period of time, then such an intervention loses its eligibility, one must now assume that the anticipated consent will not eventuate. B's personality is then presumably constituted in such a way that a positive identification with the new life situation is not compatible with his personal integrity.

- (b) The ethical weighing up turns out differently in the second case. Here the intervention in B's competent decision and the implementation of a medical measure would lead to such a massive change that one would have to speak of the installation of a new personality (this can e.g. be the case after a stroke or in the area of mental illnesses).

---

<sup>48</sup>The interval [t<sub>1</sub>-t<sub>2</sub>] only begins when it has been established that B has made his decision knowingly and deliberately. Before that, an intervention is not paternalistic action according to my definition.



It is expedient to differentiate two sub-cases. In the first one (E.2.1) refraining from the intervention leads to the death of B. In the second one (E.2.2), in contrast, the original personality is retained. If in E.2. one decides in favor of the paternalistic intervention, then one cannot be sustained either by anticipated consent or by personality-based justifications, due to my definitional differentiation. In my view there is no sufficient reason in the former case (E.2.1) to intervene paternalistically. Though this would be justifiable in a purely consequentialist conception which uniformly heeds only B's wellbeing (regardless of the personality change). But such a conception does not adequately respect personal autonomy.<sup>49</sup> B expresses his negative valuation explicitly and respect for his personal autonomy and personality should therefore tip the scales here. If one defines B's wellbeing with the inclusion of his personality, then such an intervention is not minimal and therefore no longer tolerable. The same applies to E.2.2, a case in which a medical measure spawns a new personality, whereas without the intervention the previous personality is retained. If e.g. a sex offender resists an intervention for a personality change, then such an intervention cannot be justified paternalistically. It can either be justified with recourse to the Harm-to-Others-Principle if the 'patient' in question is a danger to others, or one will have to justify it through general considerations of social desirability as regards the new personality.<sup>50</sup> However, there is no paternalistic justification for these interventions.

- (c) In the third case, B wants to induce a personality-change in himself through medical measures (e.g. through drug consumption). A paternalistic intervention (prevention of drug consumption) results in B's actual (used in the strictly modal sense) personality being retained. So the alternative consists in either conserving the actual personality or allowing a new personality to be generated. There is paternalistic justification if B's want represents a considerable risk for his wellbeing. In such a case one can possibly reckon with his later consent as soon as B has learnt how to calculate the risks properly. Also possible, if there is no such considerable risk, is a paternalistic intervention which aims at making it clear to B that his own valuation of the new personality is misguided. However, as soon as such an intervention is motivated solely by the social desir-

<sup>49</sup>This could also be justified on the misguided basis of a metaphysics of "successive selves", in which the new personality would be counted as a new individual. However, presupposing such a conception the Harm-to-Others-Principle would be appropriate and no paternalistic intervention would be on hand.

<sup>50</sup>It is not my aim to discuss here whether such interventions can be ethically justified, although in both cases one must be very cautious with the intervention in personal autonomy; cf. Walker (1994).

ability of the actual personality, this is no longer paternalistic action.<sup>51</sup> In all other cases an intervention in B's decision must be justified differently, whereby I am leaving it open here whether there can be a nonpaternalistic justification which prohibits B from precipitating a personality change.

My differentiation and valuation of these three cases is based for one thing on the thesis that the difference between the retention of the actual and the installation of a new personality is seen intuitively as being ethically significant and in all should be. It follows from my definition of paternalism that in the installation it can only involve a paternalistic intervention when the patient's wellbeing is the primary reason for the intervention. In borderline cases it is possible that the quality of life of the patient with the new personality is so good that a paternalistic intervention can be justified. But in general I am of the opinion that respect for the personality and integrity of a patient should tip the scales. That is why, in my view, interventions which include the installation of a new personality, normally cannot be justified paternalistically.<sup>52</sup> Admittedly, these cases lead my ethical intuition into a border area. What is more, the transition between the further development and change in a personality is blurred. Real cases are correspondingly confusing and continuous. Nevertheless, my general thesis, that the concept of personality is factually operative here and can take over a justifying function, has proved durable. In many cases it can be proven not only conceptually, but also empirically that through paternalistic interventions a patient has been placed in a position to retain his personal integrity or retrieve his personality (cf. Hope 1994).

## 8.5 From Paternalism to Cooperation – A Brief Conclusion

The strategy of argumentation of this last chapter can be summarized thus: medical paternalistic action can be understood as the intervention in the competent decision or action of a patient B, which is carried out to promote B's wellbeing. Herewith, allowance is made for the attempt to define medical paternalism so that only those cases are included with respect to which there is substantial ethical dissent. This dissent can be reconstructed as the question of the emphasis on doing good and the principle of respect for autonomy. However, what seems at first glance – and

---

<sup>51</sup> Alongside the case of personality-changing drugs, an aspired sex change is also perhaps a case of this third type. However, here the intervention mostly serves to clarify whether B is certain enough and adequately informed about the various aspects of his desire (and is thus not paternalistically motivated). However, I do not wish to allege that a sex change results in a personality change in every case. The more the orientation to the other sex formed B's personality before the sex change, the more probable it is that that is not the case.

<sup>52</sup> The possibility that "it could be the case that undergoing preference change against one's will can improve one's life all things considered" is also conceded by Arneson (1994, p. 74 f. fn. 27), whereby one must not interpret "life" as biography. His ethical valuation of the permissibility of such an intervention turns out to be less clear cut, however.

unfortunately also in many contributions to discussions in biomedical ethics – to be a simple antagonism proves on closer scrutiny to be a more complex structure. The reason for this is that differentiations are essential both on the side of the principle of beneficence and on that of the principle of respect for autonomy. (i) The *wellbeing* of a human person cannot be reduced to her experiential interests, but must also include enabling her to lead her own life as a person and ensure her personal integrity. Contrary to a crude hedonistic conception, every appropriate ethics of the successful personal life must be based on a comprehensive conception of the good, which also contains free space for individual concepts and articulations of this good. Respect for the integrity of the personality stands for this aspect of the good. (ii) Contrary to a widespread tendency, even human *autonomy* must not be reduced to the ability to form and exercise free decisions or actions (cf. Miller 1981). The dimension of personal life which has been named in this investigation of personal autonomy must be recognized as a central element of our concept of autonomy and our way of realizing autonomy as human persons and incorporated in the principle of respect for autonomy.

If one bolsters the discussion of medical paternalism with this more complex interpretation of wellbeing and autonomy, it transpires that firstly there are three kinds of paternalistic action: Interventions in competent decisions, interventions in autonomous actions and interventions in personal autonomy. Secondly, the multi-layered wellbeing of human persons allows for the following two cases: (i) Interventions in competent decisions or autonomous actions can be understood as paternalistic action, since they are interventions in B's autonomy for the benefit of B's wellbeing. Such interventions are justified under recourse to personal autonomy which, however, is cited in its constitutive function for wellbeing. For this reason it is a question of paternalistic action which at the same time can be and is justified by personality-based reasons. (ii) Cases in which there is intervention in personal autonomy not only in the competent decision or autonomous action, but also in personal autonomy, likewise count as paternalistic action.<sup>53</sup> In this constellation, the aspects of B's wellbeing which do not depend on the identity of his personality are placed higher than the conservation of his biographical identity. In my view, these interventions also qualify as medical paternalism. But as a rule they cannot be justified because the disregard of personality is so grave that the achievable gain in miscellaneous wellbeing can hardly compensate it.

If one allows for the increasing significance of the rating of autonomy in our culture and correspondingly also the weight of the principle of respect for autonomy in biomedical ethics, then it becomes understandable why the personality-based strategy of justification of medical paternalism is the most promising. Although justified actions are thereby rightly understood as paternalistic action, this justifica-

---

<sup>53</sup>For simplicity's sake, I am here assuming that one can only intervene in B's personal autonomy paternalistically by intervening in competent decisions or autonomous actions. This must then be colligated so broadly that the limitation of the capacity for competent decisions and autonomous actions through the administration of drugs or unnoticed and undesired manipulations of B's convictions and ideals can also be subsumed under these.

tion of paternalistic interventions at the same time also pays respect to the autonomy of the patient. This becomes possible through the multilayering of human autonomy and human wellbeing. Acknowledging this not only allows a better grasp of our ethical intuitions and the complexity of medical reality but also enables implementation to be improved.

Despite the recognition of the internal complexity of patient wellbeing and patient autonomy, due to which multilayered constellations can be encountered, I assume in my deliberations that the crucial point of paternalistic interventions rests on the presumption of a contradiction. To contradict the impression that therewith the general thesis of a thoroughgoing contradiction between personal autonomy and medical action should be defended in all areas of medicine, I would finally like to point out briefly that the concepts of personality and personal autonomy I have developed are also suitable for developing a model of physician-patient relationship based on cooperation.<sup>54</sup> In such a model of cooperation, one starts on the assumption that the medicating and medicated compile a “collaborative, shared plan for the future” (Spitz 1996, p. 239), in which physician and patient define the targets together. The physician’s role is not thereby reduced to merely supplying the patient facts for his value decisions, but rather, he will also have to contribute his own value decisions. As long as this happens openly and explicitly, such a line of action is perfectly acceptable: “It is a mistake to believe that the only appropriate role for the physician is the delivery of facts” (Brock 1993, p. 69).<sup>55</sup> Moreover, if one in doing so considers the complexity of human autonomy, then many cases of apparent paternalism can be understood as autonomy-conserving measures of a “delegated autonomy” (Agich 1993, p. 43). A person can delegate decisions for individual actions or decisions regarding certain areas of life to other persons. Her autonomy is conserved because of such a “second-order choice” (VanDeVeer 1986, p. 205). Personal autonomy permits the partial waiver of autonomy in the framework of a kind of division of labor. Such shared decisions, processes of decision-taking and developments of objectives are not to be understood overall as paternalistic action (in my sense), since they do not represent an intervention in the autonomy of the patient.<sup>56</sup>

However, the recognition of the multilayering of personal autonomy not only allows the development of such a cooperative physician-patient relationship, but can also increase awareness that e.g. care is taken to respect the partial autonomy of the patient in long-term care. If one no longer understands autonomy as merely a

---

<sup>54</sup> Such a model is above all necessary in the area of the treatment of the chronically sick or patients with permanently only limited competence, since the treatment target cannot here consist in the restoration of the patient’s complete autonomy; cf. Caplan (1992, Chap. 14 and 15) and above all Agich (1993).

<sup>55</sup> Such a cooperative model can thereby find its limits in an insoluble conflict of values or incompatible convictions between physician and patient. Then, if the Harm-to-Others-Principle cannot be implemented, the question of paternalistic intervention arises, whereby the respect for personal integrity should normally tip the scales; cf. O’Neill (1984, p. 176 f.).

<sup>56</sup> To characterize this difference, the concept of “parentalism” (Kultgen 1995) has been suggested. This should not only eliminate the sexist connotation of paternalism, but also express that no paternalistic action is existent here.

one-sided competence for decisions, but also as the greatest possible conservation of personal integrity then one can also conserve the partial autonomy of such persons with only limited competence as far as they are capable of practicing it.<sup>57</sup> Personal autonomy is multilayered and has many facets, as does the personality of a patient. The ethics of respect for personal autonomy with its emphasis on the integrity of human persons is in a position to make these facts visible and declare them ethically relevant. The development of an appropriate ethics of rehabilitative medicine or long-term care that bases on a rich and sophisticated concept of personality and personal autonomy and which renders paternalistic action as far as possible superfluous through cooperation, mutual recognition and an “environment of trust” (Caplan 1992, p. 279), can no longer be tackled in this study. In my view, it can be developed and justified on the basis of the conception of personality developed here. In this context, the explicit recognition of the biographical identity of human persons will also be able to grasp our factually effective ethical intuitions better and justify more appropriate practice. I agree thereby with James F. Childress (1997, p 62) when he maintains that “respecting persons becomes very complex when their temporality is properly included.” However, this complexity does not seem to me to be too high a price – even at the expense of the elegance and simplicity of the ethical theory – if we obtain for it the basis of a biomedical ethics which can satisfy the needs of human persons.

---

<sup>57</sup> Caplan (1992, Chap. 13–16) and Agich (1993) warn about the ethically unacceptable consequences of unnecessary incapacitation of patients with limited competence that are indebted to the use of a depleted conception of autonomy of decision and agency.

## Chapter 9

# The Interlacing of Persistence and Personality

If one dissolves the assumed uniform phenomenon of the identity of human persons over time into the persistence of the human organism and the biographical or narrative ‘identity’, a series of interrelated questions arises. For one thing, these concern the interrelation of observer and participant perspectives. Since a detailed discussion of these questions is not possible in the framework of this study, I must leave it as a brief outlook here.

In our everyday understanding and experience, the identity of human persons over time presents itself as a uniform phenomenon. This unity is thanks to the nomological character of persistence, underlying the personal lifeform opening up in the participant perspective as a causal enabling condition. Only in actually occurring special cases such as psychopathies or irreversible coma does the complexity and heterogeneity of this *prima facie* homogeneous phenomenon of ‘personal identity’ over time become visible in the manifest image (cf. Quante 2012, Chapter X). In addition, the tensions between the conditions of persistence of the human organism and our concept of person also become tangible in the context of beginning human life, as the public discussions about the appropriate ethical handling of human embryos or human embryonic stem cells show. Within the philosophical debates, the tensions and fissures in this phenomenon are additionally exposed in thought experiments designed specifically for this purpose. These thought experiments have in common almost continuously that they violate the nomological conditions of persistence valid for human beings. In this way a confrontation of three different influencing factors that are interlocked in the normal situations is instigated: now the first-person perspective and our everyday intuitions based on the normal function of human persistence stand opposite the logically possible. The analysis proposed in this study explains why in our manifest image we only rarely need to separate the levels of persistence and personality. It also provides an explanation of why there are no longer reliable intuitions and clear answers in the scenarios induced by the philosophical thought experiments.

Whilst the dimension of personality cannot play a role in the framework of the analysis of persistence for methodological reasons (here the only exception are the considerations on potentiality, cf. Chap. 3), the dimension of human persistence within the participant perspective that constitutes personality comes into focus in several ways.

In order to differentiate these different ways it is helpful to distinguish between ‘by nature’ and ‘natural’ conditions and aspects: *Nature* conditions or aspects are those which are grasped solely in the observer perspective and are regarded only as causal enabling conditions. In contrast, *natural* conditions or aspects are conditions by nature interpreted from the participant perspective.<sup>1</sup> Generally speaking, the level of persistence proves to be a causal enabling and nomological framework condition for the development and shaping of personal life. Inasmuch as it is viewed in its function of causal enabling conditions, it concerns by nature conditions which are indispensable for the formation of personhood and personality. Alongside this there are ‘natural’ aspects of personality such as e.g. facts referring to the body, the basic psychological configuration of the human being or even biological regularities such as growing up, aging or sickness. These general human life conditions, which are in part culturally invariant, form our understanding of personal life in a basic way. Individual properties of the respective individual, such as temperament, individual physical capacities, looks etc. can then play a role that is specific to the respective individual personality. These ‘natural’ specifications, out of which a human being can develop and unfold her personality, have always been grasped through the evaluative participant perspective with regards to their possible function for the development of an individual personality. Only when, in the participant perspective, they are integrated in the respective personality as specifications, are these specifications natural.<sup>2</sup> As natural entities, i.e. grasped and interpreted as entities by nature in the participant perspective, they display an evaluative-descriptive double structure and can thus gain immediate access to ethical argumentations, without there automatically having to be a false naturalistic conclusion on hand.<sup>3</sup>

Different from anti-realist (Rorty 1989) or instrumentalist-constructivist (Dennett 1990, 1992) reconstructions of personality, I understand human persistence and personality in equal measure as *real* aspects of our personal life. This, for one thing, ties in the thesis that reality cannot be understood exhaustively in either the participant or the observer perspective alone. For another, I assume in addition that the *Lebenswelt* made accessible in the participant perspective presented as a purely descriptive-evaluative entity is underlying, as opposed to the reality presented in the purely descriptive observer perspective. The reality made accessible in the observer perspective can be understood as an abstraction from the *Lebenswelt* constituted in the participant perspective (Sellars’s “manifest image” (1991, p. 6 ff.)). But the elements and aspects of this abstracted reality

---

<sup>1</sup> In contrast, I use the term “naturalistic” for philosophical constructs which evolve when elements or aspects which belong essentially to the participant perspective are reduced to the observer perspective.

<sup>2</sup> For the question of the connection between genome and personality discussed in the context of the ethical evaluation of human genetics, this distinction between entities by nature and natural entities is mostly overlooked. The human genome concerns a natural entity through which the personality of a human individual remains under-determined, since the genome is a matter of a causal enabling condition; as an example of a dispute in which this difference is blurred cf. Elliot (1993, 1997) and Persson (1995, 1997).

<sup>3</sup> A realist interpretation of these “natural” phenomena which I assume also approximates the cautiously defended rehabilitation by Siep (2004) of a commonsensical concept of nature, which embraces the normative aspects of ‘cosmos’ and ‘physis’.



are not mere constructs, but form a real, integral part of reality.<sup>4</sup> If, as I here suggest, one avoids the thesis that only one of the two perspectives can comprehend aspects of reality, then the question arises of how the relation of these two perspectives to one another can be defined. Here it is not possible to answer this question, as it concerns a complex problem which requires its own investigation.<sup>5</sup> In the predominant majority of cases there is no conflict between the two perspectives, so that the aspects being developed in them respectively can be understood as aspects of one reality. Only when, through naturalistic encroachments, the attempt has been made to reduce central aspects of the *Lebenswelt* to the observer perspective, do insoluble conflicts arise. So those kinds of naturalistic endeavors, to be found in wide areas of the philosophy of mind, the cognitive sciences and even e.g. in some approaches to evolutionary ethics and epistemology, must be rejected. But it would be an impermissible simplification to want to reduce the question of the relationship between the participant and observer perspective, of the 'manifest' and scientific worldview to a theoretical issue. Just as the interplay of both perspectives shows in our everyday interpretation of the identity of human persons over time, these two dimensions of reality do not simply lie side by side free of contact. In fact there are manifold permeations. This is above all the case because the nomological regularities provide the basis and the framework for understanding ourselves as persons in the manifest image. In addition, there are two other influential factors. For one thing, in our secular culture the (mostly) diffuse belief in an epistemological and ontological primacy of the observer perspective and the natural sciences meanwhile belongs to many people's every day understanding of the world.<sup>6</sup> For another, technology is a breach in the *Lebenswelt* for the scientific world view. Many of the problems being dealt with in biomedical ethics in fact arise because old familiar natural roots of our traditional self-understanding are being undermined by new options for action. This is particularly conspicuous in the fields of human genetics, reproduction technologies, but also in transplantation medicine. New knowledge or possibilities for action in these areas are changing the constants inscribed in our everyday self-understanding and thereby provoking an erosion of our ethical ontology, which manifests itself in irritation and conflicting ethical intuitions.<sup>7</sup> Marginalizing such ethical qualms by pointing out that they are relics of an unenlightened, unscientific understanding of the world and of ourselves does not go far enough in my view. The question of the correct relation of the everyday and scientific world view is not only undecided philosophically, but in fact quite essentially not only a theoretical but also an ethical one.<sup>8</sup> Ultimately, this debate is a matter of how we understand ourselves as human persons and in what kind of world we want to live.

---

<sup>4</sup>Naturalism, which only acknowledges as real what can be grasped from the observer perspective, is contradicted herewith; cf. in greater detail on this Quante (2000a).

<sup>5</sup>I took initial steps on this in Quante (2000a, b).

<sup>6</sup>The fierce debate conducted in Germany over the brain death criterion can be reconstructed e.g. as the question of whether a scientifically anchored understanding of death should take the place of an everyday understanding thereof; cf. Quante and Vieth (2000).

<sup>7</sup>I have analyzed this, using xenotransplantation as paradigmatic example together with Andreas Vieth in Vieth and Quante (2005).

<sup>8</sup>Cf., in more detail, Quante (2000a und b).

# Bibliography

- Ach, J.S. 1993. Embryonen, Marsmenschen und Löwen: Zur Ethik der Abtreibung. In *Herausforderung der Bioethik*, Hrsg. J.S. Ach and A. Gaidt, 71–136. Stuttgart: Frommann-Holzboog.
- . 1999. Leben für Leben. In *Gene, Klone und Organe—Neue Perspektiven der Biomedizin*, Hrsg. R. Paslack and H. Stolte, 125–140. Frankfurt a.M.: Peter Lang.
- Ach, J.S., and M. Quante Hrsg. 1999. *Hirntod und Organverpflanzung*. Zweite, erweiterte Auflage, Stuttgart: Frommann-holz- boog.
- Ach, J.S., et al. 2000. *Ethik der Organtransplantation*. Erlangen: H. Fischer, cop.
- Agich, G.J. 1993. *Autonomy and long-term care*. New York: Oxford University Press.
- Anscombe, G.E.M. 1975. The first person. In *Mind and language*, ed. S. Guttenplan, 45–65. Oxford: Clarendon Press.
- . 1984. Were you a zygote? In *Philosophy and practice*, ed. A.P. Griffiths, 111–115. Cambridge: Cambridge University Press.
- Appelbaum, P.S. 1982. Case studies: Can a subject consent to a ‘ulysses contract’? *Hastings Center Report* 12 (4): 27–28.
- Arneson, R.J. 1980. Mill versus paternalism. *Ethics* 90: 470–489.
- . 1994. Autonomy and preference formation. In *In harm’s way*, ed. J.C. Coleman and A. Buchanan, 42–75. Cambridge: Cambridge University Press.
- Ashley, B. 1976. A critique of the theory of delayed hominization. In *An ethical evaluation of fetal experimentation*, ed. D.G. McCarthy and A.S. Moraczewski, 113–133. St. Louis: Pope John Center.
- Atkins, Kim. 2008. *Narrative identity and moral identity*. New York: Routledge.
- Atkins, Kim, and Catriona Mackenzie, eds. 2008. *Practical identity and narrative agency*. New York: Routledge.
- Ayers, M. 1993. *Locke*. London: Routledge.
- Bartlett, E.T., and S.J. Younger. 1988. Human death and the destruction of the neocortex. In *Death: Beyond whole-brain criteria*, ed. R.M. Zaner, 199–215. Dordrecht: Springer.
- Battin, M.P. 1994. *The least worst death*. New York: Oxford University Press.
- Baumgartner, H.M. et al. 1997. Menschenwürde und Lebensschutz: Philosophische Aspekte. In *Beginn, Personalität und Würde des Menschen*, Hrsg. G. Rager, 161–242, Freiburg: Alber.
- Bayertz, K. 1996. Human dignity: Philosophical origin and scientific erosion of an idea. In *Sanctity of life and human dignity*, ed. K. Bayertz, 73–90. Dordrecht: Kluwer.
- . 1999. Ethik, Tod und Technik. In *Hirntod und Organverpflanzung*, Hrsg. J.S. Ach and M. Quante, 75–99, Zweite, erweiterte Auflage, Stuttgart: Frommann-holz- boog.
- Bayne, Tim. 2010. *The unity of consciousness*. Oxford: Oxford University Press.
- Beauchamp, T.L. 1977. Paternalism and biobehavioral control. *The Monist* 60: 62–80.

- . 1978. A reply to Rachels on active and passive euthanasia. In *Ethical issues in death and dying*, ed. T.L. Beauchamp and S. Perlin, 246–258. Upper Saddle River: Englewood Cliffs.
- . 1979. On justifications for coercive genetic control. In *Biomedical ethics and the law*, ed. J. Humber and R.F. Almeder, 383–396. New York: Plenum Press.
- . 1995. Paternalism. In *Encyclopedia of bioethics*, vol. 4, revised edition, ed. W.T. Reich, 1914–1920, New York: Simon and Shuster.
- . 1996. Introduction. In *Intending death*, ed. T.L. Beauchamp, 1–22. Upper Saddle River: Prentice Hall.
- . 1999. The failure of theories of personhood. *Kennedy Institute of Ethics Journal* 9: 309–324.
- Beauchamp, Tom. 2009. The concept of paternalism in biomedical ethics. *Jahrbuch für Wissenschaft und Ethik* 9: 77–92.
- Beauchamp, T.L., and J.F. Childress. 1994. *Principles of biomedical ethics*. 4th ed. New York: Oxford University Press.
- Beauchamp, T.L., and A.I. Davidson. 1979. The definition of euthanasia. *The Journal of Medicine and Philosophy* 4: 294–312.
- Becker, L.C. 1981. Human being: The boundaries of the concept. In *Medicine and moral philosophy*, ed. M. Cohen et al., 23–48. Princeton: Princeton University Press.
- Bedate, C.A., and R.C. Cefalo. 1989. The zygote: To be or not to be a person. *The Journal of Medicine and Philosophy* 14: 641–645.
- Benn, S.I. 1976. Freedom, autonomy, and the concept of a person. *Proceedings of the Aristotelian Society* 76: 109–130.
- . 1990. *A theory of freedom*. Cambridge: Cambridge University Press.
- Bermúdez, José L. 2012. Memory judgements and immunity to error through misidentification. *Grazer Philosophische Studien* 48: 123–141.
- Berofsky, B. 1995. *Liberation from self*. Cambridge: Cambridge University Press.
- Billings, J.J. 1989. “When did I begin” by Norman M. Ford. *Anthropotes Rivista di Studi sulla persona e la famiglia* 5: 119–127.
- Birnbacher, D. 1988. *Verantwortung für zukünftige Generationen*. Stuttgart: Reclam.
- . 1994. Einige Gründe, das Hirntodkriterium zu akzeptieren. In *Wann ist der Mensch tot?* Hrsg. J. Hoff and J. In der Schmitt, 28–40, Reinbeck: Rowohlt-Verlag.
- . 1995. *Tun und Unterlassen*. Stuttgart: Reclam.
- . 1996. Ambiguities in the concept of Menschenwürde. In *Sanctity of life and human dignity*, ed. K. Bayertz, 107–121. Dordrecht: Kluwer.
- . 1997. Das Dilemma des Personenbegriffs. In *Personsein aus bioethischer Sicht* (= Archiv für Rechts- und Sozialphilosophie Beiheft 73), (Hrsg.) P. Strasser and E. Starz, 9–25.
- . 1999. Fünf Bedingungen für ein akzeptables Todeskriterium. In *Hirntod und Organverpflanzung*, Hrsg. J.S. Ach and M. Quante, 49–74. Zweite, erweiterte Auflage, Stuttgart: Frommann-holz-boog.
- . 2001. Selbstbewußte Tiere und bewußtseinsfähige Maschinen—Grenzgänge am Rand des Personenbegriffs. In *Person*, Hrsg. D. Sturma, 301–321. Paderborn: Mentis.
- Bodden-Heidrich, R. et al. 1997. Beginn und Entwicklung des Menschen: Biologisch-medizinische Grundlagen und ärztlich-klinische Aspekte. In *Beginn, Personalität und Würde des Menschen*, Hrsg. G. Rager, 15–159. Freiburg: Alber.
- Bok, S. 1998. Part Two. In *Euthanasia and physician-assisted suicide*, ed. G. Dworkin et al., 83–139. Cambridge: Cambridge University Press.
- Bole, T.J. III. 1989. Metaphysical accounts of the zygote as a person and the veto power of facts. *The Journal of Medicine and Philosophy* 14: 647–653.
- Bosanquet, B. 1910. *The philosophical theory of the state*. London: Macmillan.
- Brennan, A. 1988. *Conditions of identity*. Oxford: Oxford University Press.
- Brock, D.W. 1993. *Life and death*. Cambridge: Cambridge University Press.
- . 1994. Good decision making for incompetent patients. *Hastings Center Report* 24 (6): S8–S11.

- . 1998. Commentary on 'the time frame of preferences, dispositions, and the validity of advance directives for the mentally ill'. *Philosophy, Psychiatry, and Psychology* 5: 251–253.
- Brook, D.H.M. 1994. *The unity of the mind*. New York: Palgrave Macmillan.
- Brueckner, A. 1993. Parfit on what matters in survival. *Philosophical Studies* 70: 1–22.
- Buchanan, A.E. 1988. Advance directives and the personal identity problem. *Philosophy and Public Affairs* 17: 277–302.
- Buchanan, A.E., and D.W. Brock. 1990. *Deciding for others*. Cambridge: Cambridge University Press.
- Buckle, S. 1993. Arguing from potential. In *Embryo experimentation*, ed. P. Singer et al., 90–108. Cambridge: Cambridge University Press.
- Butler, J. 1836. *The whole works*. London: Thomas Tegg.
- Callahan, D. 1992. Dementia and appropriate care: Allocating scarce resources. In *Dementia and aging*, ed. R.H. Binstock et al., 141–152. Baltimore: Paul H. Brookes.
- Cantor, N.L. 1992. Prospective autonomy: on the limits of shaping one's postcompetent medical fate. *Journal of Contemporary Health Law and Policy* 13: 13–48.
- Caplan, A.L. 1992. *If I were a rich man could I buy a pancreas?* Bloomington: Indiana University Press.
- Carter, R. 1977. Justifying paternalism. *Canadian Journal of Philosophy* 7: 133–145.
- Carter, W.R. 1982. Do zygotes become people? *Mind* 91: 77–95.
- Cassam, Q. 1997. *Self and world*. Oxford: Clarendon Press.
- Castañeda, H.-N. 1982. *Sprache und Erfahrung*. Frankfurt a.M: P. Lang.
- Childress, J.F. 1997. *Practical reasoning in bioethics*. Bloomington: Indiana University press.
- Chisholm, R.M. 1969. The loose and popular and the strict and philosophical senses of identity. In *Perception and personal identity*, ed. N.S. Care and R.H. Grimm, 82–106. Cleveland: The Press of Case Western Reserve University.
- . 1970a. Identity through time. In *Language, belief, and metaphysics*, ed. H.E. Kiefer and M.K. Munitz, 163–182. Albany: State University of New York Press.
- . 1970b. Events and propositions. *Nous* 4: 15–32.
- . 1971a. Problems of identity. In *Identity and individuation*, ed. M.K. Munitz, 3–30. New York: New York University Press.
- . 1971b. States of affairs again. *Nous* 5: 179–189.
- . 1985. The structure of states of affairs. In *Essays on Davidson's actions and events*, ed. B. Vermazen and M.B. Hintikka, 107–114. Oxford: Clarendon Press.
- . 1986. Self-profile. In *Roderick M. Chisholm*, ed. R.J. Bogdan, 3–77. Dordrecht: Reidel.
- . 1994. On the observability of the self. In *Language, metaphysics, and death*, ed. J. Donnelly, 2nd ed., 195–210. New York: Fordham University Press.
- Chodoff, P., and R. Peele. 1983. The psychiatric will of Dr. Szasz. *Hastings Center Report* 13 (2): 11–13.
- Christman, J. 1991. Autonomy and personal history. *Canadian Journal of Philosophy* 21: 1–24.
- . 1993. Defending historical autonomy: A reply to Professor Mele. *Canadian Journal of Philosophy* 23: 281–290.
- Christman, John. 2009. *The Politics of Persons*. Cambridge: Cambridge University Press.
- Churchill, L.R. 1989. Trust, autonomy, and advance directives. *Journal of Religion and Health* 28: 175–183.
- Clark, Stephen R.L. 2013. Personal identity and identity disorders. In *The Oxford handbook of philosophy and psychiatry*, ed. K.W.M. Fulford et al., 911–928. Oxford: Oxford University Press.
- Corradini, A. 1994. Goldene Regel, Abtreibung und Pflichten gegenüber möglichen Individuen. *Zeitschrift für philosophische Forschung* 48: 21–42.
- Culver, C.M., and B. Gert. 1981. The morality of involuntary hospitalization. In *The law-medicine relation: A philosophical exploration*, ed. S.F. Spicker et al., 159–175. Dordrecht: Reidel.
- Currie, B.S. 1978. The redefinition of death. In *Organism, medicine, and metaphysics*, ed. S.F. Spicker, 177–197. Dordrecht: Springer.
- Davidson, D. 1982. *Actions and Events*. Oxford: Clarendon Press.

- Dawson, K. 1993. Fertilization and moral status: A scientific perspective. In *Embryo experimentation*, ed. P. Singer et al., 43–64. Cambridge: Cambridge University Press.
- DeGrazia, D. 1999. Advance directives, dementia, and ‘the someone else problem’. *Bioethics* 13: 373–391.
- . 2005. *Human identity and bioethics*. Cambridge: Cambridge University Press.
- Den Hartogh, G. 1998. The slippery slope argument. In *A companion to bioethics*, ed. H. Kuhse and P. Singer, 280–290. Oxford: Blackwell Publishers.
- Dennett, D.C. 1981. *Brainstorms*, 267–285. Brighton: Harvester Press.
- . 1990. The interpretation of texts, people and other artifacts. *Philosophy and Phenomenological Research* 50: 177–194.
- . 1992. The self as the center of narrative gravity. In *Self and consciousness: Multiple perspectives*, ed. F. Kessel et al., 103–115. New York: Psychology Press.
- Doepke, F. 1990. The practical importance of personal identity. *Logos* 11: 83–91.
- Dresser, R. 1984. Bound to treatment: The Ulysses contract. *Hastings Center Report* 14 (3): 13–16.
- . 1986. Life, death, and incompetent patients: Conceptual infirmities and hidden values in the law. *Arizona Law Review* 28: 373–405.
- . 1989. Advance directives, self-determination, and personal identity. In *Advance directives in medicine*, ed. C. Hackler et al., 155–170. New York: Praeger.
- . 1994a. Missing persons: Legal perceptions of incompetent patients. *Rutgers Law Review* 46: 609–719.
- . 1994b. Advance directives. Implications for policy. *Hastings Center Report* 24 (6): S2–S5.
- . 1995. Dworkin on dementia. Elegant theory, questionable policy. *Hastings Center Report* 25 (6): 32–38.
- Dresser, R., and P.J. Whitehouse. 1994. The incompetent patient on the slippery slope. *Hastings Center Report* 24 (4): 6–12.
- Dübler, Dominik. 2016. *Selbstbestimmung und das gute Leben im demokratischen Staat*. Münster: Mentis.
- Dubler, N.N. 1995. The doctor-proxy relationship: The neglected connection. *Kennedy Institute of Ethics Journal* 5: 289–306.
- Dupré, John. 1993. *The disorder of things*, Metaphysical foundations of the disunity of science. Cambridge: Harvard University Press.
- Dworkin, R. 1986. Autonomy and the demented self. *Milbank Quarterly* 64 (Suppl. 2): 4–16.
- Dworkin, G. 1988. *The theory and practice of autonomy*. Cambridge: Cambridge University Press.
- Dworkin, R. 1993. *Life’s dominion*. London: Harper Collins.
- Dworkin, G. 1997. Book review – ‘Liberation from self: A theory of personal autonomy’. *Journal of Philosophy* 94: 212–216.
- Elliot, R. 1993. Identity and the ethics of gene therapy. *Bioethics* 7: 27–40.
- . 1997. Genetic therapy, person-regarding reasons and the determination of identity. *Bioethics* 11: 151–160.
- Elster, J. 1981. *Logik und Gesellschaft*. Frankfurt a.M: Suhrkamp.
- , ed. 1986. *The multiple self*. Cambridge: Cambridge University Press.
- . 1987. *Subversion der Rationalität*. Frankfurt a.M: Campus-Verl.
- Engelhardt, H.T. Jr. 1974. The ontology of abortion. *Ethics* 84: 217–234.
- . 1977. Ontology and ontogeny. *The Monist* 60: 16–28.
- . 1986. *The foundations of bioethics*. New York: Oxford University Press.
- English, J. 1975. Abortion and the concept of a person. *Canadian Journal of Philosophy* 5: 233–243.
- Evans, G. 1991. *The varieties of reference*. Oxford: Clarendon Press.
- Evans, M. 1994. Against the definition of brainstem death. In *Death rites*, ed. R. Lee and D. Morgan, 1–10. London: Europa Editions.
- Faden, R.R., and Beauchamp, T.L. (in collaboration with N.M.P. King). 1986. *A history and theory of informed consent*. Oxford: Oxford University Press.
- Fairbairn, G.J. 1995. *Contemplating suicide*. London: Routledge.

- Fateh-Moghadam, Bijan, and Thomas Gutmann. 2014. Governing [through] autonomy. The moral and legal limits of 'soft paternalism'. *Ethical Theory and Moral Practice* 17: 383–397.
- Feinberg, J. 1973. *Social philosophy*. Prentice-Hall: Englewood Cliffs.
- . 1986. *Harm to self*. Oxford: Oxford University Press.
- Feldman, F. 1991. Some puzzles about the evil of death. *Philosophical Review* 100: 205–227.
- . 1992. *Confrontations with the reaper*. Oxford: Oxford University Press.
- Fennell, P. 1996. *Treatment without consent*. London: Routledge.
- Finnis, J. 1996a. A philosophical case against euthanasia. In *Euthanasia examined*, ed. J. Keown, 23–35. Cambridge: Cambridge University Press.
- . 1996b. The fragile case for euthanasia: A reply to John Harris. In *Euthanasia examined*, ed. J. Keown, 46–55. Cambridge: Cambridge University Press.
- Fischer, J.M., ed. 1993. *The metaphysics of death*. Stanford: Stanford University Press.
- . 1995. *The metaphysics of free will*. Oxford: Oxford University Press.
- Fischer, J.M., and M. Ravizza. 1998. *Responsibility and control*. Cambridge: Cambridge University Press.
- Fisher, A. 1991. Individuogenesis and a recent book by Fr. Norman Ford. *Anthropotes. Rivista di studi sulla persona e la famiglia* 7: 199–244.
- Flaman, D. 1991. When did I begin? Another critical response to Norman Ford. *Linacre Quarterly* 58: 39–55.
- Forbes, G. 1985. *The metaphysics of modality*. Oxford: Clarendon Press.
- Ford, N.M. 1991. *When did I begin?* Cambridge: Cambridge University Press.
- Foster, J. 1979. In *Self-defence*. In *Perception and identity*, ed. G.F. MacDonald, 161–185. London: Macmillan.
- Francis, L.P. 1993. Advance directives for voluntary euthanasia: A volatile combination. *The Journal of Medicine and Philosophy* 18: 297–322.
- Frankfurt, H.G. 1988. *The importance of what we care about*. Cambridge: Cambridge University Press.
- . 1999. *Necessity, volition, and love*. Cambridge: Cambridge University Press.
- Freeman, J.M., and E.D. Pellegrino. 1996. Management at the end of life: A dialogue about intending death. In *Intending death*, ed. T.L. Beauchamp, 184–187. Upper Saddle River: Prentice Hall.
- Friedman, M.A. 1986. Autonomy and the split-level self. *The Southern Journal of Philosophy* 24: 19–35.
- Friedrich, Orsolya, and Michael Zichy, eds. 2014. *Persönlichkeit*. Münster: Mentis.
- Fulford, K.W.M. 1989. *Moral theory and medical practice*. Cambridge: Cambridge University Press.
- Garrett, B. 1990. Personal identity and extrinsicness. *Philosophical Studies* 59: 177–194.
- Geach, P.T. 1980. *Reference and generality*. 3rd ed. Ithaca: Cornell University Press.
- Gert, B., et al. 1997. *Bioethics: A return to fundamentals*. New York: Oxford University Press.
- Gervais, K.G. 1986. *Redefining death*. New Haven: Yale University Press.
- Gordijn, B. 2000. Die Debatte um Euthanasie in den Niederlanden und Deutschland. In *Medizinethik und Kultur*, ed. B. Gordijn and H. ten Have, 303–343. Stuttgart: Frommann-Holzboog.
- Gormally, L. 1992. The living will: Some reflections of a recent report. In *The dependent elderly*, ed. L. Gormally, 53–69. Cambridge: Cambridge University Press.
- . 1995. Walton, Davies, Boyd and the legalization of euthanasia. In *Euthanasia examined*, ed. J. Keown, 113–140. Cambridge: Cambridge University Press.
- Graham, G., and G.L. Stephens. 1994. Mine and mine. In *Philosophical psychopathology*, ed. G. Graham and G.L. Stephens, 91–110. Cambridge: MIT Press.
- Green, M.B., and D. Wikler. 1981. Brain death and personal identity. In *Medicine and moral philosophy*, ed. M. Cohen et al., 49–77. Princeton: Princeton University Press.
- Grisez, G., and J.M. Boyle. 1979. *Life and death with liberty and justice*. Notre Dame: University of Notre Dame Press.
- Grobstein, C. 1988. *Science and the unborn*. New York: Dana Press.

- Guckes, B. 1997. *Das Argument der schiefen Ebene*. Stuttgart: G. Fischer.
- Gunnarsson, Logi. 2010. *Philosophy of personal identity and multiple personality*. New York: Routledge.
- Gutmann, Thomas. 2008. Struktur und Funktion der Menschenwürde als Rechtsbegriff. In *Lebenswelt und Wissenschaft*, ed. C.F. Gethmann, 309–330. Hamburg: Meiner.
- . 2014. Paternalismus und Konsequentialismus. In *Paternalismus und Konsequentialismus*, ed. M. Kühler and A. Nossek, 27–66. Münster: Mentis.
- . 2015. Donation after circulatory determination of death: Regelungsoptionen. In *Hirntod und Organtransplantation – zum Stand der Diskussion*, ed. U. Körtner and C. Kopetzki, 179–193. Wien: Verlag Österreich.
- Hacking, I. 1995. *Rewriting the soul. Multiple personality and the science of memory*. Princeton: Princeton University Press.
- Hamilton, A. 1995. A new look at personal identity. *Philosophical Quarterly* 45: 332–349.
- Hanley, R. 1993. On valuing radical transformation. *Pacific Philosophical Quarterly* 74: 209–220.
- Hare, R.M. 1981. *Moral thinking*. Oxford: Oxford University Press.
- . 1988. When does potentiality count? A comment on Lockwood. *Bioethics* 2: 214–226.
- . 1990. Abtreibung und die Goldene Regel. In *Um Leben und Tod*, Hrsg. A. Leist, 132–156. Frankfurt a.M.
- . 1993. A kantian approach to abortion. In *Essays on bioethics*, ed. R.M. Hare, 168–184. Oxford: Clarendon Press.
- . 1995a. Zum moralischen Status potentieller Personen. Replik auf Schöne-Seifert. In *Zum moralischen Denken*. Band 2, Hrsg. C. Fehige and G. Meggle, 356–360. Frankfurt a.M.: Suhrkamp.
- . 1995b. Abtreibung, Empfängnisverhütung und Zeugungspflicht. Replik auf Lenzen. In *Zum moralischen Denken*. Band 2, Hrsg. C. Fehige and G. Meggle, 356–360. Frankfurt a.M.: Suhrkamp.
- . 1998. Preferences of possible people. In *Preferences*, ed. C. Fehige and U. Wessels, 399–405. Berlin: de Gruyter.
- Harris, J. 1983. In vitro fertilization: The ethical issues. *Philosophical Quarterly* 33: 217–237.
- . 1985. *The value of life*. London: Routledge.
- . 1996a. Euthanasia and the value of life. In *Euthanasia examined*, ed. J. Keown, 6–22. Cambridge: Cambridge University Press.
- . 1996b. The philosophical case against the philosophical case against euthanasia. In *Euthanasia examined*, ed. J. Keown, 36–45. Cambridge: Cambridge University Press.
- . 1998. *Clones, genes, and immortality*. Oxford: Oxford University Press.
- . 1999. The concept of the person and the value of life. *Kennedy Institute of Ethics Journal* 9: 293–308.
- Häyry, H. 1991. *The limits of medical paternalism*. London: Routledge.
- Herrmann, M. 1995. *Identität und Moral*. Berlin: De Gruyter.
- Hoff, J., and J. In der Schmitt. Hrsg. 1994. *Wann ist der Mensch tot?* Reinbeck: Rowohlt-Verlag.
- Hoffman, J.C. 1979. Clarifying the debate on death. *Soundings* 62: 430–447.
- Hope, T. 1994. Personal identity and psychiatric illness. In *Philosophy, psychology, and psychiatry*, ed. A.P. Griffiths, 131–144. Cambridge: Cambridge University Press.
- Horan, D.J. 1978. Euthanasia and brain death: Ethical and legal considerations. *Annals of the New York Academy of Sciences* 315: 363–375.
- Hughes, Julian C., et al., eds. 2006. *Dementia*. Oxford: Oxford University Press.
- Hurley, S.L. 1998. *Consciousness in action*. Cambridge: Harvard University Press.
- Izenberg, G.N. 1992. *Impossible individuality*. Princeton: Princeton University Press.
- Jobs, J. 1984. The idea of a personal history. *International Philosophical Quarterly* 24: 179–187.
- Johnston, M. 1987. Human beings. *Journal of Philosophy* 84: 59–83.
- Jonas, H. 1985. *Technik, Medizin und Ethik*. Frankfurt a.M.: Suhrkamp.
- Kahn, J.P. 1991a. Genetic harm: Bitten by the body that keeps you? *Bioethics* 5: 289–308.

- . 1991b. Commentary on Zohar's 'Prospects for' genetic therapy '— Can a person benefit from being altered?'. *Bioethics* 5: 312–317.
- Kamisar, Y. 1995. Physician assisted suicide: The last bridge to active voluntary euthanasia. In *Euthanasia examined*, ed. J. Keown, 225–260. Cambridge: Cambridge University Press.
- Kaplan, D. 1989. Demonstratives. In *Themes from Kaplan*, ed. J. Almog et al., 481–563. New York: Oxford University Press.
- Kenny, A. 1973. Freedom, spontaneity, and indifference. In *Essays on freedom of action*, ed. T. Honderich, 87–104. Boston: Routledge and Kegan Paul.
- Keown, J. 1995a. Some reflections on euthanasia in the Netherlands. In *Euthanasia, clinical practice and the law*, ed. L. Gormally, 193–218. London: The Linacre Centre for Health Care Ethics.
- . 1995b. Further reflections on euthanasia in the Netherlands in the light of the Rummelink Report and the Van Der Maas Survey. In *Euthanasia, clinical practice and the law*, ed. L. Gormally, 219–240. London: The Linacre Centre for Health Care Ethics.
- . 1995c. Euthanasia in the Netherlands: Sliding down the slippery slope? In *Euthanasia examined*, ed. J. Keown, 261–296. Cambridge: Cambridge University Press.
- Kim, J. 1993. *Supervenience and mind*. Cambridge: Cambridge University Press.
- Kleinig, J. 1983. *Paternalism*. Totowa: Rowman & Allanheld.
- Kleinig, John. 2009. Paternalism and personal identity. *Jahrbuch für Wissenschaft und Ethik* 9: 93–106.
- Knell, Sebastian, and Marcel Weber. 2009. *Menschliches Leben*. Berlin: de Gruyter.
- Koch, H.-G. et al., Hrsg. 1996. *Informed consent in psychiatry*. Stuttgart: Verlag Österreich.
- Kolak, D., and R. Martin. 1987. Personal identity and causality: Becoming unglued. *American Philosophical Quarterly* 24: 339–347.
- Korsgaard, C.M. 1996. *Creating the kingdom of ends*. Cambridge: Cambridge University Press.
- Kripke, S.A. 1981. *Name und Notwendigkeit*. Frankfurt a.M.: Suhrkamp.
- Kristjánsson, K. 1996. *Social freedom*. Cambridge: Cambridge University Press.
- Kuczewski, M.G. 1994. Whose will is it, anyway? A discussion of advance directives, personal identity, and consensus in medical ethics. *Bioethics* 8: 27–48.
- Kuhse, H. 1987. *The sanctity-of-life-doctrine in medicine – A critique*. Oxford: Clarendon Press.
- . 1990. The in vitro embryo. Questions of individuality and moral status. *International Journal of Bioethics* 1: 92–99.
- . 1996. Sanctity of life, voluntary euthanasia and the Dutch experience: some implications for public policy. In *Sanctity of life and human dignity*, ed. K. Bayertz, 19–37. Dordrecht: Kluwer.
- . 1999. Some reflections on the problem of advance directives, personhood, and personal identity. *Kennedy Institute of Ethics Journal* 9: 347–364.
- Kuhse, H., and P. Singer. 1993. Individuals, humans and persons: The issue of moral status. In *Embryo Experimentation*, ed. P. Singer et al., 65–75. Cambridge: Cambridge University Press.
- . 1994. Abortion and contraception: The moral significance of fertilization. In *The beginning of human life*, ed. F.K. Beller and R.F. Weir, 145–161. Dordrecht: Springer.
- Kultgen, J. 1995. *Autonomy and intervention*. New York: Oxford University Press.
- Kurthen, M., and D.B. Linke. 1994. Vom Hirntod zum Teilhirntod. In *Wann ist der Mensch tot?* Hrsg. J. Hoff & J. In der Schmitten, 82–94. Reinbeck: Rowohlt-Verlag.
- Kurthen, M., et al. 1989. Teilhirntod und Ethik. *Ethik in der Medizin* 1: 134–142.
- Lamb, D. 1985. *Death, brain death and ethics*. Albany: State University of New York Press.
- Leibniz, G.W. 1958. *Metaphysische Abhandlung*. Hamburg: F. Meiner.
- . 1996. *Neue Abhandlungen über den menschlichen Verstand*. Erster Band (= Philosophische Schriften Band 3.1). Herausgegeben von W. von Engelhardt und H.H. Holz, Frankfurt a.M.
- Leist, A. 1990. *Eine Frage des Lebens*. Frankfurt a.M.: Campus Verlag.
- Lenzen, W. 1995. Hare über Abtreibung, Empfängnisverhütung und Zeugungspflicht. In *Zum moralischen Denken*. Band 2, Hrsg. C. Fehige and G. Meggle. 225–239 Frankfurt a.M.: Suhrkamp.



- Lewis, D. 1976. Survival and identity. In *The identities of persons*, ed. A.O. Rorty, 17–40. Berkeley: University of California Press.
- . 1986. *On the plurality of worlds*. Oxford: Blackwell Publishers.
- Linke, D.B. 1993. *Hirnverpflanzung*. Reinbeck: Rowohlt.
- Lloyd, G. 1993. *Being in time*. London: Routledge.
- Locke, J. 1975. *An essay concerning human understanding*. Oxford: Oxford University Press.
- Lockwood, M. 1985. When does a life begin? In *Moral dilemmas in modern medicine*, ed. M. Lockwood, 9–31. Oxford: Blackwell Publishers.
- . 1988a. Warnock versus Powell (and Harradine): When does potentiality count? *Bioethics* 2: 187–213.
- . 1988b. Hare on potentiality: A rejoinder. *Bioethics* 2: 343–352.
- . 1990. Der Warnock-Bericht: eine philosophische Kritik. In: *Um Leben und Tod*, Hrsg. A. Leist, 235–265. Frankfurt a.M.
- Loux, M.J. 1998. *Metaphysics*. London: Routledge.
- Lowe, E.J. 1989. *Kinds of being*. Oxford: Blackwell Publishers.
- Lund, D.H. 1994. *Perception, mind and personal identity*. Lanham: University Press of America.
- Lynn, J. 1991. Why I don't have a living will. *Law, Medicine and Health Care* 19: 101–104.
- Macklin, R. 1982. Refusals of psychiatric treatment: Autonomy, competence, and paternalism. In *Who decides?* ed. N.K. Bell, 119–131. Clifton: Humana Press.
- . 1983. Treatment refusals: Autonomy, paternalism, and the 'best interest' of the patient. In *Ethical questions in brain and behavior: Problems and opportunities*, ed. D.W. Pfaff, 41–56. New York: Springer.
- . 1984. Personhood in the abortion debate. In *Abortion*, ed. J.L. Garfield and P. Hennessy, 81–102. Amherst: University of Massachusetts Press.
- Macklin, A. 1987. Bound to freedom: The Ulysses contract and the Psychiatric will. *University of Toronto Faculty of Law Review* 45: 37–68.
- Maddy, P. 1979. Is the importance of identity derivative? *Philosophical Studies* 35: 151–170.
- Madell, G. 1981. *The identity of the self*. Edinburgh: The University Press.
- Martin, R. 1987. Memory, connecting, and what matters in survival. *Australasian Journal of Philosophy* 65: 82–97.
- . 1990. Identity and survival: The person we most want to be. In *The experience of philosophy*, ed. D. Kolak and R. Martin, 97–108. Belmont: Wadsworth and the American Studies Association.
- . 1991. Identity, transformation, and what matters in survival. In *Self and identity*, ed. D. Kolak and R. Martin, 289–301. New York: St. Martin's Press.
- . 1993. Having the experience: The next best thing to being there. *Philosophical Studies* 70: 305–321.
- . 1998. *Self-concern*. Cambridge: Cambridge University Press.
- Martin, R.J., and S.J. Post. 1992. Human dignity, dementia, and the moral basis of caregiving. In *Dementia and aging*, ed. R.H. Binstock et al., 55–68. Baltimore: Johns Hopkins University Press.
- McCullagh, P. 1993. *Brain dead, brain absent, brain donors*. Chichester: Wiley.
- McNeill, P.M. 1993. *The ethics and politics of human experimentation*. Cambridge: Cambridge University Press.
- Meran, J.G. 1996. Advance directives and surrogate decision making: Ethical questions, legal response and clinical summary. *Jahrbuch für Recht und Ethik* 4: 381–443.
- Mill, J.S. 1998. *On liberty and other essays*. Oxford: Oxford University Press.
- Miller, B.L. 1981. Autonomy and the refusal of lifesaving treatment. *Hastings Center Report* 11: 22–28.
- Mishler, B.D., and R.N. Brandon. 1998. Individuality, pluralism, and the phylogenetic species concept. In *The philosophy of biology*, ed. D.L. Hull and M. Ruse, 300–318. Oxford: Oxford University Press.

- Moraczewski, A.S. 1983. Human personhood: A study in person-alized biology. In *Abortion and the status of the fetus*, ed. W.B. Bondeson et al., 301–311. Dordrecht: D. Reidel Publishing Company.
- Morowitz, H.J., and J.S. Trefil. 1992. *The facts of life*. Oxford: Oxford University Press.
- Mulkey, M. 1997. *The embryo research debate*. Cambridge: Harvard University Press.
- Müller, A.W. 1997. *Tötung auf Verlangen – Wohltat oder Untat?* Stuttgart: W. Kohlhammer.
- Nagel, T. 1986. *The view from nowhere*. New York: Oxford University Press.
- Nida-Rümelin, M. 1997. Chisholm on personal identity and the attribution of experiences. In *The philosophy of Roderick M. Chisholm*, ed. L.E. Hahn, 565–586. Chicago: Open Court.
- . 2006. *Der Blick von innen*. Frankfurt am Main: Suhrkamp.
- Noonan, H. 1991. *Personal identity*. London: Routledge.
- Nozick, R. 1981. *Philosophical explanations*. Cambridge: Cambridge University Press.
- O’Neill, O. 1984. Paternalism and partial autonomy. *Journal of Medical Ethics* 10: 173–178.
- Olson, E.T. 1997. *The human animal*. New York: Oxford University Press.
- Parfit, D. 1976. Lewis, Perry, and what matters. In *The identities of persons*, ed. A.O. Rorty, 91–107. Berkeley: University of California Press.
- . 1989. *Reasons and persons*. Oxford: Clarendon Press.
- Perry, J. 1972. Can the selve divide? *Journal of Philosophy* 73: 463–488.
- . 1976. The importance of being identical. In *The identities of persons*, ed. A.O. Rorty, 67–90. Berkeley: University of California Press.
- . 1979. The problem of the essential indexical. *Nous* 13: 3–21.
- . 1983. Castañeda on He and I. In *Agent, language, and the structure of the world*, ed. J.E. Tomberlin, 15–42. Atascadero: Ridgeview Publication.
- Persson, I. 1995. Genetic therapy, identity and person regarding reasons. *Bioethics* 9: 16–31.
- . 1997. Genetic therapy, person-regarding reasons and the determination of identity – A reply to Robert Elliott. *Bioethics* 11: 161–169.
- Post, S.P. 1995. Alzheimer disease and the ‘then’ self. *Kennedy Institute of Ethics Journal* 5: 307–321.
- . 1997. *The moral challenge of Alzheimer disease*. Baltimore: Johns Hopkins University Press.
- President’s Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research. 1981. *Defining death*. Washington, DC: U.S. Government Printing Office.
- Putnam, H. 1979. The meaning of ‘meaning’. In *Mind, language and reality* (= Philosophical papers, vol. 2), ed. H. Putnam, 215–271. Cambridge: Cambridge University Press.
- Quante, M. 1995a. Die Identität der Person: Facetten eines Problems. *Philosophische Rundschau* 42: 35–59.
- . 1995b. “Wann ist ein Mensch tot?” Zum Streit um den menschlichen Tod. In: *Zeitschrift für philosophische Forschung* 49: 167–193.
- . 1997a. Ist die diachrone Identität der Person infallibel? In *Feld-Zeit-Kritik*, Hrsg. M. Willaschek, 124–133. Münster: LIT.
- . 1997b. Personal autonomy and the structure of the will. In *Right, morality, ethical life*, ed. J. Kotkavirta, 45–74. Jyväskylä: University of Jyväskylä.
- . 1999a. Personale Identität als Problem der analytischen Metaphysik. In *Personale Identität*, Hrsg. M. Quante, 9–29. Paderborn: Schöningh.
- . 2000a. Manifest versus scientific worldview: Uniting the perspectives. *Epistemologia* (Italien) 23: 211–242.
- . 2000b. Zurück zur verzauberten Natur—ohne konstruktive Philosophie?. *Deutsche Zeitschrift für Philosophie* 48: 953–965.
- . 2000c. “The things we do for love”. Zu den neueren Entwicklungen in H.G. Frankfurts Analyse personaler Autonomie. In *Autonomes Handeln*, Hrsg. M. Betzler and B. Guckes, 117–135. Berlin: Akademie Verl.
- . 2001a. Menschliche Persistenz. In *Person*, Hrsg. D. Sturma, 223–257. Paderborn: Mentis.

- . 2001b. Ethische Aspekte der Xenotransplantation. In *Xenotransplantation*, Hrsg. M. Quante and A. Vieth, 15–66. Paderborn.
- . 2001c. On the limits of construction and individualism in social ontology. In *On the nature of social and institutional reality*, eds. E. Lagerspetz et al., 136–164. Jyväskylä: University of Jyväskylä.
- . 2004a. Which Intrinsicness for weak moral realism? In *Moral realisms*, ed. J. Kotkavirta and M. Quante, vol. 76, 171–187. Helsinki: Acta Filosofica Fennica.
- . 2004b. “The personality of the will” as the principle of abstract right: An analysis of §§34–40 of Hegel’s ‘philosophy of right’ in terms of the logical structure of the concept. In *Hegel on Ethics and Politics*, ed. R.B. Pippin and O. Höffe, 81–100. Cambridge: Cambridge UP.
- . 2007a. The social nature of personal identity, In: *Journal of Consciousness Studies* 14: 56–76.
- . 2007b. Autonomy for real people, In *Intentionality, deliberation and autonomy—The action-Theoretic basis of practical philosophy*, Hrsg. Christoph Lumer and Sandro Nannini, 209–226. Aldershot: Ashgate Publishing Ltd.
- . 2011a. In defence of personal autonomy. *Journal of Medical Ethics* 37: 597–600.
- . 2011b. Identifikation in Relation: Anmerkungen zum evaluativen Selbstverhältnis menschlicher Personen. In *Lebenswelt und Wissenschaft*, Hrsg. C.F. Gethmann, 603–620. Hamburg: Meiner.
- . 2012. *Person*. Berlin: Walter de Gruyter, zweite Auflage.
- . 2013a. Personale Identität und Tiefenhirnstimulation. In *Therapie und Person*, Hrsg. A. Manzeschke and M. Zichy, 27–46. Münster: Mentis.
- . 2013b. Autonomous by default: Assessing ‘non-alienation’ in John Christman’s conception of personal autonomy. *Grazer Philosophische Studien* 87: 177–195.
- . 2013c. Einführung in die allgemeine Ethik. Darmstadt: Wissenschaftliche Buchgesellschaft, fünfte Auflage.
- . 2013d. Being identical by being (treated as) responsible. In *Autonomy and the Self*, ed. M. Kühler and N. Jelinek, 253–271. Dordrecht: Springer.
- . 2014a. Die Perspektiven der Anthropologie. In *Sisäisyys Jasuunistautumisen*, Jyväskylä: SoPhi, Ed. A. Laitinen et al., vol. 125, 169–188.
- . 2014b. *Menschenwürde und personale Autonomie. Demokratische Werte im Kontext der Lebenswissenschaften*. Hamburg: Meiner Verlag, zweite Auflage.
- . 2017a. Zeitlichkeit, personale Lebensform und Ethik. In *Zeit—eine normative Ressource?*, ed. F. Dietrich et al., forthcoming. Frankfurt am Main: Klostermann.
- . 2017b. The logic of essence as internal reflection. In *The Oxford handbook of Hegel*, ed. D. Moyal, forthcoming. Oxford: Oxford University Press.
- Quante, M., and A. Vieth. 2000. Von der Gewissheit des Todes. *Die Psychotherapeutin* 13: 109–124.
- . 2002. In defence of principlism well understood. *Journal of Medicine and Philosophy* 27: 621–649.
- Quine, W.V.O. 1979. Was es gibt. In *Von einem logischen Standpunkt*, ed. W.V.O. Quine, 9–25. Frankfurt a.M: Ullstein.
- . 1992. Events and reification. In *Actions and events*, ed. E. LePore and B. McLaughlin, 162–171. Oxford: Blackwell Publishers.
- Quinn, W. 1984. Abortion: Identity and loss. *The Philosopher’s Annual* 7: 140–170.
- Radden, J. 1994. Second thoughts: Revoking decisions over one’s own future. *Journal of Philosophical and Phenomenological Research* 54: 787–801.
- . 1996. *Divided minds and successive selves: Ethical issues in disorders of identity and personality*. Cambridge: MIT Press.
- Rapp, C. 1995. *Identität, Persistenz und Substantialität*. Freiburg: Karl Alber.
- Regan, D.H. 1983. Paternalism, freedom, identity, and commitment. In *Paternalism*, ed. R. Sartorius, 113–138. Minneapolis: University of Minnesota.

- Reichenbach, B.C. 1987. Euthanasia and the active-passive distinction. *Bioethics* 1: 51–73.
- Reid, T. 1983. *Inquiry and essay*. Indianapolis: Hackett Publishing Company.
- Rheinwald, R. 1994. Causation and intensionality: a problem for naturalism. *European Journal of Philosophy* 2: 41–64.
- Ricoeur, P. 1996. *Das Selbst als ein Anderer*. München: W. Fink.
- Roberts, Melinda A., et al., eds. 2009. *Harming future persons*. Dordrecht: Springer.
- Rogers, J.A., and J.B. Centifanti. 1991. Beyond ‘self-paternalism’: Response to Rosenson and Kasten. *Schizophrenia Bulletin* 17: 9–14.
- Rohs, P. 1988a. Die transzendente Deduktion als Lösung von Invarianzproblemen. In *Kants transzendente Deduktion und die Möglichkeit von Transzendentalphilosophie*, Hrsg. Forum für Philosophie Bad Homburg, 135–192. Frankfurt a.M: Suhrkamp.
- . 1988b. Über Sinn und Sinnlosigkeit von Kants Theorie der Subjektivität. *Neue Hefte für Philosophie* 27/28: 56–80.
- . 1996. *Feld-Zeit-Ich*. Frankfurt a.M: Suhrkamp.
- . 1997. Entgegnungen. In *Feld-Zeit-Kritik*, Hrsg. M. Willaschek, 222–262. Münster: LIT.
- . 1998. *Abhandlungen zur feldtheoretischen Transzendentalphilosophie*. Münster: LIT Verlag.
- Rorty, R. 1989. *Contingency, irony, and solidarity*. Cambridge: Cambridge University Press.
- Rosenberg, J.F. 1983. *Thinking clearly about death*. Englewood Cliffs: Prentice Hall.
- Rosenson, M.K., and A.M. Kasten. 1991. Another view of autonomy: Arranging for consent in advance. *Schizophrenia Bulletin* 17: 1–7.
- Savulescu, J., and D. Dickenson. 1998. The time frame of preferences, dispositions, and the validity of advance directives for the mentally ill. *Philosophy, Psychiatry, and Psychology* 5: 225–246.
- Schechtman, M. 1996. *The constitution of selves*. Ithaca: Cornell University Press.
- Schechtman, Marya. 2014. *Staying Alive*. Oxford: Oxford University Press.
- Scherr, Judith. 2015. *Umgang mit Zwangsmaßnahmen in Krankenhäusern, Psychiatrien und Pflegeeinrichtungen*. Düsseldorf: Deutsche Krankenhaus Verlagsgesellschaft.
- Schöne-Seifert, B. 1995. Zum moralischen Status potentieller Personen. In *Zum moralischen Denken*. Band 2, Hrsg. C. Fehige and G. Meggle, 210–224. Frankfurt a.M: Suhrkamp.
- Schöne-Seifert, Bettina. 2009. Paternalismus. Zu seiner ethischen Rechtfertigung in Medizin und Psychiatrie. *Jahrbuch für Wissenschaft und Ethik* 14: 107–127.
- Sellars, W. 1991. *Science, perception and reality*. Atascadero: Ridgeview.
- Shoemaker, S. 1963. *Self-knowledge and self-identity*. Ithaca: Cornell University Press.
- . 1996. *The first-person perspective and other essays*. Cambridge: Cambridge University Press.
- Siep, L. 1987. Identität der Person und die Beziehung auf den Anderen in der sprachanalytischen Philosophie. In *Das Andere und das Denken der Verschiedenheit*, Hrsg. H. Kimmerle, 249–260. Amsterdam: B.R. Grüner.
- . 1992. *Praktische Philosophie im Deutschen Idealismus*. Frankfurt a.M: Suhrkamp.
- . 1997. *Zwei Formen der Ethik*. Westdeutscher Verlag GmbH: Opladen.
- . 2001. Der Begriff der Person als Grundlage der biomedizinischen Ethik: Zwei Traditionslinien. In: *Person*, Hrsg. D. Sturma, 445–460. Paderborn: Mentis.
- Siep, Ludwig. 2004. *Konkrete Ethik*. Frankfurt am Main: Suhrkamp.
- Siep, L. 2014. *Anerkennung als Prinzip der praktischen Philosophie*. Hamburg: Meiner, zweite Auflage.
- Siep, L., and M. Quante 1999. Ist die aktive Herbeiführung des Todes im Bereich des medizinischen Handelns philosophisch zu rechtfertigen? In *Das medizinisch assistierte Sterben*, Hrsg. A. Holderegger. 37–55. Freiburg, Universitätsverlag.
- Singer, P., and K. Dawson. 1993. IVF technology and the argument from the potential. In *Embryo experimentation*, ed. P. Singer et al., 76–89. Cambridge: Cambridge University Press.
- Snowdon, P.F. 1990. Persons, animals, and ourselves. In *The person and the human mind*, ed. C. Gill, 83–107. Oxford: Clarendon Press.

- . 1991. Personal identity and brain transplants. In *Human beings*, ed. D. Cockburn, 109–126. Cambridge: Cambridge University Press.
- Sosa, E. 1990. Surviving matters. *Nous* 24: 305–330.
- Spittler, J.F. 1999. Krankheitsbedingte Bewußtseinsstörungen. *Fortschritte der Neurologie und Psychiatrie* 67: 37–47.
- Spitz, D. 1996. Collaboration between psychiatrist and patient: how avoidable is paternalism? *Jahrbuch für Recht und Ethik* 4: 233–248.
- Sprigge, T.L.S. 1988. *Personal and impersonal identity*. *Mind* 97: 29–49.
- Steigleder, K. 1994. Die Unterscheidung zwischen dem ‚Tod der Person‘ und dem ‚Tod des Organismus‘ und ihre Relevanz für die Frage nach dem Tod des Menschen. In *Wann ist der Mensch tot?* Hrsg. J. Hoff & J. In der Schmitten, 95–118, Reinbeck: Rowohlt-Verlag.
- Steinbock, B. 1992. *Life before birth*. New York: New American Library.
- Stoecker, R. 1999. *Der Hirntod*. Freiburg: Verlag Karl Alber.
- Stoffell, B. 1998. Voluntary euthanasia, suicide and physician-assisted suicide. In *A companion to bioethics*, ed. H. Kuhse and P. Singer, 272–279. Oxford: Blackwell Publishers.
- Stone, J. 1987. Why potentiality matters. *Canadian Journal of Philosophy* 17: 815–830.
- Strawson, Peter F. 1959. *Individuals*. London: Routledge.
- Strawson, G. 1986. *Freedom and belief*. Oxford: Oxford University Press.
- Sturma, D. 1997. *Philosophie der Person*. Paderborn: Schöningh.
- . Hrsg. 2001. *Person*. Paderborn: Mentis.
- Suarez, A. 1990. Hydatidiform moles and teratomas confirm the human identity of the preimplantation embryo. *The Journal of Medicine and Philosophy* 15: 627–635.
- Sumner, L.W. 1981. *Abortion and moral theory*. Princeton: Princeton University Press.
- Swinburne, R. 1986. *The evolution of the soul*. Oxford: Oxford University Press.
- Szasz, T. 1982. The psychiatric will. A new mechanism for protecting persons against ‘psychosis’ and psychiatry. *American Psychologist* 37: 762–770.
- Taylor, C. 1976. Responsibility for self. In *The identities of persons*, ed. A.O. Rorty, 281–299. Berkeley: University of California Press.
- . 1991. *The ethics of authenticity*. Cambridge: MIT Press.
- . 1992. *Sources of the self*. Cambridge: Cambridge University Press.
- . 1993. *Human agency and language*. Cambridge: Cambridge University Press.
- Thalberg, I. 1983. *Misconceptions of mind and freedom*. Lanham: University Press of America.
- Thomasma, D.C., et al. 1998. *Asking to die*. Dordrecht: Kluwer Academic Publishers.
- Tonti-Filippini, N. 1989. A critical note. *Linacre Quarterly* 56: 36–50.
- Tooley, M. 1983. *Abortion and infanticide*. Oxford: Oxford University Press.
- . 1990. Abtreibung und Kindstötung. In *Um Leben und Tod*, Hrsg. A. Leist, 157–195. Frankfurt a.M.: Suhrkamp.
- . 1998. Personhood. In *A companion to bioethics*, ed. H. Kuhse and P. Singer, 117–126. Oxford: Blackwell Publishers.
- Tugendhat, E. 1979. *Selbstbewußtsein und Selbstbestimmung*. Frankfurt a.M.: Suhrkamp.
- Unger, P. 1979a. I do not exist. In *Perception and identity*, ed. G.F. MacDonald, 235–251. London: Macmillan.
- . 1979b. Why there are no people. *Midwest Studies in Philosophy* 4: 177–222.
- . 1990. *Identity, consciousness, and value*. Oxford: Oxford University Press.
- van Inwagen, P. 1990. *Material beings*. Ithaca: Cornell University Press.
- VanDeVeer, D. 1986. *Paternalistic intervention*. Princeton: Princeton University Press.
- Veatch, R.M. 1976. *Death, dying, and the biological revolution*. New Haven: Yale University Press.
- . 1988. Whole-brain, neocortical, and higher brain related concepts. In *Death: Beyond whole-brain criteria*, ed. R.M. Zaner, 171–186. Dordrecht: Springer.
- Vieth, A., and M. Quante 2005. Chimäre Mensch? In *Die menschliche Natur*, Hrsg. K. Bayertz, 192–218. Paderborn: Mentis.
- . 2010. The structure of perception in particularist ethics. *Ethical Perspectives* 17: 5–39.

- Waddell Ekstrom, L. 1993. A coherence theory of autonomy. *Philosophy and Phenomenological Research* 53: 599–616.
- Walker, N. 1994. Dangerousness and mental disorders. In *Philosophy, psychology, and psychiatry*, ed. A.P. Griffiths, 179–190. Cambridge: Cambridge University Press.
- Walton, D. 1979. *On defining death*. Montreal: McGill-Queen's University Press.
- . 1980. *Brain death*. West Lafayette: Purdue University Press.
- . 1992. *Slippery slope arguments*. Oxford: Oxford University Press.
- Warnock, M. 1983. In vitro fertilization: The ethical issues (II). *Philosophical Quarterly* 132: 238–249.
- . 1990. Haben menschliche Zellen Rechte? In *Um Leben und Tod*, Hrsg. A. Leist, 215–234. Frankfurt a.M.: Suhrkamp.
- . 1994. *Imagination and time*. Oxford: Blackwell Publishers.
- Warren, M.A. 1997. *Moral status*. Oxford: Oxford University Press.
- . 1998. Abortion. In *A companion to bioethics*, ed. H. Kuhse and P. Singer, 127–134. Oxford: Blackwell Publishers.
- Wartofsky, M.W. 1988. Beyond a whole-brain definition of death: Reconsidering the metaphysics of death. In *Death: Beyond whole-brain criteria*, ed. R.M. Zaner, 219–228. Dordrecht: Springer.
- Watson, G. 1987. Free action and free will. *Mind* 96: 145–172.
- . 1989. Free agency. In *The inner citadel*, ed. J. Christman, 109–122. New York: Oxford University Press.
- Wettstein, H.R. 1995. *Leben- und Sterbenkönnen*. Bern: Peter Lang.
- Wiggins, D. 1967. *Identity and spatio-temporal continuity*. Oxford: Blackwell Publishers.
- . 1976. Locke, Butler and the stream of consciousness: And men as a natural kind. In *The identities of persons*, ed. A.O. Rorty, 139–173. Berkeley: University of California Press.
- . 1979. The concern to survive. *Midwest Studies in Philosophy* 4: 417–422.
- . 1980. *Sameness and substance*. Oxford: Blackwell Publishers.
- Wilkes, K.V. 1988. *Real people*. Oxford: Oxford University Press.
- Willems, U. 2016. *Wertkonflikte als Herausforderung der Demokratie*. Wiesbaden: Springer.
- Williams, B. 1990. Who might I have been? In *Human genetic information: Science, law and ethics*, ed. D. Chadwick, 167–179. Chichester: Wiley. (Diskussion des Beitrags eingeschlossen).
- Wils, J.P. 1994. Person und Leib. In *Wann ist der Mensch tot?* Hrsg. J. Hoff and J. In der Schmitt, 119–149. Reinbeck: Rowohlt-Verlag.
- Wilson, J. 1999. *Biological individuality*. Cambridge: Cambridge University Press.
- Wolbert, W. 1998. Das Potentialitätsargument in der Diskussion um den Lebensbeginn. In *Ethische Aspekte der genetischen Diagnostik*, Hrsg. A. Marcol, 41–61. Opole.
- Wolf, U. 1986. Was es heißt, sein Leben zu leben. *Philosophische Rundschau* 33: 242–263.
- Wolf, S. 1990. *Freedom within reason*. New York: Oxford University Press.
- Wollheim, R. 1984. *The thread of life*. Cambridge: Cambridge University Press.
- Young, R. 1980. Autonomy and socialization. *Mind* 86: 565–576.
- . 1986. *Personal autonomy*. London: Routledge.
- Zaner, R.M., ed. 1988. *Death: Beyond whole-brain criteria*. Dordrecht: Springer.
- Zohar, N.J. 1991a. Prospects for 'genetic therapy' – Can a person benefit from being altered? *Bioethics* 5: 275–288.
- . 1991b. Commentary on Khan's 'Genetic harm: Bitten by the body that keeps you?'. *Bioethics* 5: 309–311.
- Zuboff, A. 1978. Moment universals and personal identity. *Proceedings of the Aristotelian Society* 78: 141–155.
- . 1990. One self: The logic of experience. *Inquiry* 33: 39–68.