

BLAMEWORTHY BELIEF

SYNTHESE LIBRARY

STUDIES IN EPISTEMOLOGY,

LOGIC, METHODOLOGY, AND PHILOSOPHY OF SCIENCE

Editor-in-Chief:

VINCENT F. HENDRICKS, *Roskilde University, Roskilde, Denmark*
JOHN SYMONS, *University of Texas at El Paso, U.S.A.*

Honorary Editor:

JAAKKO HINTIKKA, *Boston University, U.S.A.*

Editors:

DIRK VAN DALEN, *University of Utrecht, The Netherlands*
THEO A.F. KUIPERS, *University of Groningen, The Netherlands*
TEDDY SEIDENFELD, *Carnegie Mellon University, U.S.A.*
PATRICK SUPPES, *Stanford University, California, U.S.A.*
JAN WOLEŃSKI, *Jagiellonian University, Kraków, Poland*

VOLUME 338

BLAMEWORTHY BELIEF

A STUDY IN EPISTEMIC DEONTOLOGISM

by

Nikolaj Nottelmann
*University of Southern Denmark,
Denmark*

 Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-1-4020-5960-5 (HB)

ISBN 978-1-4020-5961-2 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

All Rights Reserved

© 2007 Springer

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

For Fred Schmitt –
Great Teacher and Friend

CONTENTS

Preface	xi
1. Introduction	1
1.1. The General Aims of this Study	1
1.2. The Justification of Reactive Attitudes	3
1.3. Epistemically Blameworthy Belief and Morally Blameworthy Action	3
1.4. Epistemic Blameworthiness and Criminal Liability	10
1.5. Epistemic Blameworthiness and the Notion of Epistemic Justification	16
1.6. Alternative Notions of Epistemic Blameworthiness	17
1.7. The Structure of this Study	19
2. Belief and Acceptance	23
2.1. The Doxastic Attitudes of Belief and Acceptance	23
2.2. The Deontic Evaluation of Beliefs and Acceptances	28
3. Approaching a Conception of Epistemic Blameworthiness	33
3.1. Epistemic Blameworthiness as a Fundamental Notion for Epistemic Deontology. Some Terminological Points	33
3.2. The Notion of Responsibility	36
3.3. Some Overriding Distinctions	37
3.3.1. Prospectivism Versus Retrospectivism	37
3.3.2. Substantialism Versus Non-Substantialism	41
3.3.3. Sentimentalism Versus Non-Sentimentalism	42
4. Blameworthy Belief as Inexcusably Undesirable Belief	47
4.1. A First Analysis of the Notion of Epistemic Blameworthiness	47
4.2. The Problem of Epistemic “Accuses”	48
5. Epistemic Undesirability	53
5.1. Epistemic Desiderata	53
5.2. The Importance of the Truth-Goal	55
5.3. Grounds, Reasons and Evidence	57
5.4. Truth-Conduciveness or Reliable Formation	63
5.5. The Notion of Epistemic Undesirability	69

6. Bruce Russell's Basic Analysis of the Notion of Epistemic Blameworthiness	73
6.1. A Critique of Russell's Analysis	73
7. Doxastic Control	79
7.1. The Importance of the Issue of Doxastic Control	79
7.2. Some Basic Requirements for Deontological Adequacy	83
7.3. A Taxonomy of Modes of Doxastic Control	86
7.4. Preview of the Argument Ahead	95
8. Direct Content-Directed Doxastic Control <i>or</i> Doxastic Voluntarism	99
8.1. The Import of Doxastic Voluntarism	99
8.2. The Case Against Doxastic Voluntarism	105
8.2.1. The Conceptual Impossibility Claim	105
8.2.1.1. The nature of the conceptual impossibility claim	105
8.2.1.2. Three bad arguments in favour of the conceptual impossibility claim	105
8.2.1.3. Bernard Williams' "self-defeat" arguments	107
8.2.1.4. Williams' retrospectivist incompatibility claim	110
8.2.1.5. Doxastic voluntarism and self-deception	111
8.2.1.6. Dion Scott-Kakures' prospective argument	115
8.2.1.7. Heinrich Wansing against the conceptional impossibility claim	116
8.2.2. The Psychological Impossibility Claim	118
8.2.2.1. Self-deception revisited	118
8.2.2.2. The psychological impossibility of doxastic action	120
8.3. The Case <i>for</i> Doxastic Voluntarism	123
8.3.1. The Compliant Strategy	123
8.3.1.1. The special analogy argument	123
8.3.1.2. Hobartian voluntarism	128
8.3.2. The Defiant Strategy	131
8.3.2.1. The ostensive argument	131
8.3.2.2. The general analogy argument	136
9. Direct Property-Directed Doxastic Control <i>or</i> Property Voluntarism	143
9.1. The Import of Doxastic Property Voluntarism	143
9.2. John Heil as a Pioneering Property Voluntarist	144
9.3. James Montmarquet's Property Voluntarism	146
9.4. The Limitations of Property Voluntarism	147

10. Indirect Content-Directed Doxastic Control <i>or</i> Doxastic Pascalianism	153
10.1. William Alston's Conception of Epistemic Blameworthiness	153
10.2. Doxastic Pascalianism	157
10.3. A First Glimpse of the Foresight Problem	159
10.4. A Frankfurt-Style Objection	161
10.5. The Limitations of Doxastic Pascalianism	163
10.6. Third-Person Indirect Doxastic Content Control	165
11. Indirect Property-Directed Doxastic Control <i>or</i> Property Pascalianism	169
11.1. Property Pascalianism	169
11.2. John Heil and Hilary Kornblith as Pioneering Property Pascalianists	170
11.3. How to Exercise Indirect Property-Directed Doxastic Control	172
12. Intellectual Obligations	177
12.1. The Importance of Intellectual Obligations	177
12.2. Institutional and Non-Institutional Intellectual Obligations	178
12.3. The Source of Intellectual Obligations	179
13. Foresight and Blameworthy Inadvertence to Risk	189
13.1. Defining Foresight	189
13.2. Blameworthy Inadvertence to Risk	193
13.3. Relevant Foresight and Relevant Inadvertence to Risk	195
13.4. Larry Alexander's Objections	197
13.5. Foresight, Inadvertence and Harm-Doing	200
14. Epistemic Blameworthiness Analysed	203
14.1. A Detailed Analysis of the Notion of Epistemic Blameworthiness	203
14.2. What Counts as an Adequate Excuse for Violating an Intellectual Obligation?	207
14.3. Particular Types of Excuses from Epistemic Blameworthiness	208
14.3.1. Cultural Isolation	208
14.3.2. Cognitive Deficiency	214
14.4. Does the Proposed Analysis of the Notion of Epistemic Blameworthiness have a Vicious Regress Problem?	217
15. Epistemic Autonomy	223
15.1. The Kantian Concept of Practical Autonomy	223
15.2. Approaching a Notion of Epistemic Autonomy	225

15.3. Epistemic Agent-Autonomy and Epistemic Role-Autonomy	227
15.4. Epistemic Agent-Autonomy Examined	228
Notes	233
References	259
Name Index	263
Subject Index	265
Brief Summary	267

PREFACE

The present work has been a long way in the making. After having finished an essay on the epistemology of Austrian sociologist-cum-philosopher-of-science Otto Neurath, in the year 2000 I set out to investigate more fully the broad web of sociological studies of cognitive practices that has continuously infiltrated epistemology for most of the twentieth century.

However, after months of laborious reading it dawned on me that the related academic fields commonly known under the names of the sociology of knowledge and social epistemology, were not well-groomed parks ready for the mapmaker. Rather they appeared to be dense jungles, where the smoke from one campfire was barely visible from the location of its neighbour, the narrow pathways between them being only of use to the detached helicopter or the undiscerning bulldozer. Much preferring a helicopter in such terrain, I set out to gain some measure of conceptual clarity that would perhaps help me map out portions of the impassable morasses. For my own convenience I began my explorations in the province of social epistemology occupied with the notion of testimony. Here most debaters were professional academic philosophers, a fact ensuring some consistency in the use of traditional epistemological terminology.

However, even within this orderly territory I soon ran into dire difficulties. A central debate within the said domain concerns the conditions under which a cognizer is *justified* in holding a belief induced in her by a testimonial source. Debaters invoke decisively different standards of evaluation when discussing such matters: Some stress the reliability of testimonial sources or the rationality of the audience as crucially important to the justification of the ensuing beliefs, while others seemed solely occupied with believers' *rights* to put trust in testimony under various circumstances.

I realized that a fruitful perspective on this debate could only be gained, if one had an adequate conception of the notion of an epistemic right to hold a belief. Here it seemed natural to presume that a believer had such a right, only if, in some epistemic sense, she could not be appropriately *blamed* for holding the belief in question under the relevant circumstances. Further, an adequate conception of the conditions for such epistemic blamelessness could hardly dispense with a notion of *epistemic blameworthiness*.

While browsing the recent literature relevant to an understanding of the latter notion, it became increasingly evident that analysing it was more than enough to occupy an entire book. It did not detract from this point that I soon discovered the affinities between the debates over a so-called deontological conception of epistemic justification (conceiving of justification in terms of rights, permissions,

blamelessness, etc.) and the debates in ethics over the notion of culpable ignorance. From the latter discussion, the link to the disputes over the legal notion of criminal negligence, not least to the resourceful debate about criminal negligence in rape cases, was hard to miss once discovered (I am particularly grateful to Marcia Baron for guiding my research in this direction).

Thus the present work came into being. I seek here to provide a general framework for the study of deontic epistemic normativity from which future endeavours may proceed. In terms of ambition, this work is greatly indebted to predecessors such as James A. Montmarquet's 1993 *Epistemic Virtue and Doxastic Responsibility* and David Owens's 2000 *Reason Without Freedom: The Problem of Epistemic Normativity*. However, I deviate from these works in many respects, not least by explicitly preferring the notion of epistemic blameworthiness as my basic deontic notion. In the run of my argument it shall hopefully emerge that this choice allows me to set to labour almost all recent contributions to the debate over deontic epistemic normativity. It is my sincere hope that the present work will fuel and inspire this debate such that a rewarding discussion of its key issues can take place within a suitably unified framework.

I am deeply grateful to several institutions and individuals for aiding me in the process of writing this study. First and foremost I wish to thank the Danish Research Council for the Humanities for providing me with a three-year research grant in 2000 and the Carlsberg Foundation which sponsored my research from 2005 up until now. I also wish to thank the Department of Philosophy, Education and Rhetoric at the University of Copenhagen for housing me the period when this work came into being. Thirdly, I owe many thanks to the Department of Philosophy at Indiana University, Bloomington for welcoming me as a visiting research scholar in the fall term of 2003.

The present work has benefited immensely from discussions with colleagues and friends. A unique thank goes to professors Frederick F. Schmitt and Marcia Baron, Indiana University, whose level of professional skill, hospitality, and unselfishness will never cease to amaze me.

Sadly, I cannot hope to mention everyone, whose encouragements and critical comments have helped this work into being. However, apart from the above, I am especially indebted to Robert Audi, Finn Collin, Richard Feldman, Jan Riis Flor, Steve Fuller, John Heil, Vincent F. Hendricks, Pamela Hieronymi, Adam Leite, Neil Levy, James Montmarquet, David Owens, Wlodek Rabinowitz, Jay Rosenberg, Benjamin Schnieder, Matthias Steup, Thomas Uebel, and Heinrich Wansing.

Lastly, I wish to express my deepest gratitude towards my family: My true love and lawful wedded wife Ditte and our wonderful children Klement and Jakobe. Without your love and caring support, I would never have made it this far.

Nikolaj Nottelmann,
14-12-2006

CHAPTER 1

INTRODUCTION

“But”, says one, “I am a busy man; I have no time for the long course of study which would be necessary to make me in any degree a competent judge of certain questions, or even able to understand the nature of the arguments.”

Then he should have no time to believe.

W.K. Clifford, 1877¹

Abstract. In this introductory chapter I present some central areas of relevance for the results of this work. I first establish in Section 1.1 that it primarily aims at providing an analysis of the notion of *epistemic blameworthiness*: The epistemic dimension of blameworthiness that may attach to an agent for her holding of a particular belief under particular circumstances. I briefly present the outline of my preferred analysis. I then claim that such an analysis may help shed light on some vexing philosophical questions, not least within moral and legal philosophy: In Section 1.2 I argue that the analysis will be of practical relevance towards establishing a fair standard of evaluation for our reactive attitudes towards each other as believers. In Section 1.3 I argue that it provides a way of understanding an agent’s moral blameworthiness for her harmful actions based on so-called “honest mistakes”; i.e. harmful actions prompted by false beliefs. I then argue in Section 1.4 that the analysis paves the way for a fruitful understanding of the problematic legal notion of criminal negligence, and this in fact whether or not one accepts the influential formal definition of the notion codified by the drafters of the American Model Penal Code. Finally, in Section 1.5 I submit that the results should also satisfy some meta-epistemological curiosity, yielding an analysis of what has been taken by some authors to constitute at least one central aspect of the controversial notion of epistemic justification. In Section 1.6 I briefly present some alternative notions that may more or less appropriately be labelled “epistemic blameworthiness,” notions with which the notion discussed in the present study should not be confused. The chapter concludes in Section 1.7 with a brief overview of the argument ahead.

1.1. THE GENERAL AIMS OF THIS STUDY

We often dislike each other’s beliefs. We judge that the next person’s beliefs are unreasonable, based on inadequate evidence, or simply foolish. Sometimes we are even prone to judge certain beliefs in stronger terms. Consider a person, who, as a result of reading a book by some perverse conspiracy theorist, has come to believe firmly that the queen of England is a mean Martian reptile in disguise. Here, at least some of us will be ready to declare this person genuinely *blameworthy* for believing such a stupid thing as that, even if we are less than certain exactly why this judgement is appropriate.

The present study aims to provide a working analysis of the epistemic dimension² of blameworthiness that may attach to an agent for holding a particular belief under particular circumstances. It will work out in some detail the conditions under which an agent may appropriately be made the object of such blame. It will maintain that actual agents are, have been, and may be epistemically blameworthy in the relevant sense.

As such this work constitutes a defence of a central tenet of *epistemic deontology*, as this position has been ordinarily understood in the recent literature³: An agent's holding of a belief may be an appropriate object of deontic attitudes, an epistemic dimension of blame in particular.

Epistemic deontology is not in itself a very controversial position. As shall emerge it has been defended by a number of authors. Still, the question of how to account for the legitimacy of ascriptions of epistemic blameworthiness has been a matter of untiring dispute. Most authors agree that an agent's blameworthiness for an occurrence of any kind presupposes that she could at least hypothetically have exercised some kind of control over that occurrence. However, the issue of doxastic control, the control that an agent enjoys over doxastic attitudes such as her beliefs, is a highly controversial issue with wildly divergent opinions in the recent literature, not least concerning the relation between an agent's power to exercise such control and the blameworthiness of her beliefs.

I shall aim to settle this debate and work out in some detail the ways in which an agent's doxastic control is in fact connected to her epistemic blameworthiness under various circumstances. This account will then pave the way for a detailed overriding analysis of the notion of epistemic blameworthiness, rich enough to deal satisfactorily with a full repertoire of specific cases. In brief outline, I shall argue that an epistemically blameworthy belief is a belief that is undesirable from an epistemic perspective (see Chapter 5) and for whose epistemic undesirability the believer has no adequate excuse. I shall maintain that a believer lacks an adequate excuse of the relevant kind, if, and only if, the epistemic undesirability of her belief is a result of an exercise of doxastic control constituting an inexcusable violation of an intellectual obligation performed with a relevant type of foresight or inadvertence to risk.

The key notions of this analysis, of course, shall be discussed at length below. In the run of this discussion I shall tie such notions as doxastic control, intellectual obligation and relevant foresight together in a more extensive fashion than has so far been attempted in the published literature. Thus, I hope to establish a firm conceptual platform for epistemic deontology, from which future explorations in deontic epistemic normativity may safely take off.

Firstly, however, in the present introductory chapter I shall point to three reasons why a notion of epistemic blameworthiness is virtually indispensable in our stock of normative concepts:

1. It provides a way of justifying many of our ingrained reactive attitudes toward one another as believers.
2. It provides a way of making sense of certain intuitively compelling, but theoretically puzzling, cases of moral blameworthiness.
3. It provides a way of making sense of the juridical concept of *criminal negligence* as justly establishing the criminal liability of an offender for certain types of criminal offences.

For these reasons, an accurate grasp of the notion is a matter of much more than theoretical curiosity: It may substantially inform our social interactions with other

agents, not least the circumstances under which we are ready to submit them to punishments ranging from mild resentment to harsh terms of incarceration, insofar as we want such punishments to be fair and just.

1.2. THE JUSTIFICATION OF REACTIVE ATTITUDES

Below I shall argue that the notion of epistemic blameworthiness intimately relates to issues of moral blameworthiness and criminal liability. However, it has a certain importance, even when leaving these issues out of the question. Consider the following example:

“The educated racist.” An intelligent college student believes on the basis of testimony presented by some obscure demagogue that people with a dark skin colour are generally inferior to people with lighter skin.⁴

I take it that many of us, if confronted by this college student, would feel at least a mild resentment towards her for holding the racist belief under these circumstances; a negative attitude markedly different in character from, e.g. the dissatisfaction we might feel about her eye colour or other aesthetic features. In fact, this resentment seems much closer in kind to the reactive attitudes we typically take against actions, which we consider morally blameworthy. Further, I take it, this resentment may be present even if we do not believe that the educated racist will in fact perform any harmful actions prompted by her racist belief.

Now certainly, if holding a belief were not something an agent could be appropriately blamed for in any sense, reactive attitudes such as resentment would be out of place. For example resenting a person for her natural skin colour seems to constitute a severe cruelty towards that person. Plausibly, this is because our natural skin colour is not something we can be blamed for in any sense.

In contrast, it seems highly plausible that an agent’s epistemic blameworthiness may sufficiently justify attitudes such as resentment or indignation taken towards her, at least mild degrees of such attitudes. In short then, if we are in fact prone to feel resentment or even indignation towards one another for holding certain beliefs, an accurate grasp of the notion of epistemic blameworthiness could spell out conditions under which such attitudes are justified. This constitutes perhaps the most obvious area of relevance for the conclusions presented in this work.

1.3. EPISTEMICALLY BLAMEWORTHY BELIEF AND MORALLY BLAMEWORTHY ACTION

In 1975 the rape case *Director of Public Prosecutions v. Morgan*⁵ made major headlines in Great Britain and Australia. According to the testimony presented in court, roughly the following incidents gave rise to the trial:

Mr. Morgan, a senior officer in the Royal Air Force, was out drinking with three male junior colleagues. In the run of the evening, he invited the three men to come home with him and have intercourse with his wife. According to the colleagues, Morgan informed them that his wife was “kinky” and would appreciate having intercourse

with them, however she would probably feign resistance and dissent as part of the sexual game. He claimed that he had brought home colleagues for similar purposes in the past, and that his wife had enjoyed it. Upon arriving at the Morgan household the four men dragged Mrs. Morgan from a room in which she was sleeping to another room and placed her on a double bed. The four men took turns in having intercourse with her, while the others forcibly restrained her. Mrs. Morgan resisted persistently and screamed to her children to call the police. After the men had left, Mrs. Morgan drove to a hospital and reported a rape. Medical evidence supported her report.

What happened that night in the Morgan household constituted the *actus reus* of a rape by any standard⁶: The four men submitted Mrs. Morgan to unwanted, non-consensual sex by the use of massive physical force. However, after some initial confusion, the three colleagues all pleaded not guilty of rape on the defence that, at the time of intercourse, they all *believed* that Mrs. Morgan consented to the grisly proceedings. In holding the belief that she consented on the basis of Mr. Morgan's explanations, so they claimed, they had simply made an *honest mistake* and could not be guilty of rape.

Mr. Morgan could not be convicted of rape, as at that time, according to British legislation, by definition a man could not rape his wife (he was later convicted of assisting a rape, though). However, all three colleagues were convicted of rape. If this had been the end of the matter, the public's interest in the case would probably have quickly faded. However, the case went to The Court of Appeal, which confirmed the convictions. Still, it had some doubts about the soundness of the instructions given to the jury deciding the original case. The judge had instructed the jury that the "honest mistake" defence would not acquit the three men unless their belief that Mrs. Morgan consented was both "honestly held" and "a reasonable belief; such a belief as a reasonable man would entertain if he applied his mind and thought about the matter."⁷ The Court of Appeal therefore asked the highest instance in the British juridical system, The House of Lords, to consider the question: Can you properly convict someone of rape if he honestly but *unreasonably* believed, at the time of the alleged rape, that the woman was consenting?

By three votes against two the five Law Lords answered this question in the negative and thereby installed the so-called "Morgan rule"⁸: An offender is not guilty of rape if he honestly believed at the time of intercourse that his victim consented, irrespective of the reasonableness of this belief. Still, the Law Lords dismissed the appeal on the grounds that the jury, even if properly instructed, would not have accepted the defence's claim that the three men believed that Mrs. Morgan consented. However, only a week after this ruling an alleged rapist in a very similar case, *Regina v. Cogan*,⁹ was acquitted because he was found to have honestly but unreasonably believed that his victim, although sobbing, consented to have intercourse with him.

The ruling of the Law Lords, together with its imminent consequences, caused what may best be described as a massive public outrage in Great Britain as well as in Australia,¹⁰ where at that time the rulings of the Law Lords were still held in considerable respect. In the press, the Morgan Rule was widely referred to as a "rapists' charter."¹¹ The outrage is certainly understandable. The Morgan rule raises

pressing concerns of a moral as well as a legal nature. As the relation between these concerns is somewhat complicated, I shall deal with the strictly moral issue first.

Personally I consider it very unlikely that the Morgan rapists actually believed throughout the intercourses that Mrs. Morgan consented to have sex with them. However, let us assume for the present purposes that the following was the case in *Morgan*:

1. Each of the three men physically forced Mrs. Morgan to unwanted, non-consensual sex. Thus, uncontroversially, the *actus reus* of a rape was instantiated in each intercourse.
2. However, each of the three men, during the activities mentioned in (1), believed that Mrs. Morgan consented¹² to have intercourse with him under the conditions present (i.e. including physical constraint, etc.).
3. Each of the three men performed the activity mentioned in (1) prompted by his belief mentioned in (2).¹³
4. Each of the three men based the belief mentioned in (2) entirely on Mr. Morgan's erratic testimony.

I take it that most of us, on considering this case, immediately feel a certain amount of resentment towards the three men, markedly different from the sadness we would feel if the harm inflicted upon Mrs. Morgan had been merely an unfortunate accident. I take this to support the claim that each of the three men is *morally blameworthy* for inflicting the harm mentioned in (1) on Mrs. Morgan, unless he can provide some very strong and exotic defence. However pointing to the *locus* of their moral blameworthiness for the harm caused is no easy matter. Notice first that their blameworthiness can hardly rest on the simple fact that they were moved to have sex with Mrs. Morgan partly because of their mistaken beliefs. Given that they really held on to those beliefs during the entire tragic proceedings they could hardly have been morally required to choose alternative actions under the circumstances.

Holly Smith, recognizing that acts based on an "honest mistake" like the above, are "unwittingly wrongful,"¹⁴ has opted for the solution that such actions may instead be deemed morally blameworthy due to the efficacy of a "reprehensible configuration of desires and aversions." She supports this claim by the following analysis:

S is to blame for performing act A if, and only if,

1. Act A is objectively wrong,
2. S had a reprehensible configuration of desires and aversions, and
3. This configuration gave rise to the performance of A.¹⁵

However, this analysis is far from convincing, particularly when applied to cases like the above. In the present case the men need have had nothing but the desire to have (violent) sex with a *consenting* partner, hardly a desire "reprehensible" in itself.¹⁶ Many men and women have had this desire without ever committing blameworthy actions like the horrible one presently discussed.

Another option would be to locate the blameworthiness of the men's action in their morally bad characters. Surely, one might think, men capable of such mistakes

must be somehow “tainted.” However, it is not clear that the men need have been very “tainted” in the case as presented above. They were just incredibly credulous and probably not too sharp-witted either. However, being stupid and gullible *per se* hardly makes one blameworthy for committing a rape. As pointed out by Holly Smith judgments of blameworthiness are at bottom very different from judgments of “moral” character.¹⁷ First it seems possibly to judge a person’s character without reference to particular actions, second:

Nor are blaming judgments simply assertions that the person has a bad character which has manifested itself in action. For we blame people for their acts even though those acts do not manifest any traits stable or long-lasting enough to be ascribed to “character.”¹⁸

We might well add here that even the most saintly of characters may occasionally perform a blameworthy action, and, more to the present point, plausibly even the most “tainted” of characters may occasionally perform actions she cannot appropriately be blamed for. “Tainted” character does not make for blameworthiness in itself.

Thus, leaving out their choice of action, their desires, and their characters as proper loci of moral blameworthiness, in lack of alternative options the source of the men’s moral blameworthiness must be tracked to their mistaken *belief*, which prompted the harm-doing.

The conclusion is thus that if the men are morally blameworthy for their actions, this must rest on the fact that, under the circumstances, they were *blameworthy* for holding their mistaken belief. Montmarquet presents a similar insight thus:

A seventeenth century Spanish priest orders Mayan babies to be taken from their mothers, baptized then killed – so as to ensure the salvation of their souls. Now such a man’s desires (considered as such) are not the root of the problem here. Rather, if we find this act highly culpable, ultimately it must be in virtue of the culpability of this man’s beliefs.¹⁹

In order to vindicate the strong intuition that the *Morgan* rapists and Montmarquet’s Spanish priest are, at least given suitable further specifications of the cases, morally blameworthy for the horrible harm they inflicted upon others, there is no way around offering an account of how such agents may be blameworthy for holding the mistaken beliefs that led them to inflict this harm.

At this stage, it is vital to notice that the following claim about epistemic blameworthiness due to E.M. Curley flies in the face of the above function of the notion:

What is culpable is not the belief, but acting on a belief which has not been fairly examined, in circumstances, where the possible consequences of acting on the belief – and in particular, the possible consequences for other people are so momentous.²⁰

Curley here tries to eschew the notion of epistemic blameworthiness proper in favour of a species of blameworthiness for actions. However, as seen above, in cases of action based on “honest mistakes,” the source of blameworthiness simply cannot be traced to the harmful action, while the prompting belief is acquitted. Even if I have not yet shown that the notion of epistemic blameworthiness has a relevance outside the framework of blameworthy action, surely within this framework, the

blameworthiness of a belief cannot be reduced to the blameworthiness of the action based on that belief: Rather, in order to make sense of the blameworthiness of the harmful action prompted by some belief, we need a substantial account of epistemic blameworthiness strong enough to *explain*, why the agent is morally blameworthy for inflicting the harm under the given circumstances. I shall aim to provide such an account in what follows.

In fact already in 1877 W.K. Clifford in his famous essay “The Ethics of Belief” pointed out the problems inherent in Curley’s dismissal of a notion of (epistemic) blameworthiness applying to beliefs in isolation. Clifford first presents his celebrated “ship-owner” example, in effect an “honest mistake” case very similar to the ones discussed above: A ship-owner owns a ship that is far from seaworthy. However, his slight and reasonable doubts about the ship’s sea-worthiness he manages to put aside by a self-deceptive routine. Ensuingly he sends the ship to sea with a load of passengers, sincerely believing at the time of embarkation that his ship is sound. Nevertheless, it sinks and everybody aboard drowns. “Surely,” Clifford submits, “he [the ship-owner] was verily guilty of the death of these men.”²¹

Now, Clifford anticipates: “It may be said, however, that in both²² these supposed cases it is not the belief which is judged to be wrong, but the action following upon it.”²³ Against this suggestion, which is effectively Curley’s, he submits:

[I]t is not possible so to sever the belief from the action it suggests as to condemn the one without condemning the other. No man holding a strong belief on one side of a question, or even wishing to hold a belief on one side, can investigate it with such fairness and completeness as if he were really in doubt and unbiased; so that the existence of a belief not founded on fair inquiry unfits a man for the performance of this necessary duty. Nor is it truly a belief at all which has not some influence upon the actions of him who holds it. He who truly believes that which prompts him to an action has looked upon the action to lust after it, he has committed it already in his heart. If a belief is not realized immediately in open deeds, it is stored up for the guidance of the future.²⁴

This passage, held in Clifford’s characteristic grand style, raises a host of questions, not least concerning the “duty of fair inquiry” that Clifford stresses here. This is an issue I cannot take up until much later (Chapter 12). However, the second part of the above passage (from “nor is it truly...”) is immediately relevant to the present proceedings. Here Clifford seems to offer the following point: Once an agent holds a belief, on the basis of which it seems to him reasonable to perform some action, he has no morally significant choice about whether to perform that action or not. Thus the only way to account for the blameworthiness “condemnability” of such an action is to hold the prompting belief *in itself* blameworthy (condemnable). The blameworthiness of the action cannot be accounted for if severed from the blameworthiness of the belief that prompted it.

The question now arises which dimension of blameworthiness Clifford invokes in condemning the ship-owner’s mistaken belief. Could it be that he attempts to push a narrowly *moral* dimension of culpability into the doxastic realm? I believe not, for reasons I shall now make clear.

As noted by Susan Haack, Clifford does not in his famous essay distinguish between epistemic and moral evaluations of beliefs.²⁵ Still, Haack interprets Clifford

as offering his “ship-owner example” as an argument for the so-called “correlation thesis_J”: The thesis that a belief’s being “epistemically justified” is invariably correlated with its being “ethically justified.”²⁶ Accordingly, and despite her acknowledgment of Clifford’s persistent resistance to distinguish between moral and epistemic evaluation, she takes Clifford as offering his “ship-owner example” as a case of “morally culpable ignorance.”²⁷ She then, in my view plausibly, pushes the objection to this supposed claim that if the ship-owner’s belief seen in isolation were to be “morally culpable,” it should also be so under different circumstances, where it did not bring about harmful actions:

But the case [Clifford’s “ship-owner example”] has a number of features, that are not to be found whenever someone believes unjustifiedly, and some of which are essential to the unfavourable moral appraisal appropriate here. The unjustified belief is false; the proposition concerned is of great practical importance; the person concerned is in a position of special responsibility; the false belief leads to dramatically harmful consequences; and the belief is wilfully self-induced. The correlation thesis_J is false unless the ignorance would still be morally culpable even if all these features were absent.

I take it that Haack’s considerations here effectively defeat “the correlation thesis_J.” In fact less could do it: It is far from clear, how beliefs, considered *per se*, can be morally blameworthy. Perhaps then, a Curley-style position is really a misguided attempt to appropriate the observation that it is hard to see how holdings of beliefs can be, e.g. *evil, malicious, harmful* or otherwise undesirable from a moral perspective, in anything but the derivative sense that they may give rise to actions evil, malicious, harmful or otherwise morally undesirable.

However, to attribute to Clifford anything like a Haack-style “correlation thesis_J” is certainly highly implausible, as well as very uncharitable. First, it is not clear what Clifford would count as the “justification” of a belief, and whether he needs such a notion at all.²⁸ After all the headline of the relevant section in “The Ethics of Belief” is “The duty of inquiry,” not “The duty of having justified beliefs.” Second, and most importantly, I will submit against Haack that the most plausible explanation why Clifford did not distinguish between moral and epistemic evaluations in Clifford (1999) was neither confusion, nor an oblique endorsement of some kind of “correlation thesis.” Rather the text suggests that he simply took epistemic and moral evaluation to be equally basic dimensions of normative evaluation with no need to “vindicate” the importance of epistemic evaluations by “correlating” them with moral evaluations. The independent significance that Clifford puts on epistemic evaluations is brought out several times in his essay, perhaps most prominently in the following lofty passage:

And no one man’s belief is in any case a private matter which concerns himself alone. Our lives are guided by that general conception of the course of things which has been created by society for social purposes. Our words, our phrases, our forms and processes and modes of thought, are common property, fashioned and perfected from age to age; an heirloom which every succeeding generation inherits as a precious deposit and a sacred trust to be handed on to the next one, not unchanged but enlarged and purified, with some clear marks of its proper handiwork. Into this, for good or ill, is woven every belief of every man who has speech of his fellows. An awful privilege, and an awful responsibility, that we should help to create the world in which posterity will live.²⁹

In short: We should not pursue the True in order to pursue the moral Good, but because pursuing the True³⁰ is valuable in itself. By not serving this goal appropriately, according to Clifford, we are simply offending “the sacred tradition of humanity!”³¹ I shall not here venture further into Clifford’s historic-cum-idealistic conception of epistemic value. However, I believe there is certainly something to be said for his setting moral and epistemic evaluations on an equal footing.

Most importantly I believe it a gross mistake to take the observation that holdings of beliefs cannot *per se* be *morally* blameworthy to rule out that holdings of beliefs may in themselves be blameworthy in a sense strong enough to underwrite moral blameworthiness for harm-doing brought about by such beliefs.

As we saw above attempts to track the moral blameworthiness for harm-doing prompted by honestly mistaken beliefs to further instances of moral blameworthiness, simply did not work. On the contrary, it is commonly agreed, e.g. that no matter which evil or harmful actions had resulted from the *Morgan* rapists’ belief in Mrs. Morgan’s consent, the men would *not* have been morally blameworthy for these actions or their consequences, if only their belief that Mrs. Morgan consented had been *reasonable*.³² Thus, on this view the moral blameworthiness of the *Morgan* rapists is linked to the *unreasonableness* of their belief that Mrs. Morgan consented. Now, it seems clear that unreasonableness as the notion is employed here is an *epistemic* notion: It is because the rapists had inadequate reasons to believe it *true* that Mrs. Morgan consented, that their blameworthiness may be tracked to their fateful belief, not because, e.g. that they had reasons to believe this proposition morally bad (see also Section 5.3).

Now, even if *necessary* in order to establish the blameworthiness of the *Morgan* rapists, it is still not *sufficient* to point out that their belief that Mrs. Morgan consented to intercourse was not “reasonable” in, e.g. the sense invoked by the judge in the *Morgan* case: Supposedly, the men might be *excused* for not living up to the proposed standards of reasonability under the circumstances. Excuses that might be invoked could, e.g. appeal to the men’s very dim wits that prevented them from being “reasonable” men or made Mr. Morgan’s malicious testimony irresistibly credible by their lights. Whether or not such excuses are appropriate in the case, the central point to notice is that, unless no such excuse is appropriate, the men can hardly be blameworthy for the harm brought about by their actions. Given an appropriate excuse, the harm they inflicted on Mrs. Morgan would be simply due to *force majeure*, very sad and unfortunate, but not something anyone could appropriately be blamed for (here I leave Mr. Morgan’s obvious blameworthy role in the actual proceedings out of the picture).

Clearly, the above argument holds no matter which undesirable properties (unreasonableness in diverse senses, irrationality, formation by an falsity-conducive belief-forming mechanism, etc.) the belief is stipulated to have. If the belief’s having such a property is to make it the case that the men were blameworthy for the harm inflicted by their actions, it cannot be the case that the men were appropriately excused for the fact that their belief had this property.

Thus the source of the rapists’ blameworthiness for their harm-doing is neither the fact that the harmful action was prompted by the belief that Mrs. Morgan consented,

nor the fact that that belief was false “mistaken.” It is rather the fact that the relevant belief had epistemically undesirable properties (such as unreasonableness) together with the fact that, under the circumstances, the men had no excuse for holding such an epistemically undesirable belief.³³ In other words it has been established that a satisfactory understanding of the blameworthiness of harmful action prompted by “honestly mistaken” beliefs cannot do without a sufficient grasp of the notion of *epistemic* blameworthiness.

This reveals a much tighter relation between practical and epistemic values than has generally been recognized, at least on the deontic level: An agent can be blameworthy for her harm-doing simply in virtue of being epistemically blameworthy for holding certain beliefs on which she acted. In this sense practical and epistemic “duties” are on a par, violating the one kind may sometimes bring about the violation of the other: In lofty terms: In order to pursue the Good as I should, I must also sometimes pursue the True as I should. This conclusion is forced upon us by the compelling moral blameworthiness inherent in “honest mistake” cases like Clifford’s “ship-owner example”³⁴ and the *Morgan* case. Perhaps then we could simply let go of the specifications of diverse types of blameworthiness and speak of blameworthiness or culpability *simpliciter*, as did Clifford. To keep it absolutely clear that the basis of epistemic blameworthiness is epistemic, not moral, dimensions of undesirability, I shall however stick to common usage here.

1.4. EPISTEMIC BLAMEWORTHINESS AND CRIMINAL LIABILITY

Except in rare circumstances, some modern jurisdictions require that, in order for a defendant to be judged guilty of a criminal offence, the prosecution must not only prove that the defendant brought about an *actus reus*, i.e. a “social harm”³⁵ that it was her legal duty not to bring about; it must also prove that the defendant did so with a suitable *mens rea*, literally a “guilty mind.” The criminal offences not requiring a *mens rea* proof for convictions are then specifically indicated in the legislature as being under “strict liability.”³⁶

Arguably, requirements of *mens rea* proofs for convictions were introduced in jurisdictions to ensure that a convict is at least morally blameworthy for the harm-doing for which she is punished, the “broad meaning” of *mens rea* being simply that of “a morally blameworthy state of mind.”³⁷ The introduction of *mens rea* requirements meant a departure from “coarser”³⁸ legal practices, caring little about moral desert, but punishing only to get rid of harm-doers, deter prospective harm-doers or revenge harm done. However, in courtroom practice, the notion of *mens rea* proved a “chameleon-like”³⁹ concept fraught with severe ambiguity. In order to install a consistent legal practice, documents such as the American Law Institute’s Model Penal Code (MPC) were drafted, offering a restricted taxonomy of *mens rea* types with technical and seemingly rigid definitions of each type.

I shall argue that the Model Penal Code, taken here as a standard taxonomy, simply fails to provide a compelling account of the *mens rea* involved in “honest mistake of fact” cases like the *Morgan* rape case from Section 1.3: In such cases even the

“broad meaning” given to *mens rea* above cannot be stretched to cover what is really the source of moral blameworthiness and arguably also of criminal guilt: An epistemically blameworthy belief!

The Model Penal Code operates with four types of *mens rea* presented in order of declining severity: purpose, knowledge, recklessness and negligence. Each are defined with respect to “elements of an offence.” In *Morgan* each of the men’s non-consensual intercourse with Mrs. Morgan constituted such an element. As, in the type of jurisdiction relevant here, defendants are convicted and sentenced individually, I shall now limit myself to considering one of the *Morgan* offenders, call him “Smith,” although of course the points are valid for each of them. Let us presume that the events actually took place as presented by the defence in *Morgan*, i.e. that Smith made an “honest mistake” about Mrs. Morgan’s consent at the time he forced her to non-consensual sex. The crucial question now becomes: Which, if any, was Smith’s *mens rea* with respect to his non-consensual intercourse with Mrs. Morgan? I shall present reasons to doubt that the Model Penal Code commands the resources to adequately answer this question.

Firstly, it is uncontroversial that Smith did not act *purposely* or *knowingly* with aspect to the element in question, the non-consensual intercourse. To have acted *purposely*, according to the MPC, with regard to the non-consensual sex Smith should have had a “conscious object to engage in conduct of that nature or to cause such a result.”⁴⁰ Clearly this was not the case. At no point did Smith believe that he was doing any unlawful harm, he believed rather that the bizarre intercourse was something Mrs. Morgan enjoyed and consented to. Thus it cannot have been his conscious object to engage in a non-consensual intercourse.

Secondly, to have acted *knowingly*, according to the MPC, with regard to the non-consensual sex at least Smith should “have been *aware* that his conduct is of that nature or that such circumstances exist.”⁴¹ However, Smith was not so aware. Supposedly, he did not believe or even suspect that Mrs. Morgan did not consent to have sex with him.

This leaves two possibilities, if Smith is to be culpable according to MPC’s “minimal requirements of culpability”⁴²: In submitting Mrs. Morgan to the violent intercourse, he must have acted either *recklessly* or *negligently*. MPC defines “recklessness” thus:

A person acts recklessly with respect to a material element of an offence when he consciously disregards a substantial and unjustifiably risk that the material element exists or will result from his conduct. The risk must be of such a nature and degree that, considering the nature and purpose of the actor’s conduct and the circumstances known to him, its disregard involves a gross deviation from the standard of conduct that a law-abiding person would observe in the actor’s situation.⁴³

An example of reckless action on this definition would be a hell-bent youngster’s driving in the wrong direction on a highway, just for the thrill of it. If the youngster causes a traffic accident, surely he acted recklessly with respect to this accident; he consciously disregarded a substantial and clearly unjustifiable risk that a traffic accident would result from his thrill ride. Now, Smith’s mental state at the time of the non-consensual intercourse does not seem easily comparable to that of this reckless

driver. Still, E.M. Curley, in an early commentary on *Morgan*, has explicitly argued that the *Morgan* rapist *did* act recklessly on MPC's definition,⁴⁴ even if he made an "honest mistake":

But given this account of recklessness [MPC], it is difficult to see why it should be thought that in rape any belief in consent, no matter how unreasonable, excludes recklessness as to consent.⁴⁵

This is supposedly because:

Yet even if we assume that Morgan's three colleagues did believe Mrs. Morgan was consenting, it is difficult to resist concluding that they could not have cared much whether that belief was correct, if they persisted in it in the face of her behaviour.⁴⁶

There are two serious problems with this account. Firstly, we may doubt whether Smith did really consciously "advertently"⁴⁷ disregard a substantial risk at all: He was convinced that Mr. Morgan spoke the truth about the sexual signals of his wife and during the rape, perhaps he simply did not even consider the possibility that Mrs. Morgan did not consent. Secondly, even if Curley is right that, by trusting Mr. Morgan, Smith consciously disregarded a risk of *getting the facts of Mrs. Morgan's consent wrong*, this conscious disregard of risk did *not* concern any material element of the criminal offence of rape: Having mistaken beliefs about the sexual consent of another person, I trust, does not count as a criminal offence or an element thereof in any known legislation. Smith stood accused on charges of rape, not on charges of having a mistaken belief. Thus Curley simply fails to establish on MPC that Smith acted recklessly with regard to any element of the rape of Mrs. Morgan. Smith did not consciously disregard the risk of raping Mrs. Morgan, as given his mistaken belief, he was plausibly not aware of such a risk at all.

We are stuck then with trying to make sense of the proposal that Smith acted *negligently* when raping Mrs. Morgan. MPC defines "negligence" thus:

A person acts negligently with respect to a material element of an offence when he should be aware of a substantial and justifiable risk that the material element exists or will result from his conduct. The risk must be of such a nature and degree that the actor's failure to perceive it, considering the nature and purpose of his conduct and the circumstances known to him, involves a gross deviation from the standard of care that a reasonable person would observe in the actor's situation.⁴⁸

There are two main problems with this definition. Firstly, a *mens rea*, with its literal meaning of a "guilty mind," intuitively must refer to states of the offender's mind, which somehow makes for her criminal liability. However, on MPC "negligence" refers, not to a state of mind that the offender is in, but rather to a state of mind that the offender is *not* in: her *failure* to be perceptive of a certain risk. In a sense MPC's definition takes the "mens" out of "mens rea!" It allows that the "culpable" offence was one that the actor was not even aware that she committed.

The second main problem concerns the nature of the "reasonable person" referred to in MPC's definition of negligence. In "honest mistake" cases, the "reasonable person" surely cannot be the agent herself. If the agent had been "the reasonable person" she would have been aware of the risk that she failed to be aware of. However, neither is "the reasonable person" any particular other agent; supposedly no actual agent is

always reasonable in every situation. Rather than, “the reasonable person” is a fictive character “in the same situation as the agent,” including, among other things the “circumstances known to her.”

However, it is not hard to see that if “knowledge of circumstances” is here taken to mean simply that the “reasonable person” has the same beliefs as the offender, it immediately becomes clear that, e.g. the rapist Smith was not negligent at all, as he certainly did nothing unreasonable given his wildly mistaken belief. “Knowledge of circumstances” must then mean something different from the doxastic perspective of the agent, i.e. the “reasonable person” must hold beliefs that the actual agent did not hold or lack beliefs that the actual agent held. The drafters of MPC now certainly owe an explanation of the nature of “the reasonable person” not least concerning which beliefs the “reasonable person” holds in a given situation and why the offender should be judged against this particular background.

The first of the above problems has led some authors to despair of using MPC’s definition of negligence as a basis for convicting defendants at all. For example Jerome Hall, commenting on MPC’s definition of negligence, holds that “it would be a great step forward to exclude negligent behaviour from the scope of penal liability.”⁴⁹ He supposes that “No one should be punished unless he has clearly acted immorally, i.e. voluntarily harmed someone, and unless a criminal sanction is both suitable and effective,”⁵⁰ and then simply denies that an unawareness of a risk *per se* constitutes moral blameworthiness in an action or omission of an action. I shall not go into his arguments to this conclusion here. Rather I shall take as my point of departure the intuition that the “honestly mistaken” rapist Smith *was* morally blameworthy for the rape of Mrs. Morgan and really *had a mens rea*, understood in the very broad sense of “a blameworthy state of mind.” The only way to make good this intuition is then to abandon MPC’s *mens rea* definitions, since the best MPC could offer was an account based on a state of mind that Smith did *not* have; an awareness of the risk of raping Mrs. Morgan.

Given my earlier conclusion that the locus of Smith’s moral blameworthiness could only be the epistemic blameworthiness of his belief that Mrs. Morgan consented, it would seem then that the only candidate for a *mens rea* in Smith’s case *is* his very epistemically blameworthy belief that Mrs. Morgan consented. Thus, the only way to avoid the unappealing conclusion that in “honest mistake” cases, there can be no *mens rea* except from a kind, which is not really a mental state at all, the notion of epistemic blameworthiness is indispensable. On this line of approach then, Joshua Dressler’s account of the broad “culpability meaning” of the term “*mens rea*” is still too narrow:

Broadly speaking, “*mens rea*” is defined as “a general immorality of motive,” “vicious will,” or an “evil-meaning mind.” Although each of these phrases has a slightly different connotation, “*mens rea*” as used here suggests a general notion of moral blameworthiness, i.e., that the defendant committed the *actus reus* of an offence with a morally blameworthy state of mind. For current purposes, this may be termed the “culpability” meaning of “*mens rea*.”⁵¹

However, Smith did not have an “evil-meaning mind” or even a “morally blameworthy state of mind” at the time he raped Mrs. Morgan. What he had was a mistaken belief.

However, to restate an earlier point, this does not mean that the source of Smith's blameworthiness for the rape cannot be tracked to this mistaken belief: Smith can be held *epistemically* blameworthy for holding that belief, and for that reason he can be held morally blameworthy for the harm he inflicted by the actions prompted by his mistaken belief. In order then to capture the "very broad" sense of "*mens rea*" as a state of mind that makes for the agent's moral blameworthiness for an offensive action, epistemically blameworthy beliefs must be allowed to count as a *mens rea* insofar as they moved the offender to inflict the harm that it was his legal duty not to inflict.

As I shall now proceed to argue, even jurisdictions upholding MPC's definition of negligence need the notion of epistemic blameworthiness in order to avoid total arbitrariness in their invocation of "the reasonable person" when passing judgments of criminal negligence. I have already once touched upon the problem of the cognitive perspective of "the reasonable person": Since in "honest mistake" cases the reasonable person cannot have the same beliefs as the supposedly negligent agent had in the relevant situation, which beliefs *would* the fictitious reasonable person in fact have in this situation? I shall argue that the only non-arbitrary answer to this question seems to be: The same beliefs as had the negligent agent, apart from her epistemically blameworthy beliefs!

If, e.g. the rapist Smith is to be judged negligent in his actions, "the reasonable man" could hardly be allowed to believe persistently throughout the non-consensual intercourse that Mrs. Morgan consented and that he did not have a reason to doubt that she consented, for in that case he would never have stopped, even for a moment, to be convinced that Mrs. Morgan consented, and would have perceived no reason for abandoning the intercourse. Supposedly even "the reasonable man" (if he has a human psychology) does not typically do what he does not take himself to have any reason to do.⁵² Now why should Smith's actions be judged against a cognitive background not presupposing his actual beliefs that Mrs. Morgan consented and that he did not have a reason to doubt this? It does not seem right here to simply point to the unreasonableness of those beliefs. Suppose for instance that Smith has an appropriate excuse for holding such beliefs, e.g. he could simply be irredeemably cognitively deficient⁵³ (very dumb and unable to improve). Would it still be just to judge Smith against a standard that presupposes cognitive powers he could not possibly have had? Many jurisdictions in effect have answered "yes" to this question, holding only physical cognitive impairments such as blindness relevant to the cognitive standards against which a negligent defendant can be held.⁵⁴

There is, however, something deeply disturbing about this answer. As pointed out by Larry Alexander, this practice effectively amounts to holding cognitively deficient persons strictly liable for the harm caused by their cognitive deficiency, which flies in the face of the idea that the purpose of *mens rea* requirements for criminal liability is exactly to alleviate the injustice of imposing strict liability for most types of offences. In Alexander's words, there is simply "no moral difference between punishing for inadvertent negligence [e.g. harmful actions due to mistake of fact] and punishing on the basis of strict liability."⁵⁵ Alexander makes his point using the infamous 1971 case

State v. Williams.⁵⁶ In this case a pair of poorly educated parents was convicted of the statutory manslaughter of their baby son, who died of gangrene after his parents had failed to bring him to a doctor. The evidence suggested that the parents had simply been unable to appreciate the seriousness of the child's condition, believing him to suffer only from a benign "tooth ache" in no need of professional medical attention. However, both the trial court and the appeal court found it proved that the parents were guilty of statutory manslaughter due to their criminal negligence⁵⁷ in failing to bring the child under professional treatment. The appeal court judged that "a man of reasonable prudence under the same or similar situations" would have taken the child to a doctor.⁵⁸

I agree to Alexander's conclusion that in this sad case the defendants might as well have been held strictly liable for the death of their son, since, plausibly, the epistemic standards to which they were held were such as they could never have aspired to: Even if their belief that the child did not need professional medical attention was false and perhaps even highly unreasonable, it does not seem farfetched to assume that they were appropriately excused for holding this belief given the kind of agents they were, and thereby also excused for their unfortunate omissions, since, on the basis of this belief, their actions were as conscientious as could be demanded. This suggests that only if they were in fact blameworthy for holding the unfortunate belief under the circumstances, could there be any justice in holding them to standards not hypostatizing this belief. If this had been acknowledged by the judges of the case, perhaps they would not have added the misery of incarceration to the Williams's grief of losing their son.

Considering another famous "honest mistake" case might add further support to this claim. In the 1875 case *Gordon v. State*⁵⁹ a young man was charged with attempts of under age voting at a general election. The young man was convicted, despite evidence that at the time he deposited his ballot, he believed himself to be old enough to vote. He based this belief entirely on the testimony of his mother and another older relative. It seems plausible that if this defendant should be accused of acting negligently on the definition provided by MPC, it would be highly unjust to judge him against the standards of "a reasonable person" not holding the mistaken belief that he was old enough to vote. A natural way of accounting for this intuition is to point to the plausibility of the claim that the young man was not epistemically blameworthy for holding the false belief that he was old enough to vote. After all, what better base could he have had for this belief than the testimony of his mother!

In conclusion, it is strongly suggested by the above examples that the cognitive perspective or "knowledge" of "the reasonable person" relevant to a judgment of criminal liability based on MPC's definition of negligence must take into account the epistemic blameworthiness of the beliefs relevant to the offender's harmful acts, if blatant injustices are to be avoided.

An account of the notion of epistemic blameworthiness could therefore be of immediate relevance to judges and juries in "honest mistake" cases like the ones considered above, whether or not they accept MPC's definition of negligence. The topic of the present study should therefore also be of some interest to legislators and

drafters of penal codes. However, I shall not here try to suggest how the notion of epistemic blameworthiness may in fact be integrated into actual legislations or penal codes. Legislation serves much wider purposes than probing the moral blameworthiness of convicted agents. Arguably, the law must also be so formulated that it is possible to prove the guilt of most guilty defendants in actual legal settings. If the law is overloaded with complex requirements for criminal liability aiming to secure that no blameless agent is ever convicted, we risk that too many criminals get acquitted due to the overly heavy burden of proof imposed upon the prosecution. This risk may perhaps sometimes be more important than the risk of occasionally convicting a blameless defendant. However, weighing such concerns against each other in actual situations touches upon pragmatic concerns, which I cannot hope to gauge accurately in the present context.

The point stands, however, that the notion of epistemic blameworthiness provides a promising guideline for *mens rea* judgments of criminal liability in “honest mistake” cases. This establishes yet another dimension of wider relevance for the results presented in this work.

1.5. EPISTEMIC BLAMEWORTHINESS AND THE NOTION OF EPISTEMIC JUSTIFICATION

In the preceding paragraphs we have investigated various ways, in which a proper grasp of the notion of epistemic blameworthiness is immediately relevant to practical concerns. However, the notion also has a narrowly meta-epistemological relevance, insofar as several authors have simply equated a justified belief with a belief not epistemically blameworthy or insisted that at least a lack of epistemic blameworthiness constitutes an important dimension of epistemic justification.

Arguably this *deontological* conception of epistemic justification has some historical and intuitive prominence. William Alston has submitted that

In any event, the most natural understanding of ‘justified’ in application to belief is tied to the supposition that terms in the “required”-“permitted” family are applicable to doxastic attitudes like believing ...⁶⁰

And in a similar vein:

Most epistemologists who have attempted to explicate justification have set out a concept of this sort [i.e. a deontological concept]. It is natural to set out a deontological concept on the model of the justification of behavior. Something I did was justified just in case it was not in violation of any relevant duties, obligations, rules, or regulations, and hence not something for which I could rightfully be blamed.⁶¹

An example of a prominent recent epistemologist who have at least partially subscribed to a deontological conception of epistemic justification is Alvin Goldman, who coins his conception of so-called “weak epistemic justification” in perspicuously deontic terms:

On one conception, a justified belief is (roughly) a *well-formed* belief, a belief formed (or sustained) by proper suitable, or adequate methods, procedures, or processes. On another conception, a justified belief is a faultless, blameless, or non-culpable belief ... (...) ... I ... call the first conception the *strong* conception

and the second the *weak*. Each of these seems to me a legitimate conception. Each captures some chunks of intuition involving the term “justified” (in its epistemic application).⁶²

For reasons that shall emerge fully in Chapter 5, I shall generally prefer to replace the intricate notion of epistemic justification by several more specific dimensions of epistemic desirability for belief. Still, a sufficient grasp of the notion of epistemic blameworthiness remains of imminent relevance to the untiring meta-epistemological debates over the proper understanding of the concept of epistemic justification.

1.6. ALTERNATIVE NOTIONS OF EPISTEMIC BLAMEWORTHINESS

The notion of epistemic blameworthiness, which I shall primarily be concerned with here concerns an agent’s blameworthiness for her holding of a particular belief under certain circumstances. However there are at least three other perfectly legitimate senses in which an agent can be “epistemically blameworthy,” one of which shall prove of some importance later. Thus, in order to avoid unnecessary confusion I shall briefly present these notions.

First, an agent may imaginably be blameworthy for *another* agent’s holding of a particular belief under certain circumstances. For example, in the *Morgan* case as presented above Mr. Morgan intuitively bears some of the blame for his colleagues’ mistaken belief that Mrs. Morgan consented. After all he plotted to make the colleagues form this belief and was even successful in his malicious plan. This sense of epistemic blameworthiness in which an agent can be blameworthy for *other* agents’ epistemically undesirable beliefs, shall not concern me much. I shall briefly touch upon it in Section 10.6.

Another sense in which an agent can be “epistemically blameworthy” does not directly touch upon the epistemic desirability of her beliefs at all, but concerns the way in which she performs her cognitive role in her community. Here, arguably, an agent may be blamed for failing to display a certain epistemic *virtue* in her conduct. Consider, e.g. the sometimes epistemic virtue of partial [cognitive] autonomy, defined by Frederick F. Schmitt as an agent’s partial reliance on her own observations.⁶³ Schmitt observes that the virtuousness of an agent’s partial autonomy depends on the cognitive role of that agent:

[W]hether partial autonomy is required depends (among other things) on whether the subject is an expert on the topic of the given belief ... (...) ... I would propose that it comes to the claim that doctors need to know more [about medicine than laypeople] if they are to be epistemically virtuous.⁶⁴

It is soon brought out that in this passage the phrase “to know more” is not intended to mean “having more true beliefs” or even “having more true beliefs with certain epistemically desirable properties” (as, e.g. justification), but is rather intended to mean “being more epistemically virtuous.” The specific “epistemic virtues” incumbent upon an agent assuming the cognitive role of a medical expert is brought out in the following passage:

If an immunologist believes that thyroiditis is typically caused by an autoimmune attack on thyroglobulin, and she does so on the basis of testimony without much partial autonomy, she may well be justified but her

belief is not virtuous. To be a doctor is to assume a complex role, of which medical expertise is a central component. Being a medical doctor entails deep engagement in medical topics, and this generally requires partial autonomy.⁶⁵

In other words, if an agent is to assume the role of a medical expert, it is not enough that she relies entirely on the testimony of other experts, even if her beliefs based on this testimony are both true and “justified”: We further expect her to have investigated empirically into the matters on which she offers her expert opinion, i.e. we roughly expect her to have personally conducted what qualifies as successful empirical research within her supposed field of expertise. If she has not done so, imaginably she may be *blameworthy* for not having performed her cognitive role properly.

This sense of blameworthiness deserves the label “*epistemic*” due to the importance that an agent’s lack of, e.g. partial autonomy may have for her contribution to the service of her community’s epistemic goals. Schmitt submits:

Expert roles enhance the ability of the division of [cognitive] labour to serve epistemic goals by concentration the work of ascertaining the truth on a topic in the hands of a few, increasing the reliability of sources and relieving laypersons of the deep burden of partial autonomy.⁶⁶

Thus, an agent’s failure to be as partially autonomous as she should be, may damage her community’s chances of getting to the truth about certain matters, which is obviously undesirable from an epistemic perspective. However, an agent’s epistemic blameworthiness in this sense is not essentially linked to the epistemic undesirability of her particular beliefs in the sense of epistemic desirability, which I shall prefer here (see Chapter 5). Schmitt, taking *justification* to be the chief epistemic desideratum for beliefs, expresses a version of this point thus:

A doctor who gets his information from a renowned medical expert or from a medical encyclopedia is as justified as I am [in believing this information].⁶⁷

Thus, this possible sense of “epistemic blameworthiness” is not conceptually related to the sense that I shall primarily discuss, and I shall not be further concerned with it.

It is important, however, to keep the last-mentioned notion of “epistemic blameworthiness” safely apart from a notion that bears crucially on the notion of epistemic blameworthiness central to the present study: An agent’s blameworthiness for actions or omissions as a result of which she later comes to hold epistemically undesirable beliefs. As I shall later argue (Chapter 12), such actions or omissions may be held blameworthy, insofar as the agent stood under an *intellectual obligation* not to perform them.

For an initial appreciation of this point, consider again the unhappy case *State v. Williams* from Section 1.4, the case where, due to their ignorance of medical matters, two parents held a mistaken belief about the severity of their son’s illness. Now, in Larry Alexander’s words,

suppose that at some point before the baby became ill, the Williamses were invited to attend a free class for parents on how to recognize medical emergencies of children. Suppose that the Williamses realized that, given this class’s importance relative to alternative uses of their time, they should attend this class, and that not to do so would be to take a substantial and unjustifiable risk of failing to perceive a medical

emergency. Nonetheless, they do not attend the class, but watch television instead. And suppose that had they attended the class, they would have learned about abscessed tooth, gangrene etc., and would likely have recognized the need to seek prompt medical treatment for William [their baby son].⁶⁸

Given these suppositions there certainly seems to be something to the claim that the Williamses were epistemically blameworthy for their later mistaken belief about the health condition of their son, based as it were on poor-quality evidence.⁶⁹ However, if they were not blameworthy for staying home watching television instead of attending the class (e.g. they could have been very ill that night or had to watch over their sick son) the epistemic blameworthiness of their mistaken beliefs again evaporates: Certainly they would then be *excused* for their belief's epistemic undesirability, insofar as they would then had had no past opportunities to improve their cognitive dispositions with regard to children's health matter that they should have pursued.

The kind of appropriate excuses I suggested above strongly suggest that moral values play into ascriptions of blameworthiness in the sense relevant here: moral obligations to perform alternative actions such as taking care of a child or saving other people from harm may alleviate an agent of blame for performing or omitting certain belief-influencing actions. If this is correct, epistemic and moral deontic evaluations stand in a two-tiered relation: an agent may be morally blameworthy for certain actions or omissions in virtue of acting on epistemically blameworthy beliefs (see Section 1.3) and those beliefs again may be epistemically blameworthy due to certain previous actions or omissions, which are in themselves blameworthy due to a lack of morally acceptable excuses. As shall later emerge, David Owens, in a recent study of deontic epistemic normativity, has argued that the blameworthiness of beliefs cannot hinge on the blameworthiness of actions.⁷⁰ I shall hope to make clear below that this view is mistaken and perhaps based on a failure to appreciate the close connection between practical and epistemic values sketchily brought out above.

1.7. THE STRUCTURE OF THIS STUDY

In an academic study it is often clarifying to offer a detailed presentation of the main conclusions in the introduction. However, this is hardly the case here: As shall emerge, almost all the notions I shall employ in my analysis of the notion of epistemic blameworthiness are seriously prone to misinterpretation. Further, the detailed analysis, which I present in Section 14.1 is complex. Presenting it conscientiously at this early stage would require me to pre-empt so many central matters that it would be impossible to maintain anything remotely resembling a transparent argumentative structure.

I shall thus proceed in a different manner, presenting in the course of my argument three analyses of the notion of epistemic blameworthiness in order of increasing accuracy and sophistication. This progression of analyses shall then act as the framework within which I can discuss the import of the notions that prove central to the present concerns.

The argument will have the following over-riding structure:

In Chapters 2–6 I shall lay down the conceptual groundwork for dealing with the bulk of the current literature on deontic epistemic normativity as well as for presenting my positive contribution to the field. Even though these chapters contain much of indispensable importance to the argument of the later chapters, less scrupulous readers may want to skip them.

In Chapter 2 I shall discuss the notion of belief, which is of central relevance, since it is an agent's holding of a *belief* that is the object of the sort of deontic epistemic evaluation I shall aim to understand more fully. I shall contrast the doxastic attitude of *belief* with the related doxastic attitude of *acceptance*, and argue that beliefs are indeed appropriate objects of deontic evaluations in the relevant sense.

In Chapter 3 I shall then pave the way for my first and basic analysis of the notion of epistemic blameworthiness. I shall argue that the notion of epistemic blameworthiness is indeed fit to act as a fundamental notion in the context of epistemic deontology. I shall then lay down a number of meta-level requirements that an adequate basic analysis of this notion must satisfy.

In Chapter 4 I shall present a basic analysis satisfying these meta-level requirements. Roughly I shall argue that an epistemically blameworthy belief is a belief which is undesirable from an epistemic perspective, and which the believer has no appropriate excuse for holding nevertheless. I shall dismiss as unfounded the possible objection that a belief might be epistemically blameworthy, even if it not epistemically undesirable in the relevant sense. In Chapter 5 I shall then present a substantial analysis of the relevant notion of epistemic undesirability. This will allow me, in Chapter 6, to offer good reasons for preferring my basic analysis of the notion of epistemic blameworthiness to a rival basic analysis of the notion recently advanced by Bruce Russell.

Chapter 7 is a vital chapter. Here I shall engage with the entangled topic of the relation between ascriptions of epistemic blameworthiness and an agent's power to exercise various modes of control over her beliefs. I shall argue that an agent cannot be epistemically blameworthy for holding some belief, unless she enjoys or enjoyed some mode of control over it. I proceed to offer a basic taxonomy of modes of doxastic control, and argue that this taxonomy covers the territory better than similar taxonomies offered in the literature.

In Chapters 8–11 I shall deal with specific modes of doxastic control and their relevance to ascriptions of epistemic blameworthiness. I shall reject in Chapter 8 the influential position commonly known as “doxastic voluntarism” holding, among other things, that normal agents may form and suspend a belief as a matter of performing an action. I shall argue that doxastic actions, if they are possible at all, are simply too exotic to be of relevance to ascriptions of epistemic blameworthiness. In Chapters 9–11 I shall then discuss how alternative modes of doxastic control may in fact matter crucially to ascriptions of epistemic blameworthiness. Chapter 11 contains most of my positive contribution to this discussion.

In Chapter 12 I proceed to discuss the notion of an *intellectual obligation*: I shall argue that agents may in fact stand under obligations to exercise certain modes of

doxastic control over their beliefs and that appeals to such obligations play a crucial role in ascriptions of epistemic blameworthiness. However, it shall also be argued that an agent is not epistemically blameworthy in virtue of violating the intellectual obligations incumbent of her, unless further conditions are met.

One of these conditions, I shall argue, is that the agent must have violated her intellectual obligations with a relevant kind of foresight or inadvertence to the risk of coming to hold epistemically undesirable beliefs. In Chapter 13 I shall engage with the difficult task of shedding at least a dim light on the notions of foresight and inadvertence relevant here.

At last then, in Chapter 14 I shall present my most elaborate and conclusive analysis of the notion of epistemic blameworthiness. I shall defend this analysis against some general worries and attempt to highlight some of its interesting consequences. Not least I shall here have the opportunity to discuss various important types of excuse for epistemic blameworthiness.

Finally, in the appendix-like Chapter 15 I then argue that freedom from epistemic blameworthiness of the type analysed here makes a good candidate for capturing at least one viable sense of the notoriously elusive notion of epistemic autonomy.

Readers only interested in my positive contributions should focus primarily on Chapter 7 and Chapters 11–14. If not interested in problems regarding foresight and blameworthy inadvertence to risk, the slightly technical Chapter 13 may also be dispensed with. However, I lay down a broad foundation for the study of deontic epistemic normativity. This will require me to introduce quite a bit of novel terminology and sometimes operate on a high level of abstraction. Each chapter will therefore need to presuppose the mastery of concepts and analyses earlier introduced.

BELIEF AND ACCEPTANCE

Abstract. In this chapter I take on the task of specifying suitable uses for the terms “belief” and “acceptance” in the present context. I argue that a belief is best construed a mental state of conviction to a certain high degree. On the other hand acceptance is best construed as the mental action of “going along with” a proposition in one’s reasoning. I bring out the wider importance of these conceptions to the aims of this study. In Section 2.2 I take on the objection that really the deontic evaluation of acceptances is the only type of evaluation important to socially significant concerns like the ones presented in Chapter 1. I conclude that deontic evaluations of beliefs remain of central relevance to such concerns.

2.1. THE DOXASTIC ATTITUDES OF BELIEF AND ACCEPTANCE

It should be obvious that its stipulated objects of normative evaluation must crucially inform any normative theory. For example a theory of the good-making features of airplanes is unlikely to resemble a corresponding theory for French cooking. However in epistemology many authors have been reluctant in scrutinizing the nature of basic objects of epistemic evaluation such as *beliefs*. The attitude is even encountered that no such analysis is strictly required, since the term “belief” has a “family of perfectly legitimate meanings.”⁷¹

In the present context I cannot acquiesce in such laxity due to the imminent danger of confusing the attitude of belief with its “grand cousin” acceptance. As shall emerge, conflating these doxastic attitudes is fatal to a fruitful approach to the issue of doxastic control, which is central to epistemic deontology (see in particular Section 8.3.2.1). I do not here deny that “belief” and “acceptance” are in fact highly ambiguous terms with a rich family of meanings. Thus I shall not attempt to offer anything like an account of the ultimate nature of either belief or acceptance. Rather I shall attempt to fix uses for these terms that are at least suitable for the present purposes.

I shall consider only beliefs belonging to human individuals. This is not to deny that non-human individuals (e.g. dogs) may hold beliefs or that a collective of individuals may hold a belief not accountable for in terms of the mental properties of the individual members of that collective.⁷² However, as shall emerge, if the notion of epistemic blameworthiness should apply to such believers, I would have to argue that several other notions apply to them as well. I would, e.g. have to argue that such believers might act with foresight; that they can stand under obligations etc. This would be too tedious, given the already complex nature of the present topic.

Further, I will make two basic assumptions. Firstly, I shall assume “the eminent suitability of belief as a subject of responsibility and normative evaluation generally” (to use James A. Montmarquet’s fortunate formulation).⁷³ This is no uncontroversial assumption, neither in a historical, nor in a recent perspective. Historically, the concept of belief standardly invoked in recent epistemological literature is far from

commonplace. On the modern conception, belief is an attitude *common* to the states of unjustified and justified belief as well as to at least some instances of knowledge. Arguably, e.g. Kant, for all his conceptual subtlety, lacked a corresponding notion.⁷⁴

In recent time, the concept of belief as a fundamental concept in epistemology has come under fire from two quarters. Firstly there has been a trend exemplified by Williamson (2000) that denies the conceptual fundamentality of the concepts of belief and justification and suggests taking the concept of knowledge as fundamental instead, even to the extreme of declaring it an “unexplained explainer” for epistemology.⁷⁵ I shall not evaluate the merits of this claim here. If Williamson’s approach to epistemic normativity gains general favour, I can only hope that my points translate into the new idiom without too much damage.

A different approach hostile to the fundamentality of the notion of belief has it that this notion is better replaced with a notion of acceptance as “a proper subject of normative epistemic scrutiny.”⁷⁶ I will evaluate this claim below (Section 2.2) and reject it, although the notion of acceptance remains highly important to my over-all concerns.

My second basic assumption will be that any belief or acceptance is a *propositional attitude* in the sense that it is appropriately labelled by an English phrase of the form “the belief that . . .” or the “the acceptance that . . .,” where the sub-phrases “that . . .” express propositions.⁷⁷ This is an assumption standardly made in the recent epistemological literature. It has however been subjected to severe criticism from some quarters. For example Daniel Dennett has argued that an agent cannot appropriately be ascribed some propositional attitude, unless she fully *understands* some linguistic expression of the relevant propositional content.⁷⁸ This spells trouble for my commonplace assumption, since in many cases an agent with an undeveloped language cannot really be claimed to fully understand the sentences standardly taken as linguistic expressions of the propositional contents of her doxastic states.

Dennett employs the example of a small boy who somehow takes his father to be a doctor, but has only a very rudimentary understanding of what a doctor is. Does this boy really *believe* or even *accept* that his father is a doctor? Insofar as he believes or accepts *something* to be the case concerning his father, what *does* he believe or accept? I agree that such questions raise legitimate concerns. However, exploring them would arguably overload the discussion with too many intricacies. In the formulations I shall provide throughout, it shall be assumed that we are talking of beliefs and acceptances with full-blown propositional contents. If such are in fact rarer than generally assumed, the analyses I shall provide will be of a similar limited applicability. For ease of expression I shall use throughout the strictly incorrect locutions “belief that p” and “acceptance that p” to denote either a belief or an acceptance with the proposition <p> as its content. I here sacrifice terminological correctness in order to achieve a maximum of consistency with the literature I shall quote throughout.⁷⁹

Belief as a propositional attitude must be held strictly apart from *belief in* someone or something (e.g. God or the government), not least because *belief in*, in Wilfrid Sellars’ words, often implies “some kind of involvement or commitment or hope, faith, charity, expectation etc.”⁸⁰ not involved in propositional belief. An obvious

case is *belief in* God, in the familiar sense of trust in a divine providence, as contrasted with, e.g. a *belief that* God exists, which may obviously be entertained in a much more detached fashion than the former attitude (as is, e.g. typically the case with rationalist deists). The difference between the two senses of belief is not, as one might perhaps suspect, satisfactorily explained by simply noticing that propositional belief involves a proposition. Arguably one could *believe in* the truth of a proposition (in the sense in which one believes in God or in postmodernism). Whatever this means, however, it expresses a commitment to the proposition different from that expressed by the *belief that* locution. It is the latter commitment, which shall concern me here.

With the above in place, it is now time to get a clearer perspective on the doxastic attitudes of belief and acceptance and their differences. This is immensely important to the objects of this study: Not only will my final analysis of the notion of epistemic blameworthiness (Section 14.1) employ both notions; I will also need to rely on a fairly substantial understanding of the difference between beliefs and acceptances in the run of my later argument against the claim that belief-formations may count as actions (see in particular Section 8.3.2.1).

I shall adopt L. Jonathan Cohen's understanding of the notion of acceptance:

But in my sense to accept that *p* is to have or adopt a policy of deeming, position or postulating that *p* – that is of going along with that proposition (either for the long term or for immediate purposes only) as a premise in some or all contexts for one's own or other's proofs, argumentations, inferences, deliberations, etc. whether or not one assents and whether or not one feels it to be true that *p*.⁸¹

Accepting that *p* in this sense clearly does neither require one to believe that *p*, nor to form the belief that *p* in the future.⁸² James A. Montmarquet has suggested that, since an agent may accept a proposition for years and thus be in a “continuous state” of acceptance, acceptance is best construed as a kind of *disposition*.⁸³ This however seems unconvincing: An agent may, e.g. swim for a considerable period of time and thus be in a continuous state of swimming without swimming being anything like a disposition; swimming is clearly a kind of action or pattern of action and the same seems to go for acceptance. In Cohen's words:

Acceptance is thus a mental act (as what was called “judgement” often used to be), or a pattern, system or policy of mental action.⁸⁴

It should be noted that acceptance in the present sense may well be truth-oriented, as is most clearly the case when an act of acceptance is motivated by an agent's desire to accept true propositions.⁸⁵ Some authors have even restricted the notion of acceptance to truth-oriented acceptance. For example Mark Kaplan has submitted:

X accepts *p* is just shorthand for “X would defend *P* were her sole aim to defend the truth.”⁸⁶

In the present context I see no need for this restriction, as acceptance remains a distinctive attitude irrespective of its particular motivation. For example in a plausible scenario, a medical doctor under primitive circumstances may need to quickly administer some drug or other to a fatally wounded patient. Despite her lack of decisive evidence, she may be strongly motivated to accept the proposition that a certain drug be effective against the patient's pains. This acceptance is perfectly consistent

with a radical reluctance on the doctor's part to defend the truth of that proposition should that be her sole aim.

Let us turn now to the notion of belief. I shall conceive of a belief that *p* as a mental state of conviction to a high degree that *p*. However, the English language may well lack a single term for the relevant strength of conviction short of "belief" itself. Nevertheless, Wilfrid Sellars has made a valiant attempt at such a characterization when he states that, if an agent believes that a statement is true, "in a certain sense the believer takes it for granted, regards it as already settled, as not a matter of debate or of internal puzzlement that"⁸⁷ this statement is true.

Some further considerations in the wake of Sellars' above point will perhaps help us improve our grip on the notion of belief. It should be noted that neither I, nor seemingly Sellars, regard a conviction in the relevant sense as a *feeling*⁸⁸ in any straightforward sense. My primary reason for this stand is the following: Expressions of belief simply seem markedly different from expressions of garden-variety feelings that are propositional attitudes such as, e.g. gratitude. In a crucial sense, self-ascription of belief is a matter of "extrospection" rather than introspection: In reporting my beliefs I report how I take *the world* to be, rather than how I take my mental life to be. In this sense, expressions of beliefs are not "auto-biographical statements" at all.⁸⁹

The strength of conviction required for belief still deserves some investigation. Most importantly, it would seem that the minimal degree of conviction required for belief must be set rather high. Nevertheless, e.g. L. Jonathan Cohen seems committed to the thesis that even "weak" degrees of conviction that *p*, such as hunches, qualify as beliefs, insofar as he states that "a person who accepts everything that he believes is dangerously credulous."⁹⁰ This only seems to be the case if the degree of conviction required for belief is set at a very low level: Surely a person that accepts every whimsical hunch that pops into her mind is seriously at fault and perhaps even dangerous under certain circumstances. On the other hand, a person that, *ceteris paribus*, accepts only propositions of which she is strongly convinced, seems overly meticulous rather than credulous.

A mere hunch that *p* does not constitute anything like a belief that *p* in the relevant sense. This contention seems consonant with the way we ordinarily employ the notion of belief: Surely it is psychologically implausible that a detective should believe the chamber maid to be the professor's murderer, while at the same time having a hunch that the butler did him in. It would seem that his hunch precludes him from believing in the butler's innocence, and would perhaps even typically induce him to believe that the butler is guilty. However, there is no reason why it should not be possible that the detective be without *any* belief as to who murdered the professor, while retaining his hunch and perhaps even *accepting* for purposes of investigation that the butler was at foul. In normal parlance, it seems that the detective could sincerely state: "I have a hunch that the butler did him in, but honestly I do not really believe it."

Belief requires a degree of conviction considerably stronger than a mere hunch. However, one may easily overdo in the other extreme Sellars' point that belief involves some kind of being "settled." For example David Owens, in his search for an "epistemic notion of belief" (the one that supposedly "dominates the literature

on the analysis of knowledge”⁹¹), has explicitly required belief to imply a knowledge claim:

Nothing less than thinking that you know could be the psychological state which together with justification and truth (and perhaps something else), ensures that you do know. And so the beliefs that interest me imply claims to knowledge.⁹²

Owens introduces this analysis to distinguish the notion of belief relevant to epistemological concerns from a notion often employed in ordinary parlance, where a person may utter “I believe it will rain tomorrow,” meaning really something like “If I have to make a guess, my best guess is that it will rain tomorrow.”⁹³ This notion of belief clearly does not require conviction in the truth of the weather forecast, and hence is crucially different from the notion invoked in the present context.

It is far beyond the scope of the current study to seriously evaluate the merits of the renowned analysis of knowledge as justified-true-belief advocated by Owens and others. If, however, as Owens suggests, this analysis requires beliefs to involve convictions as strong and transparent as to issue in overt knowledge claims, its domain of application seems seriously constrained compared to that probably imagined by most of its adherents. Firstly, this model would seem to rule out all instances of opaque or subconscious belief: If I am not aware that I believe *p*, I cannot very well *claim* to know it. The thesis that beliefs are always transparent to the believer, is implausible to say the least: I may well come to realize that I have for many years (mistakenly) believed Donald Duck to wear shoes, even though I have for long not given any conscious thought to the issue. Owens might here attempt to weaken his analysis and argue, in my opinion somewhat unnaturally, that an agent may still “claim” to know a proposition, even if this “claim” is subconscious or only *implicit* in her overt behaviour. However, an appeal to subconscious claims seems to do little explanatory work.

Further, very often we do not even implicitly claim to know what we believe. In an extreme case I may even be aware that all available evidence counts against the truth of *p*, and so recognize that I cannot know that *p*, while still sticking to my belief that *p* due to some strange cause or other. Such cases of *akratic* belief (or *akratically held* beliefs) are well described in the literature.⁹⁴ The following case introduced by James A. Montmarquet may count as a standard example⁹⁵: Bubba is a rather dim-witted, though not mentally retarded, citizen of the American state of Tennessee. Bubba’s son, Billy Bob, stands accused of murder and the prosecution has presented considerable evidence indicating his guilt. Despite admitting the obvious strength of the case against Billy Bob, Bubba sticks to his belief that Billy Bob is innocent, although faced with the massive evidence he would not claim to know this.

Now, on Owens’ conception of belief, akratic belief is simply impossible. Doxastic akrasia becomes a contradiction in terms. Consequently, in the above case, Owens is forced to conclude that Bubba does not *believe* that Billy Bob is innocent (in the sense relevant to epistemological concerns).

This conclusion is highly implausible: It is far from clear that Bubba would be one inch more convinced of Billy Bob’s innocence, if the legal evidence supported this

conviction: In this case, one could well imagine the blockheaded Bubba's degree of conviction to be exactly the same as in the former scenario. However, in the latter case Bubba's belief that Billy Bob is innocent would appear to be in some sense *justified*, insofar as his belief satisfies the epistemic desideratum of standing in no conflict with the available evidence. Hence in this case, Owens could hardly deny that Bubba's belief is of the kind relevant to epistemological concerns. But, as Owens has expressly specified belief to be a psychological state, and Bubba's state of conviction is unaltered, it follows that Bubba's akratic belief from the former case is also of the right kind for epistemology. Consequently it would seem that Owens' requirement that belief involves a degree of conviction as firm as to issue in knowledge claims is simply too strong.

The best we can say about belief in the present context then seems to be the following: A subject holds a belief that *p* if, and only if, she is convinced to a high degree that *p*. Conviction here is best characterized as a type of mental state, without however being a type of feeling in any straightforward sense. The degree of conviction that *p* required for a belief that *p* must be significantly stronger than a mere hunch that *p* but need not be as strong as to issue in a claim to know that *p*. Within these constraints I will have to assume the notion of belief sufficiently well understood by my reader.

2.2. THE DEONTIC EVALUATION OF BELIEFS AND ACCEPTANCES

There are at least two important reasons why acceptance might be taken as a type of doxastic attitude more immediately relevant than belief to issues of epistemic blameworthiness: Firstly, since acceptance is construed as a mental action, it is on a par with other kinds of action in terms of deontic evaluation. Thus, agents may be presumed blameworthy for their acceptances in the same sense that they are blameworthy for other kinds of action, and epistemic blameworthiness may be treated as a special type of blameworthiness for actions. The problems springing from the realization of our failure to directly control the content of our beliefs (see Chapter 8) are thus avoided: According to a common conception, which I shall not challenge here, agents are typically in a position to directly control their actions. If a shift to *acceptances* as sole objects of epistemic blame is warranted, our inability to directly control the content of our *beliefs* loses any direct relevance to the issue of deontic epistemic normativity.

The second principled reason for "shifting" to acceptances is perhaps more fundamental: It would seem that it is our acceptances, rather than our beliefs, which make a difference to the way we interact socially. At a first blush then, it would seem incredible that beliefs and their evaluation should ever be of wider relevance to moral and legal concerns like those presented in Section 1.3 and Section 1.4 above.

I do not take such arguments to diminish the importance of the deontic epistemic evaluation of beliefs. The first line of reasoning is undoubtedly well guided insofar as deontic evaluations of acceptances avoid at least some of the controversies surrounding the deontic evaluation of beliefs. Inconvenience alone, however, is hardly enough to exclude the feasibility of the more problematic type of evaluation.

The second argument is more substantial. If, namely, belief is mistakenly held to be a kind of feeling on a par with, e.g. gratitude, the argument seems quite strong. Many social obligations are obligations to *express* certain feelings rather than to actually feel in certain ways: I am obliged to show gratitude towards my aunt for the dreadful pull-over she gave me for my birthday and may do this by writing her a thankful letter etc., but I am hardly obliged to actually *be* grateful for that distasteful piece of clothing. This is why the Christian command to sincerely “love thy neighbour” is so radical: Perhaps I may (sometimes) be obliged to treat my neighbour *as if* I loved him, but to *actually* always love him is quite another matter.

Equivocating belief with a feeling thus easily leads to the idea that *expressions* of belief, e.g. acceptances, rather than *states* of belief, are the real objects of deontic interest. This idea gets further support from considering some examples of supposed epistemic obligations from the literature:

Lorraine Code, e.g., considers as a prime example of an epistemic obligation the obligation to *know* in which side of the road to drive when visiting a foreign country.⁹⁶ Exactly what *knowing* such a rule amounts to is a controversial issue. However, I find it far from obvious that an obligation of the above-mentioned sort applies to propositional knowledge or even beliefs in the first place: As long as the driver consistently *accepts* the proposition that, e.g. one must drive on the right side of the road in Denmark, it would seem that her inability to *believe* this proposition is inessential, at least insofar as the traffic authorities are concerned.

Another of Code’s examples concerns the notorious “Flat Earth Society”, an association of people believing the venerable proposition that the Earth, contrary to widespread recent opinion, is in fact flat as a pancake. Now, Code submits, having a “flat earther” in the position to withhold grants for space research would pose moral – as well as epistemic problems.⁹⁷ According to Code then, a person in such a position would be obliged to believe that the Earth is not flat, or at least obliged not to believe that it is flat. However, again it is far from clear that any problems of either a moral or an epistemic nature arise, as long as the grant distributor consistently succeeds in *accepting* current mainstream cosmology when considering applications for funding, even if, at the bottom of her heart, she believes that “travelling around the globe” is just plain nonsense.

Considerations as the above make a serious case for the claim that it is mostly in order to exhibit desirable patterns of *acceptance* that obligations applying to doxastic attitudes are imposed on us in the first place. However, this is far from making beliefs irrelevant as objects of epistemic deontic evaluation, as a careful examination of Code’s examples should bring to light.

Assume that the driver from Code’s first example is indeed obliged to *accept* the proposition that one should drive at the right side of the road in Denmark. Now, consider what it would take for the driver to fulfil this obligation without *believing* the afore-mentioned proposition. Certainly it is highly implausible that the driver should have no belief concerning which side of the road to drive in, so in order for belief and acceptance to come apart in this case we would have to imagine her (falsely) believing that one should drive in the *left* side of Danish roads. Given

the obvious great danger ensuing from being mistaken on this issue, it seems clear that nothing short of threats to her life, or perhaps a strong suicidal desire, could motivate the driver to separate her acceptances from her beliefs concerning such matters.

In the “Flat Earth” example it is more plausible that the grant distributor could “overrule” her belief that the Earth is flat in order satisfactorily to consider applications for space research. Career ambitions and fears of social stigmatisation would probably suffice. Still I suspect the unease we would feel in having a “flat Earther” employed at the Space Commission, does not directly spring from general distrust in her epistemic “character” or from a consideration of her “epistemic irresponsibility” in holding the flat Earth belief, as Code would have it.⁹⁸ Certainly the flat Earther’s professional track record could be excellent and her professional conduct as “upright” as reasonably demandable.

Rather, I will submit, our unease plausibly springs from a consideration of the *frailty* of her separation of her flat-Earth beliefs from her spherical-Earth acceptances. As soon as her specific motivations to separate her acceptances from her beliefs vanish, we would expect her appropriate conduct to follow down the drain. I suspect such considerations vindicate, e.g. John Hardwig’s suggestion that “trust in the testimony of others . . . involves trust in the character of the testifiers,”⁹⁹ only that being of the right character does not immediately imply sincerity in the expression of one’s beliefs, as one should perhaps expect; rather it implies an appropriate pattern of (truth-oriented) acceptance: A “flat Earther” at the Space Commission board should pose no problem as long as we could be assured that she would never give in to her erratic beliefs when fulfilling her professional duties. However, given the specific nature of her beliefs we would probably not be willing to grant her that kind of trust: Nobody can be expected to be that “upright.”

As I hope to have sufficiently indicated by now, an agent’s acceptances very often depend upon her beliefs. It is simply not the case as stipulated above that our beliefs do not matter to our social conduct: In order to perform the mental act of accepting a proposition, under many circumstances we simply need to believe it. Further, as indicated by the “flat Earther” example, very often an agent cannot be trusted to satisfactorily maintain an important position in our community while sticking to certain beliefs.

At this point the danger of equivocating belief with our garden-variety feelings can be more accurately gauged¹⁰⁰: For most of our feelings, it is simply not the case that they matter directly to the fulfilling of our social obligations, even if harbouring certain feelings often makes fulfilling these obligations much easier. Perhaps only intense love and hate impose the immediate force on our actions exerted by even very benign beliefs. This consideration, if anything, points to the massive importance of the attitude of belief in our cognitive economy. The cases of “honest mistakes” considered in Section 1.3 and Section 1.4 were all cases of action or omission prompted by the relevant agents’ false beliefs. To submit in such cases that the agents’ acceptances are what really matters, does not convincingly locate the source of moral blameworthiness or criminal guilt: It only interposes yet another “honest mistake”: The acceptance

prompted by the relevant false belief. The deontic evaluation of this belief remains the central issue.

This does not mean that a one-to-one relationship obtains between obligations to believe and obligations to accept. For example it is highly likely that an agent could be blameworthy for failing to accept a proposition, even if she fails to be able to believe it. As, e.g. submitted by L. Jonathan Cohen: “Ideally . . . a scientist would dispense altogether with belief in the truth of his favoured hypotheses and rest content with accepting them.”¹⁰¹

Altogether the relation between obligations to believe and obligations to accept appears to be immensely complex. This, however, does not relegate obligations to believe to strange second-rank obligations as the above considerations have made clear: An agent’s beliefs typically determine her actions in a very strong sense. Thus, an agent’s management of her beliefs is inherently significant from a social perspective. In the light of this observation, it should no longer be a source of puzzlement how the notion of epistemically blameworthy belief could have the practical significance stipulated in the introduction to the present study. Neither should we be too surprised if the standards for appropriate deontic evaluations of beliefs turn out to be socially informed to a wide degree.

APPROACHING A CONCEPTION OF EPISTEMIC BLAMEWORTHINESS

Abstract. In this chapter I set out to defend my choice of the notion of epistemic blameworthiness as fundamental. In Section 3.1 I defend my choice of the basic term “blameworthiness” over such terms as “blame,” “permission,” “prohibition” and “deontic justification.” In Section 3.2 I proceed to account for my use of the notoriously ambiguous, but highly popular term “responsibility.” I thereafter set out to discuss some general constraints that may apply to an adequate analysis of the notion of epistemic blameworthiness. In Section 3.3 I discuss a set of distinctions applying to types of analyses (theories) of the notion. I argue first that an adequate analysis of the notion must be *historical* or *retrospectivist* in the sense that the legitimacy of an ascription of epistemic blameworthiness must depend entirely on matters previous in time to that ascription (Section 3.3.1). I then argue (Section 3.3.2) that an adequate analysis must be *substantialist* in the sense that it must maintain that an agent’s epistemic blameworthiness depends on *her* history, rather than, e.g. solely on features of the community imposing that ascription. Finally I argue (Section 3.3.3) that an adequate analysis must be *non-sentimentalist* in the sense that it does not entail that an epistemically blameworthy agent is the object of some reactive attitude such as resentment or indignation. This paves the way for working out a satisfactory analysis in the ensuing chapters.

3.1. EPISTEMIC BLAMEWORTHINESS AS A FUNDAMENTAL NOTION FOR EPISTEMIC DEONTOLOGISM. SOME TERMINOLOGICAL POINTS

Some authors¹⁰² have claimed that deontic notions such as blameworthiness, praiseworthiness, obligation, permission, etc. have no place in epistemology, and that all legitimate normative epistemic concepts are for that reason “purely” evaluative. It is part and parcel of the present study to work out how agents may in fact be blameworthy for their beliefs under many circumstances. In doing so it is vital to get painstakingly straight on the basic terminology, as some popular terms within the territory are fraught with dangerous ambiguities.

I shall term “a deontic epistemic evaluation” any evaluation that entails an assessment of an agent’s blameworthiness or praiseworthiness for her holding of a particular belief. I shall label “deontic” all normative terms (predicates, notions, etc.) that convey an assessment of the above type. Matters of praiseworthiness, however, shall not concern me much. Rather, I shall take *blameworthiness* as the fundamental notion for the brand of epistemic deontology advocated here. I generally prefer the term “blameworthiness” over the more versatile term “blame,” as the latter term is often used to single out an agent as the *cause* of some event or occurrence, without thereby involving a deontic evaluation of that agent: I may, e.g. correctly “blame” my clever opponent for my bad fortune at the poker table, without holding her in the least *blameworthy* for my losses.¹⁰³ Unless explicitly pointed out, I shall use the term “blame” in a deontic sense on the instances where, for ease of expression, I shall employ it.

I shall use the terms “blameworthy belief” and “blameless belief” (i.e. non-blameworthy belief) as short-hand for a belief that the relevant agent is, respectively, blameworthy or blameless in holding. The same goes for all other evaluative terms applying to beliefs. For example I shall take an “epistemically desirable” belief to be a belief that it is desirable from an epistemic perspective that the relevant agent holds. I shall discard as useless in the present context any distinction between, e.g. a belief that an agent is justified in holding and a justified belief “proper.”¹⁰⁴ In a strong sense, I submit, the branch of epistemology that I here engage in evaluates *believers*, not beliefs in themselves.¹⁰⁵

Having established “blameworthy” as my basic deontic predicate I shall now consider some other popular deontic terms. Some, as will emerge, are more useful than others. First of all, I shall abstain from using the popular terms “epistemically virtuous” and “epistemic virtue.” Virtue seems to suggest praiseworthiness, and issues of praiseworthiness, as already indicated above, shall not concern me here. Moreover issues of praiseworthiness are not immediately connected to issues of blameworthiness. For example: I may fail to be blameworthy for stealing your car, since I did not steal it, but I am hardly praiseworthy for failing to steal your car for that reason; perhaps the only reason why I did not steal your car was that I just did not have the right opportunity to steal it.

An alternative term for “blamelessness from an epistemic perspective” that has enjoyed some popularity in the literature is “deontological justification” (see Section 1.5). I shall avoid employing this notion when possible due to the immense controversies concerning the nature of epistemic justification. However, the following statement due to Matthias Steup might still be useful:

The idea underlying the deontic conception of epistemic justification is that, *whenever* a belief is deontologically justified, no epistemic obligations have been violated.¹⁰⁶

I find it safe to presume, as does probably Steup,¹⁰⁷ that obligations can be understood as norms, which an agent, insofar as she lacks an appropriate excuse, is blameworthy for not adhering to. “Obligation” I shall thus take to be a safe term in the present context.

However, e.g. the term “ought,” though included in Steup’s list of deontic terms,¹⁰⁸ is strongly ambiguous between a deontic and a merely evaluative sense. Thus it is more dangerous here, where much depends on keeping deontic and non-deontic evaluative predicates cleanly apart. As noted by Wilfrid Sellars, the ambiguity of the term “ought” is particularly dangerous, when evaluating agents¹⁰⁹: It is easy for any statement concerning how an agent *ought to be*, to presume that it is in fact deontic, and hence illegitimate, should the agent not be responsible for her failure to adhere to the norm. Such confusions are highly prone to arise within epistemological contexts, when discussing which methods an agent *ought* to employ in her inquiries concerning some subject matter. Here it might easily be thought a valid objection to a proposed non-deontic norm that agents cannot possibly be responsible for adhering to it. The dangers relating to the term “ought” also surround any use of the term “should.” For that reason the terms “ought” and “should” are best avoided as basic terms within the

present context. It almost goes without mention that the famous and much-discussed slogan of “ought implies can” exemplifies these dangers: Clearly some “oughts,” namely the non-deontic “oughts,” do not imply a “can” in any sense. When the terms “ought” and “should” are occasionally employed below, I shall always use these terms in a deontic sense.

William Alston and Matthias Steup have united in preferring “permission” and “prohibition” as primitive notions for epistemic deontologism¹¹⁰ (Plausibly, “permission” here must be taken in the sense that something is permitted, if it is not prohibited). In Steup’s case his terminological hierarchy does not seem significantly to influence his writing, insofar as conceptions of deontic justification in terms of obligations, blame and permissibility interchange freely in his writing. In Alston’s case, however, his terminological preferences seriously inform his position: He expressly states his basic analysis of the notion of deontological justification in terms of prohibition and permission respectively:

To say that S is justified in believing that *p* at time *t* is to say that the relevant rules or principles do not *forbid* S’s believing that *p* at *t*.¹¹¹ (my italics)

And further:

The justification of anything, H, consists in H’s *being permitted* by the relevant principles: epistemic, moral or whatever.¹¹² (my italics)

Responsibility and blameworthiness, on the other hand, Alston takes to be “normative consequences of an agent’s situation with respect to what is required, prohibited or permitted.”¹¹³

Alston’s terminological hierarchy is somewhat unfortunate in the present context: Prohibitions and permissions are terms that naturally apply to actions rather than to occurrences in general. For example even it is forbidden or permitted by the relevant authorities that I smoke in my office, the *occurrence* that my cigarette is lit is not forbidden or permitted in anything but a strained sense. On the other hand it makes perfect sense to say that I am blameworthy for that occurrence under suitable circumstances.

Given such natural constraints on the normative force of prohibitions, it is hardly surprising that Alston quickly jumps to the conclusion that deontic epistemic justification must require some sort of *direct* control over beliefs, a control “naturally” modelled on “the maximally direct control we have over the motions of our limbs and other parts of our body, the voluntary movements of which constitute ‘basic actions,’ actions we perform ‘at will,’ just by an intention, volition, choice or decision to do so.”¹¹⁴

Thus, for Alston the issue of deontic epistemic normativity is quickly reduced to more or less the following two questions: (1) Do holdings of belief resemble “basic actions” in terms of voluntary control? (2) If beliefs are not subject to voluntary control in the sense that “basic actions” are, can any sense be given to the notion of a “permissible belief”?

Alston resoundingly answers question (1) in the negative, and his answer to question (2) follows suit.¹¹⁵ Now, as he takes an analysis of deontological justification in terms

of epistemic obligations to be subordinate to an analysis in terms of permissibility, he rejects the viability of the former notion as well:

It is held that one is justified in believing that *p* only if one fails to violate any intellectual obligations *in believing that p*, i.e., only if correct intellectual standards permit one to believe that *p* under the conditions that do obtain. I reject this as a viable suggestion for a condition of justification, since it presupposes that belief is under effective voluntary control . . . and it seems clear to me that it is not.¹¹⁶

I believe that Alston errs here in tying the notion of an epistemic obligation closely to the notion of permissibility (from an epistemic perspective). As Alston has amply demonstrated, it makes dubitable sense to talk of beliefs being permitted or prohibited. Nevertheless, I shall argue, a working notion of epistemic *blameworthiness* is indeed feasible. Consequently the terms “permission,” “prohibition” and their correlates are best avoided in the present context.

3.2. THE NOTION OF RESPONSIBILITY

Some special attention must now be given to the term “responsibility” and its correlates. I shall only employ this term to denote the kind of *authorship* of some occurrence necessary, but not sufficient, to merit legitimate ascriptions of blameworthiness to the authoring agent for that occurrence: More precisely, I shall say that an agent is *responsible* for some occurrence if, and only if, she enjoys or enjoyed some mode of control over it (whether she has in fact exercised that control or not). In Chapter 7 below I will provide an argument to the conclusion that responsibility in the present sense is indeed required for epistemic blameworthiness as well as a full-blooded account of the ways in which an agent can control her beliefs in the relevant sense. I am convinced that the proposed sense of the term “responsibility” captures at least one of its ordinary usages, though I shall not argue the point at length here.

The conception of responsibility invoked here is *normatively neutral* in the following sense: From an agent’s being responsible for an attitude, nothing follows concerning the deontic status of that attitude: The agent may be blameworthy, blameless, praiseworthy, etc. depending on the relevant circumstances. However, in the literature, the terms “responsible” and “responsibility” have been used in a number of senses, several of which are not normatively neutral in this manner. To clear away obvious possibilities of misunderstanding I will briefly consider the most frequently occurring alternative senses.

First, “responsible” has been used by a number of authors to mean, what is here meant by “epistemically blameless.”¹¹⁷ Second, to some a “responsibility” can be what in the present study is referred to as an obligation,¹¹⁸ and “responsible” in this vein may mean “obliged.”¹¹⁹ Running these alternative senses of “responsibility” together with the one preferred here may result in near-total terminological opacity. For example the sentence “An agent is blameless for holding a belief which she has authored, if she has not thereby violated her epistemic obligations” may come to read: “An agent is responsible in holding a belief for which she is responsible, if she has not violated her epistemic responsibilities.” In the light of such obscurity in a context surely not in need of further complications, I will stick to the sense of “responsible” fixed above.

A further possibility of confusion arises from the distinction enforced by T.M. Scanlon between responsibility as attributability (roughly the sense preferred here) and so-called “substantive responsibility.”¹²⁰ The latter notion is somewhat problematic to capture accurately. Scanlon offers the following example: Suppose that a father has no duty to set up his estate, such that his son and heir cannot spend his money unwisely, as in fact the son does. Now, Scanlon contends, the results of the son’s unwise spending is his own *substantive responsibility*, insofar as the son has no one but himself to blame for that outcome.¹²¹ Scanlon hesitates to offer anything like an “iffy” analysis of the notion of substantive responsibility, which is perhaps best captured by the following analysis: An agent is *substantially responsible* (in Scanlon’s sense) for a consequence of an action if, and only if, whatever blame or praise is appropriately attributable to any agent for that consequence uniquely befalls that agent.

Arguably, in the suggested interpretation Scanlon’s notion of substantial responsibility captures the sense of responsibility sometimes invoked in ordinary parlance by phrases like “it was her responsibility” etc. Yet, whether or not this interpretation is true to Scanlon’s intentions, it is important to notice that substantive responsibility is something very different from mere responsibility in the sense invoked throughout this work. As I shall argue that responsibility for holdings of beliefs is in fact a species of responsibility for consequences of belief-influencing actions or omissions rather than a species of responsibility for actions proper, Scanlonian substantive responsibility must be kept safely out of the picture not to disturb intuitions already set to labour hard enough.

Finally it should be noted, that the term “responsibility” may in some contexts often be used in a way closely associated with the one preferred here: It might be held that an agent is responsible for an action or a state of affairs if, and only if, she is *accountable* or *answerable* for the action, in the sense that it is *appropriate* that she gives an account of (explains) her intentions in bringing about that action or that state of affairs, and if failing to do so may face legitimate corrective measures.¹²² Although responsibility as authorship and responsibility as accountability may often overlap in the consideration of responsibility for actions, they are significantly different notions. For example in a context of mandatory education, a neophyte to spelling may well be held accountable for her failure to spell correctly, even though her mistakes was not something she could have controlled in any sense. Just as in the case of Scanlonian substantial responsibility, responsibility as accountability must therefore be kept safely apart from responsibility in my preferred sense.

3.3. SOME OVERRIDING DISTINCTIONS

3.3.1. *Prospectivism Versus Retrospectivism*

The bulk of the present study shall be occupied with stating and explicating an analysis of the notion of epistemic blameworthiness, i.e. we shall mostly operate on a meta-normative level. Ensuingly I shall also aim to demonstrate that the meta-normative claims defended shed new and important light on first-order problems concerning the epistemic blameworthiness of particular agents in particular scenarios.

However, in order to engage fruitfully with the meta-normative issue, I must begin at an even higher level of abstraction: It seems possible to impose several, as it were, meta-meta-normative demands that an adequate analysis of epistemic blameworthiness should satisfy and as shall emerge, one's stand on the appropriateness of these demands matters vitally to the kind of analysis one can give and the way one must interpret its components.

In their introduction to the meta-normative debate over the nature of moral responsibility John Martin Fischer and Mark Ravizza¹²³ explicitly introduce and implicitly suggest a number of distinctions between types of analyses of the notion of moral blameworthiness. Most of these distinctions can be made immediately relevant to the present epistemological context and I cannot in the present context avoid taking a stand on them. However, one distinction I shall take the license to ignore.

The ignored distinction divides "optimistic" and "pessimistic" theories of moral blameworthiness: Here "optimist" theories are *compatibilist* in nature, insofar as they hold that agents may be blameworthy for their actions (and the consequences thereof) even if "the thesis of determinism" is true, whereas "pessimist" theories deny this. Discussing the merits of the thesis of determinism and its possible relation to issues of epistemic blameworthiness would simply take me too far in the present context. Even with this issue left beside the discussion is complex enough, and I shall therefore simply assume that human agents are responsible for their actions under normal circumstances. This means that I shall simply ignore the possible relevance of the claim that our beliefs or belief-influencing actions are determined, whatever that exactly means.¹²⁴ Still, later (Section 8.3.2.2) I shall aim to debunk a particular attempt due to Matthias Steup to make "optimism" in the above sense directly relevant to the issue of epistemic deontology.

A number of distinctions explicitly enforced by Fischer & Ravizza, however, I shall consider, e.g. their distinction between "historical" and "non-historical" theories of blameworthiness. Here *non-historical* theories hold that the blameworthiness of an action supervenes on the properties of "the current time-slice" whereas *historical* theories hold the past (but not the action's future) to be relevant also.¹²⁵ This dichotomy may naturally be supplemented with an obvious third option considered by the authors in a different context: Certain "social-regulation" theories¹²⁶ hold the *future* to be the only relevant time-interval to issues of blameworthiness.

Formulated in a perhaps more suggestive terminology we are here presented with the choice between *a-temporal*, *retrospective* and *prospective* types of theories. Of course these options do not exhaust the logical possibilities. The option remains that the future and the past may *both* essentially matter to the blameworthiness of an agent. However, as shall emerge from the below discussion, this possibility need hardly be taken seriously.

A lot is clearly at stake here for my analysis of epistemic blameworthiness. As shall later be clear (Section 4.1), this analysis hinges crucially on the dual normative notions of epistemic undesirability (for beliefs) and inexcusability, none of which is easily given a prospective reading. In particular, I shall ultimately take inexcusability to hinge crucially on matters of doxastic control, which of course must be

exercised *prior* to the point of time when a belief is submitted to deontic epistemic evaluation.¹²⁷ Further, as shall emerge, the notion of epistemic undesirability can hardly be given a prospective reading either: It is fairly obvious that some epistemic desiderata are in fact retrospective relative to the belief under evaluation, most conspicuously the desideratum of *formation* by a reliable mechanism. Since my analysis of epistemic blameworthiness is thus constructed from retrospective components, my theory belongs squarely in the historical or retrospective camp. If compelling reasons for preferring an a-temporal or prospective theory can be given, I am in dire straits indeed.

Luckily the case against prospective theories of epistemic blameworthiness is strong, which can be seen from an analogy with the corresponding moral case. Historically prospective theories of moral blameworthiness have been offered as a way of optimistically meeting the determinist challenge. Bluntly put, prospectivists of the “social-regulation” type have argued that only the *future* benefits from imposing blame on an agent matter to the appropriateness of such deontic attitudes as resentment or indignation. In their eyes, the appropriateness of such attitudes at most *contingently* depends on the actual history leading to the event or occurrence blamed. A relaxed stance towards determinism may thus be adopted. Determinism may even be claimed to assure the effectiveness of the measures taken against condemned agents.¹²⁸

Perhaps the most coherent version of the above “social-regulation” strategy would be one that *replaces* blameworthiness with something like social salience as the prime justifier for our reactive attitudes. Such a strategy would abandon the notion of blameworthiness altogether (perhaps due to a pessimist stance on compatibilism) or take a cynical line, considering issues of blameworthiness and justice entirely – or almost entirely irrelevant to the normative status of attitudes such as indignation or resentment. In other words, this strategy would either reject the viability of any “iffy” analysis of blameworthiness or simply leave it to the armchair division. I believe a case can be made that this is the strategy actually followed by most actual theorists in this domain, although I cannot defend this claim here. We should only notice that the above versions of the “social-regulation” strategy do not pose any substantial threat to the present endeavour, since they leave the notion of blameworthiness untouched.

For the present purposes, however, it is important to rule out prospectivism concerning blameworthiness, even if this requires setting up a straw man.¹²⁹ We may here take *prospectivism* concerning blameworthiness to consist in the following claim: An agent is blameworthy for something if, and only if, holding her so blameworthy has desirable future consequences.

Prospectivism in the suggested sense is of course inconsistent with the historical approach favoured here. However, it also seems an immensely implausible theory: First of all prospectivism implies that any agent, even when she has absolutely nothing to do with a supposedly blameworthy event or occurrence, may appropriately be deemed blameworthy for that event or occurrence as long as ascribing such blameworthiness has the right kind of consequences. Thus, any agent may ultimately be blamed for anything, given the relevant future! Further this conclusion seems unavoidable, even if the prospectivist attempts to narrow the relevant class of consequences down

to those that strictly concerns the blameworthy event and its author: Suppose a hardened criminal has robbed a bank. Now perhaps blaming *me*, a seemingly innocent person, for the bank robbery may actually make the bank robber a better man. This could, e.g. be the case if he suffers from an intense hatred towards me, a hatred preventing him from living an ordinary life and only mollified by the blissful observation of my severe punishment. On prospectivism I then become blameworthy for the bank robbery, a bizarre proposition, which I would hardly be wise to concede.

On a more general level, prospectivism seems to sever the crucial link between blameworthiness and excusing past conditions, generally recognized at least since Aristotle¹³⁰: If prospectivism is true, no excusing appeal to the past can ever matter directly to the question of my blameworthiness for anything. This, I shall be satisfied to submit, is simply an unacceptable conclusion.

The above point easily carries over to the consideration of a prospectivist theory of *epistemic* blameworthiness. Here again it seems intuitively compelling that, if the notion of epistemic blameworthiness is to carry any weight, it cannot sever the link between blameworthiness and inexcusability: It must be possible to be excused from epistemic blame solely due to one's past, no matter the future. I shall discuss a number of such excuses. Perhaps the most intuitively appealing type of excuse from epistemic blame is cultural isolation (see also Section 14.3.1): Consider a native who holds seemingly bizarre beliefs about the past of his community on the basis of a very unreliable oracle. Now, even if such beliefs are strongly epistemically undesirable in a number of ways, it seems obvious that the native is not blameworthy for holding them, if only he is prevented to a sufficiently high degree from questioning the authority of the oracle by his upbringing and general social situation.

For this reason alone, prospectivism concerning epistemic blameworthiness is in deep trouble. Consequently I shall not take it seriously in what follows. In fact even allowing a prospectivist *component* in one's analysis of epistemic blameworthiness seems perverse: Following the lead of the above example, plausibly one's past may *totally* excuse one from blame, come what may: The future cannot be allowed to override excuses provided by past circumstances. At most then, future circumstances could be held to sometimes either establish – or acquit from – blameworthiness, when no excusing past circumstances obtain.

However, neither of these options looks particularly inviting. The first option just rehashes the problems inherent in a “purely” prospectivist position: If the past offers insufficient material to establish blameworthiness, drawing in the future seems immensely unfair to the agent under evaluation: She has simply done nothing that could merit an ascription of blameworthiness. On the other hand, the second option allows that future circumstances may alleviate from blameworthiness otherwise earned on the basis of one's past. In effect this means that, if things turn out a certain way, an agent may later rightfully claim that she was not blameworthy at an earlier instance, as she would have been, had the future turned out differently. This establishes appeals to future outcomes as being in effect sometimes appropriate excuses from blame otherwise earned by one's past, which strikes at least this author as an oxymoron at best.

Having put prospectivism safely aside, it remains to consider the a-temporalist option presented by Fischer & Ravizza. According to their understanding of a-temporalism¹³¹ the moral blameworthiness of an action supervenes on the properties of the relevant “current time-slice.”¹³² A-temporalism is a highly controversial position concerning moral blameworthiness for actions: At least to this author it is hard to see how, e.g. the past life of a criminal cannot but matter directly to the issue of her blameworthiness for a criminal offence. However, a-temporalism concerning morally blameworthy *consequences* of actions is just plainly incoherent: How can an event be blameworthy due its status as a consequence of an action without the past being relevant to its blameworthiness?

As I shall argue at length below, the only feasible account of epistemic blameworthiness has it that blameworthy beliefs are blameworthy consequences of past belief-influencing actions and omissions, not blameworthy events in themselves. Thus a-temporalism poses no serious threat to the analysis of epistemic blameworthiness developed here.

3.3.2. *Substantialism Versus Non-substantialism*

A distinction on which Fischer & Ravizza places some emphasis is the distinction between so-called “ledger” views and so-called “Strawsonian” views concerning the nature of moral blameworthiness. They quote Michael J. Zimmermann for the “ledger” view that “someone is blameworthy if he is deserving of such [moral] blame; that is, if it is correct, or true to the facts, to judge that there is a ‘debit’ in his ‘ledger’”¹³³ and conclude:

Ledger views construe our ascriptions of responsibility as first and foremost judgments concerning an agent’s moral value; the reactive attitudes and the associated practices of praising and blaming take on a secondary role, following from these primary assessments of moral worth like practical consequences.¹³⁴

The contrasting “Strawsonian”¹³⁵ view is presented thus:

[Peter] Strawson holds that it is these reactive attitudes [resentment and moral indignation]¹³⁶ and practices [of imposing praise and blame] themselves that are constitutive of responsibility... [R]esponsibility is grounded in nothing more than in our adopting these attitudes toward one another. Jonathan Bennett puts this point well: “My feeling of indignation at what you have done is not a perception of your objective blameworthiness, nor is it demanded of me by such a perception. It expresses my emotional make-up, rather than reflecting my ability to recognize a blame-meriting person when I see one...”¹³⁷

In short, according to Fischer & Ravizza, Strawson defended the following peculiar conception of moral blameworthiness:

Strawson’s theory holds that being morally responsible is nothing other than being a recipient of these [reactive] attitudes and a participant in the associated practises.¹³⁸

This seems to suggest that an agent is morally blameworthy for an occurrence if, and only if, there is a practice in her community of blaming her for this occurrence. And this is even the more charitable interpretation. If one takes Bennett’s line from above, we may even take Strawson to claim that an agent is morally blameworthy for

something if only she is a member of *my* community and *I* resent her or feel moral indignation towards her for that something.

In essence, this type of position has it that statements about the blameworthiness of an agent are really not statements about that agent at all (as would appear from the surface grammar) but rather statements about something entirely different, namely about a practice of imposing blame on that agent or having reactive attitudes towards that agent. I will term such a position “meta-deontic non-substantialism.” This position, in effect, amounts to a strong form of antirealism concerning blameworthiness. The opposite position, which holds statements of blameworthiness to be substantial statements about the merit of the relevant agent, I shall term “meta-deontic substantialism.”

It is easily seen that non-substantialism in the versions attributed to Bennett and Strawson above is deeply problematic: They leave practices of blaming agents *infallible*, in the sense that they leave no room for the obvious possibility that we might adopt a practice of resenting an agent for performing some undesirable action only to discover later that she was in fact not blameworthy at any time: As long as the relevant practice of resenting an agent obtains, that agent *is* blameworthy, period! I take it that this observation alone suffices to defeat “Strawsonianism” in Fischer & Ravizza’s sense. As the above concerns must apply equally to the epistemic case, where we should hardly expect infallibility either, it would seem that already at this stage we can reject a “Strawsonian” stance on the nature of epistemic blameworthiness. In fact non-substantialism generally appears to be mistaken: If our practices of ascribing blameworthiness are to remain fallible, it seems overwhelmingly natural to assume that what we may mistake are features of the *object* of such ascriptions, not features of our very practice: It is because *the agent* held blameworthy did not deserve this ascription that the ascribers were *wrong*, not because they made a mistake concerning what they were doing when they ascribed blameworthiness to her.

We may thus safely commit ourselves to meta-deontic substantialism. Ascriptions of blameworthiness are substantial statements about the supposedly blameworthy agent. Still, the demise of non-substantialism does not mean that we can safely commit ourselves to a “ledger view” in the above sense, as I shall now proceed to argue.

3.3.3. *Sentimentalism Versus Non-sentimentalism*

The “Strawsonian” view as presented by Fischer and Ravizza (see above) is hardly the view defended by Peter Strawson in writing at any point, particularly not in his 1962 paper “Freedom and Resentment”¹³⁹ employed by Fischer & Ravizza to justify their ascription. I will submit that, in misreading Strawson direly, they miss out on the really interesting distinction involving reactive attitudes such as resentment and indignation, which is in effect a distinction cutting through the domain of substantialist theories of blameworthiness.

To make good the first claim, consider the actual aim of “Freedom and Resentment,” rightly hailed by Fischer & Ravizza as a “landmark essay.”¹⁴⁰ Here Strawson discusses

the relation between the thesis of determinism and the concept of moral responsibility. He formulates the supposed problem involved in this relation thus:

[T]he real question is not a question about what we actually do, or why we do it. It is not even a question about what we would in fact do if a certain theoretical conviction gained general acceptance. It is a question, about what it would be rational to do if determinism were true, a question about the rational justification of ordinary inter-personal attitudes in general.¹⁴¹

From this quotation, we should already suspect that Strawson hardly holds such attitudes to be rational just in virtue of *being present*, as Fischer & Ravizza would have him: The question of the rationality of our attitudes is expressly *not* the question about which attitudes we actually have. Still Fischer and Ravizza are correct that “Strawson holds it that it is these reactive attitudes and practices themselves that are constitutive of responsibility,”¹⁴² they just overdo the point: Strawson’s real point is that the question of the *general* rationality of our practices simply cannot be answered at all *within* the framework of human rationality, whose standards must *operate* inside – and are *constituted by* those very practices. This *naturalistic* commitment to our actual practices is something very different from a commitment to the legitimacy of particular attributions of blameworthiness, as Strawson is quick to point out:

[S]uch a question [i.e. the question of the general rationality of our practice] could seem real only to one who had utterly failed to grasp the . . . [. . .] . . . fact of our natural human commitment to ordinary inter-personal attitudes. This commitment is part of the general framework of human life, not something that can come up for review as particular cases can come up for review within this general framework.¹⁴³

Strawson’s central point is thus that the question of the particular legitimacy of a reactive attitude can only be discussed *within* very general constraints with “common roots in our human nature and our membership of human communities.”¹⁴⁴ Our actual practices thus “constitute” the notion of responsibility, not in virtue of rendering actual particular ascriptions of blameworthiness infallible, as Fischer and Ravizza’s “Strawsonianism” would have it, but rather in the much more subtle sense of constituting the *framework*, within which particular ascriptions may be legitimately and rationally discussed.

In fact, Strawson does not in “Freedom and Resentment” present anything like an “iffy” analysis of the notion of moral responsibility (actually he does not have much to say about this notion, apart from a clear acknowledgement that it admits of rational discussion within a suitable framework). Rather, he presents a substantial naturalistic theory regarding the general rationality of reactive attitudes like resentment and indignation. I hold this theory entirely neutral as to what regards the thesis of this study, where it is simply taken for granted that issues of epistemic blameworthiness may be rationally discussed, leaving safely aside the metaphysical question of the constituency of the framework, wherein such a discussion takes place.

Above I presented strong reasons for rejecting Fischer & Ravizza’s “Strawsonianism.” Nevertheless, their presentation, even if profoundly un-Strawsonian in character raises an important issue that I cannot safely ignore. The problem is that, even with the demise of “Strawsonianism,” the contrasting “ledger view” set up by Fischer & Ravizza is not home free, as the two positions hardly exhaust the

possibilities. Fischer & Ravizza has it that “ledger” views must commit themselves to a certain detachment or coolness regarding reactive attitudes:

[T]he ledger view . . . (. . .) . . . suggests that assessments of responsibility can be made from an objective, uninvolved and detached standpoint.¹⁴⁵

It seems, however, that what is really distinctive of the ledger views is not that they hold reactive attitudes to be somehow irrelevant to the issue of blameworthiness. Rather, ledger views simply take statements about the blameworthiness of an agent to *entail* statements about the status of her “ledger.” There is absolutely no reason why a “ledger” theorist in this sense could not *also* hold that an agent’s blameworthiness essentially require the presence of certain reactive attitudes taken towards her. Actually Fischer himself is such a theorist:

This type of theory . . . (. . .) . . . holds that morally responsible agents, are not just those who, as a matter of practice, are recipients of the reactive attitudes; agents are morally responsible if, and only if, they are appropriate recipients.¹⁴⁶

It emerges, then, that the important distinction raised by the present discussion concerns the link between an agent’s blameworthiness and her reciprocity of reactive attitudes. Here I shall suggest a distinction, which I take to capture what is really at stake in Fischer & Ravizza’s discussion of ledgers and reactive attitudes. To get to the point I shall bypass the issue of moral blameworthiness and jump straight to the epistemic case:

One possible epistemic meta-deontic position is that the epistemic blameworthiness of an agent *entails* that that agent is the object of some reactive attitude with regard to her relevant belief. Call this position “epistemic meta-deontic *sentimentalism*.” Epistemic meta-deontic non-sentimentalism is then the position that epistemic meta-deontic sentimentalism is false.

Here, I have left it entirely open *who* adopts the reactive attitude and *when*. Since I aim to reject sentimentalism, I want the argument to be as general as possible. It should be clear, however, that if sentimentalism is true, it is something that an “iffy” analysis of the notion of epistemic blameworthiness cannot ignore. In this case of course the specific nature, location and temporality of the relevant set of reactive attitudes should be explicitly stated.

The general problem with Fischer & Ravizza’s presentation is that it strongly suggests the idea that sentimentalism entails non-substantialism, which is clearly not the case. If so sentimentalism could demand no serious attention, since, as argued, non-substantialism is a highly implausible position, leaving our actual practices infallible.

As we saw above, sentimentalism in ethics has been embraced by would-be Strawsonians like John Martin Fischer, and may indeed command a certain minimal degree of respectability. On sentimentalism, blameworthiness is simply similar to aesthetic notions like goodness of taste in food: Arguably it is incoherent to maintain that a particular types of cheese tastes well, while denying that anybody has ever adopted a positive attitude towards its particular flavour. This is perfectly consistent

with the claim that a verdict of good taste in a cheese is a statement about the cheese, not about any agent's reactive attitude towards it.

Still, it is hard to see why anyone should want to adopt epistemic sentimentalism if not out of an ill-guided reverence for Peter Strawson. After all epistemic sentimentalism entails the highly implausible conclusion that if a belief goes unnoticed, it is not blameworthy at all, no matter what the further circumstances. The only vaguely appealing reason to embrace any type of sentimentalism I can think of is the idea defended by Fischer & Ravizza that our practice of blaming agents "essentially involves reactions to one another with feelings and attitudes that stem necessarily from our concerned involvement as participants in a moral community."¹⁴⁷ Non-sentimentalism with its apparent coolness then seems to fly in the face of the committed, emotion-soaked nature of our actual practices.

Here, however, three points are in order: First, such intuitions can be re-directed to, e.g. a *truly* Strawsonian position regarding the justification of reactive attitudes, from which it does not follow at all that all *particular* instances of blameworthiness must be accompanied by reactive attitudes. Non-sentimentalism is perfectly consistent with the assumption that if no one ever adopted reactive attitudes towards other agents (no such practice existed), the notion of blameworthiness would have no extension, which is arguably all Strawson could be committed to.

Second, non-sentimentalism can warmly embrace the proposition that as a matter of fact blameworthy agents are *typically* the objects of reactive attitudes.

Third, non-sentimentalism may even concede that, due to the human cognitive make-up, without the guidance of the relevant kind of reactive attitudes agents would be at a loss concerning which other agents to appropriately blame. The detached "coolness" of the ledger view, as presented by Fischer & Ravizza is a caricature at best. I thus believe sentimentalism to be a position rife with obvious difficulties and without substantial arguments in its favour. I shall therefore embrace non-sentimentalism.

In conclusion: I hope to have demonstrated that a viable analysis of the notion of epistemic blameworthiness must be of a retrospectivist (historical), substantialist, non-sentimentalist kind. It must maintain that the epistemic blameworthiness of an agent's belief hinges on *her* past and does not vanish, just because no one actually resents or reproaches her for her holding of that belief. I shall offer an analysis satisfying these requirements below.

BLAMEWORTHY BELIEF AS INEXCUSABLY UNDESIRABLE BELIEF

Abstract. In this chapter I present a first basic analysis of the notion of epistemic blameworthiness. Roughly, I argue that an epistemically blameworthy belief is a belief, which is in a certain sense undesirable from an epistemic perspective, and which the relevant agent has no appropriate excuse for holding nevertheless. I briefly argue in Section 4.1 that this analysis satisfies the overriding requirements laid down in Section 3.2 above. In Section 4.2 I then take on a basic challenge to my analysis: The imagined, but hardly far-fetched, objection that agents may be epistemically blameworthy, even if their beliefs are in fact not undesirable in the relevant sense. I reject this suggestion after a consideration of the merits of its ethical correlate.

4.1. A FIRST ANALYSIS OF THE NOTION OF EPISTEMIC BLAMEWORTHINESS

Finally, I am now in a position to venture my basic analysis of the notion of epistemic blameworthiness:

Blame_{epist}: An agent is *epistemically blameworthy* for her holding the belief that *p* if, and only if,

1. She believes that *p*.
2. Her belief that *p* is *epistemically undesirable*.¹⁴⁸
3. She has *no appropriate excuse* for the epistemic undesirability of her belief that *p*.

The key notions of the above analysis are obviously epistemic undesirability and lack of appropriate excusability. I shall have much more to say about these notions in what follows. Suffice it here to notice some basic points:

“Lack of appropriate excusability” must be understood *externalistically* in the following sense: An agent’s being appropriately excused does not generally imply her actually stating this excuse or even her being able to state it. This should be clear from the actual excuses discussed below. For example I shall later endorse the view that “cultural isolation” can count as an appropriate excuse from epistemic blameworthiness given certain constraints (see especially Section 14.3.2). Roughly speaking, an agent can (at least sometimes) be excused for the epistemic undesirability of her beliefs if prevented from holding a relevant epistemically desirable belief due to her confinement within a particular community. Now certainly, an excuse of this kind is not something an agent could normally be asked to state: Obviously this type of excuse may still be appropriate, even if the agent is unaware of the existence of communities distinct from her own.

The notion of epistemic undesirability shall be discussed at length below. Suffice it for now to remark that this notion will be accounted for with reference to the definitive epistemic goal of holding only true beliefs. I shall have more to say on

this “truth-goal” below, where I shall aim to demonstrate that an agent may serve this goal in virtue of aiming for beliefs which satisfy a number of merely evaluative *epistemic desiderata*, where by a “merely evaluative” epistemic desideratum I shall mean an epistemic desideratum expressible without the use of deontic notions like, e.g. blameworthiness.

Before proceeding, note in provision that $\text{Blame}_{\text{epist}}$ satisfies the requirements laid down in the preceding section:

Certainly, epistemic blameworthiness must be a *historical* notion on $\text{Blame}_{\text{epist}}$, if the in-going key notions are historical. As shall emerge below, the epistemic desiderata discussed are in fact retrospective in nature. Further, as already pointed out in Section 3.3.1, the notion of excusability is essentially retrospective.

That epistemic blameworthiness is a *substantialist* notion on $\text{Blame}_{\text{epist}}$ should be fairly evident: An agent’s epistemic blameworthiness hangs essentially on *her* excusability.

Finally, for epistemic blameworthiness to be a *non-sentimental* notion on $\text{Blame}_{\text{epist}}$, it should not entail that the blameworthy agent is the recipient of a reactive attitude. Here, it should be clear from the account of epistemic undesirability offered below that a belief may be epistemically undesirable, even if this goes undetected by any agent. Further, given the observations on sentimentalism in Section 3.3.3, a sentimentalist analysis of the notion of lack of appropriate excusability is not viable. Just as in the case of blameworthiness itself, it seems very implausible to suppose that undetected cases of inexcusability do not exist.

For these reasons, I believe it safe to conclude that $\text{Blame}_{\text{epist}}$ satisfies the relevant meta-level requirements, and that we may proceed to consider its particular merits.

An ethical correlate of $\text{Blame}_{\text{epist}}$ is the following: An agent is morally blameworthy for performing an action if, and only if, the action is morally bad and the agent is not appropriately excused for the bad-doing instantiated by the performance of that action.

This might seem so obviously true that some might well feel that I am balancing more than dangerously on the border of utter banality in offering its epistemic equivalent as a basic analysis of epistemic blameworthiness. Of course this charge is warmly welcomed. Indeed I hope and suspect that $\text{Blame}_{\text{epist}}$ is true and even in some sense banal, in the sense that, once brought out, it seems highly plausible. However, as we shall see below, my preferred analysis of the notion of epistemic blameworthiness may be subject to serious challenges, even as it stands at this point.

4.2. THE PROBLEM OF EPISTEMIC “ACCUSES”

Michael Zimmermann has offered a tough challenge to an ethical correlate of $\text{Blame}_{\text{epist}}$ by arguing that moral blameworthiness is compatible with the absence of wrong-doing.¹⁴⁹ In other words: According to Zimmermann, an agent may be to blame (morally) for her performance of an action, even if that action is not undesirable along any moral dimension of desirability. He terms a reason for imposing such “counter-evaluative” blame on a good-doer an “accuse” (in contrast to an excuse, which is a reason for acquitting a wrong-doer of blame). The challenge to $\text{Blame}_{\text{epist}}$

is this: if appropriate *epistemic accuses* exist, $\text{Blame}_{\text{epist}}$ cannot stand. $\text{Blame}_{\text{epist}}$ certainly entails that epistemically desirable beliefs are always blameless.

Luckily I do not believe that Zimmermann's views pose any serious threat to $\text{Blame}_{\text{epist}}$. First his argument for the existence of appropriate *moral* accuses is far from appealing. Zimmermann submits the following pair of premises to get his argument going:

1. An action may not be morally wrong even if acted in the belief that it is morally wrong.¹⁵⁰
2. If an action is acted freely in the belief that it is morally wrong, it is morally blameworthy.¹⁵¹

From these premises follows the desired conclusion that if an action is acted freely in a belief that it is morally wrong, it may be blameworthy, even if not morally wrong.

The premise (1) seems uncontroversially true and shall not be contended. The premise (2) however, has some rather bizarre consequences, which Zimmermann is ready to face openly. For example an action becomes blameworthy, even if the agent was in fact *mistaken* in her belief that it was morally wrong. Very conscientious agents thus get overloaded with blame, while less conscientious agents go free. Zimmermann considers the example of the saintly Sarah, obsessed with devoting her every waking hour to the service of others. Sarah one day decides to sleep for an extra hour in the morning and does so in the belief that it is morally bad. In fact, Zimmermann reasonably assumes, it is not bad at all. Now, Zimmermann argues, Sarah is blameworthy for sleeping the extra hour, since she did this "freely in the belief" that it was morally wrong.¹⁵²

To see the bizarre impact of this claim, consider that Sarah's lazy sister, Susanna, who sleeps the extra hour without even giving morality a thought, goes entirely free of blame on Zimmermann's account. How can he make good these consequences?

Zimmermann's only defence is indirect. He argues (correctly, I think) that Sarah's blameworthiness does not strictly entail that she should be "blamed out loud":

This [Sarah's blameworthiness] is consistent with its being wrong for some reason to give overt expression to this judgment (since such expression might be misconstrued or have some other undesirable effect). Hence the claim that Sarah is blameworthy does not imply even that she should be blamed "out loud", let alone that she should be adversely treated in some more serious manner. And I am quite willing to concede that it might well be wrong to engage in such behavior towards such a saintly person as Sarah.¹⁵³

However, if blaming Sarah were to make any sense at all, it would seem that there must also be scenarios where the reasons overruling her adverse treatment are not present. Thus Zimmermann, in my opinion, does not succeed in "re-directing" our intuitions: Not only does it seem bizarre to blame Sarah "out loud" and acquit Susanna, it seems equally bizarre to blame Sarah in the first place. In the lack of a better defence, Zimmermann's premise (2) cannot command much credibility. Zimmermann has hardly made a strong case for the appropriateness of moral "accuses" for actions. Thus, his example has hardly strengthened the case for "epistemic accuses" either.

However, more compelling examples are around, examples making no appeal to anything like Zimmermann's second premise. Consider, e.g. the following case presented by Bruce Russell:

[E]xamples can be given where a person is blameworthy and has no legitimate excuse but does what is objectively right. Suppose, for instance, that I have every reason to believe that doing something is objectively wrong, say, giving an innocent person something I think is poison, even though, as it turns out, it is a substance that will cure the person of her fatal disease. Then, whether intentionally or out of laziness or inattention, if I give the person what is in the bottle marked "poison" but is really the medicine that will cure her, I will do what is objectively right, even though I will be blameworthy, and have no legitimate excuse for doing it.¹⁵⁴

This case immediately seems more compelling than Zimmermann's example, although, as shall emerge, everything hinges on what is meant by "objective rightness" above. The force of this example also highlights what is wrong with blaming Zimmermann's saintly Sarah: Unlike the hazardous physician from Russell's example, call her Ivy, Sarah had absolutely no *good reason* for believing that her supposedly blameworthy action was morally wrong: Her estimation of the moral worth of her sleeping late was simply entirely unjust to herself, and intuitively should therefore carry no weight in our assessment of the blameworthiness of her action. In contrast, it would seem that Ivy has eminently good reasons for believing her action morally undesirable. The label on the poison bottle alone should present her with the evidence necessary to conclude that her action was in fact harmful, although by sheer luck it turned out not to be so.¹⁵⁵

However, these considerations also highlight that Russell's notion of objective rightness does not include the notion of having good reasons. Thus to him, from "it was objectively right for agent A to perform action φ " it does not follow that "A had good reasons for performing φ ." This choice of terminology I consider somewhat unfortunate, since "having a good reason" is in some sense¹⁵⁶ an objective notion (e.g. it hardly follow from "A believes herself to have a good reason" that "A actually has a good reason") and having good moral grounds for the actions one performs is clearly desirable from a moral perspective. Since Russell reaches a similar conclusion,¹⁵⁷ this dispute is merely terminological.

In conclusion, neither the case of Sarah nor the case Ivy violates the principle that the blameworthiness of an action entails its moral undesirability: Sarah's action was not morally undesirable, but was hardly blameworthy either. Ivy's action was genuinely blameworthy, but morally undesirable too, insofar as it was based on bad moral reasons (even if it was not "objectively wrong" in Russell's sense). As shall be argued below, having adequate reasons for what one believes is in fact an important epistemic desideratum. In consequence an epistemological parallel to Russell's example would pose no threat to $\text{Blame}_{\text{epist}}$.

Now, concerning an agent's blameworthiness for the desirable *consequences* of her actions, the possibility of accuses seems even more remote than above. Consider again the saintly Sarah. Assume that as a consequence of her sleeping late, some mean-spirited old man does not get helped across the street. Sarah foresees this, and believes it a morally bad consequence that mars her saintly conscience. In fact

her fatigue in assisting the grumpy geezer accounts for her decision to stay in bed. However, the old man has never wanted Sarah's help. In fact she is a nuisance to him, but he has never been able to successfully convey this to her. Thus, the above-mentioned consequence of Sarah's snoozing is in fact desirable in any way. Now, even if, despite appearances, Sarah is indeed to blame for her innocuous sleeping late, it certainly seems highly bizarre to blame her for the old man's good fortune. Concerning blameworthiness for consequences then, the following principle seems basic: An agent cannot be blameworthy for a consequence of an action unless this consequence is somehow undesirable.

As I shall argue at length below, blameworthy beliefs are not instances of blameworthy actions, but rather blameworthy *consequences* of belief-influencing actions or omissions. This consideration sufficiently establishes that the notion of accuses poses no serious threat to my account of epistemic blameworthiness.

The demise of epistemic accuses, however, does not mean that $\text{Blame}_{\text{epist}}$ is home free. In Chapter 6 I shall engage with a rival account of epistemic blameworthiness due to Bruce Russell, which, although similar to $\text{Blame}_{\text{epist}}$ in some respects, is obviously incompatible with it. The divergences between $\text{Blame}_{\text{epist}}$ and Russell's analysis touch on fundamental issues regarding the exact nature of certain epistemic desiderata, which means that I cannot attempt to reject it before addressing the issue of epistemic desirability, as I shall now proceed to do.

EPISTEMIC UNDESIRABILITY

Abstract. In Chapter 4 I claimed that, in order to be epistemically blameworthy, an agent's holding of a belief must be somehow undesirable from an epistemic perspective. In this chapter I present my basic considerations regarding the notion of epistemic undesirability. In Section 5.1 I present a list of non-deontic epistemic desiderata inherited from William Alston. In Section 5.2 I submit that, in order to deserve the name of "epistemic" in the present context, the desirability of a belief's fulfilment of a certain epistemic desideratum must be accountable for relative to the *truth-goal* of holding only true beliefs. I end up in Section 5.5 with three key desiderata central to standard deontic evaluations of beliefs: Reasonableness, basing on good reasons, and formation by a reliable process. The exact understanding of either desideratum is a highly controversial matter. In Sections 5.3 and 5.4 I attempt to rule out the most obvious sources of misunderstanding. I argue ensuingly that a belief may be held to be *epistemically undesirable* in the relevant sense either if it is not reasonably held, not based on good reasons or not formed by a reliable process. The chapter concludes with a consideration of the dual imagined objections that I have conceived of the relevant notion of epistemic undesirability either too broadly or too narrowly. The first objection would have it that the relevant notion of epistemic undesirability is better equated with plain falsity (or non-truth, if one eschews bivalence). The second objection would demand that it be sensitive to such broader epistemic goals as understanding and wisdom. I argue that, while the first objection puts too heavy a strain on the notion of inexcusability inherent in my understanding of epistemic blameworthiness, the second presents no imminent worry to the present concerns.

5.1. EPISTEMIC DESIDERATA

The notion of epistemic undesirability is fundamental to $\text{Blame}_{\text{epist}}$. In the present chapter I shall attempt to carve out this notion in a way fruitful to evaluations of epistemic blameworthiness. Since most authors have been preoccupied mostly with what makes a belief epistemically *desirable*, I shall follow the tradition and approach the notion of undesirability through its opposite.

Traditionally *justification* has been taken to be the chief dimension of epistemic desirability for beliefs. By some authors, justification has been taken to be a merely evaluative (i.e. non-deontic, see Section 3.1) norm, a belief being justified if, e.g. formed by a process, which is in a suitable sense generally reliable. Others have submitted that justification is in fact a deontic concept.¹⁵⁸ Finally, some have argued that justification, although a deontic concept, in fact needs to be supplanted by a wider concept, e.g. termed "epistemization,"¹⁵⁹ which does the normative work that justification was traditionally thought to do, namely that of turning true beliefs into knowledge. "Epistemization" might then be thought to involve deontic as well as merely evaluative components. For example Matthias Steup has suggested that "epistemization" consists of epistemic blamelessness and reliable formation in combination.¹⁶⁰

I believe that there are indeed valuable intuitions pointing in the direction of deontic as well as merely evaluative conceptions of epistemic justification. However, in the face of the immense controversy surrounding this concept, I believe it more fruitful in the present context to follow William Alston in replacing “plain” justification with a number of dimensions of epistemic desirability or *epistemic desiderata*. This is not to deny the meta-epistemological importance of an adequate analysis of the notion of justification (or knowledge for that matter). Only to submit that the controversy surrounding the nature of an appropriate analysis of this kind stands in the way of providing a conception of epistemic undesirability fruitful to the present context.

Also this move should help to keep clear the crucial distinction between deontic and merely evaluative evaluations, which is easily obscured by employing a notion that might suggest either kind of evaluation. Alston has catalogued five non-deontic epistemic desiderata for beliefs, which may be paraphrased as follows.¹⁶¹

1. Basing relation: It is desirable that a belief is *based on* adequate grounds (adequate evidence, good reasons). In order for some ground (evidence, reason) to act as a basis for a belief, it must be the agent’s reason for forming (sustaining, revising) the belief. It is not enough that the agent simply *has* the adequate ground (good reason, adequate evidence) for holding (sustaining, revising) the belief.
2. Truth-conducivity: If based on a ground, it is desirable that this basis of the belief is *actually* indicative of the truth of the belief. It is not enough that it is thought to be so by the agent or her epistemic community. Further it is desirable that a belief has been acquired “in a reliable way.” If based on grounds, this desideratum amounts to the requirement that basing one’s belief on such grounds is a generally reliable way of forming beliefs.¹⁶²
3. Cognitive accessibility. It is desirable that a belief is based on a basis that the agent can cognitively access (without great difficulty).
4. Higher-level requirements. If a belief satisfies one or more epistemic desiderata (this one included) it is desirable that the agent is aware of this.
5. Coherence. It is desirable that a belief coheres with a suitable set of other beliefs held by the agent (or her epistemic community).

Arguably, much of the debate over the nature of epistemic justification has really been the playing of two or more of the above desiderata out against each other. Once the notion of justification is left aside, a more fruitful perspective is gained. In the present context, I shall be concerned merely with the two first items on Alston’s list, adequate basing and truth-conducivity, which I shall take to be genuine non-deontic dimensions of epistemic desirability for beliefs. These desiderata I shall supplement with one further epistemic desideratum, the desideratum of reasonableness (properly understood). I shall not attempt to vindicate the general desirability of the last three items on Alston’s list.¹⁶³ In this respect, the account of deontic epistemic normativity I shall offer is regrettably elliptic. Still, I trust, it will prove strong enough to satisfactorily deal with a wide number of cases.

In order to avoid unnecessary complications, I shall assume throughout that the fulfilment of the above epistemic desiderata is *universally* desirable, i.e. desirable

for *any* belief held by *any* agent in *any* context. If the desiderata are in fact context- or agent-relative in the sense that my above assumption is false, it should be clear how this could be accounted for without doing any damage to the overall structure of the present argument. It should also be noticed that my account of epistemic blameworthiness, even if based on universal epistemic desiderata, is still agent- as well as context-sensitive in the following sense: In virtue of her particular history an agent may well be to blame for holding a belief, even if another agent would be blameless for holding a belief with the same content in a similar context. Also an agent may be to blame for holding a belief in a specific context, if the context is one that she is appropriately judged blameworthy for entering. These considerations shall be given extensive treatment below.

5.2. THE IMPORTANCE OF THE TRUTH-GOAL

Some may have noticed that I have so far left out the most important of epistemic desiderata, the desideratum of truth itself. Of course I do not deny the obvious desirability of having true beliefs. On the contrary, I shall submit that truth is in an important sense *the* central epistemic desideratum, insofar as the desirability of all other epistemic desiderata must ultimately derive from it. I here follow a number of authors, a.o. William Alston, in conceiving of the “epistemic point of view” in terms of the “aim of maximising truth and minimizing falsehood.”¹⁶⁴ Now, as it stands at this point this aim or goal may be subject to the following worry: It would seem that, e.g. a stone comes out cognitively optimal, as it does not hold any beliefs at all, hence no false beliefs. The stone has in a radical way maximised truth (there is no beliefs it could hold, hence no further true beliefs it could hold) and minimized falsehood (it holds no false beliefs).

Alston therefore brings in the further qualification that the definitive epistemic goal is more specifically to be “defined by the aim of maximising truth and minimizing falsehood *in a large body of beliefs*.”¹⁶⁵ He submits:

The qualification “in a large body of beliefs” is needed because otherwise one could best achieve the aim by restricting one’s belief to those that are obviously true.

And further that

It remains true that our central cognitive aim is to amass a large body of beliefs with a favourable truth-falsity ratio.

However, Alston’s qualification, even if self-consciously “rough,”¹⁶⁶ seems beside the point. It must suffice to stress that, as a matter of course, the aim of maximising truth and minimizing falsehood (which I shall henceforth refer to as “the truth-goal”) only concerns *active* cognizers, i.e. agents who pursue cognitive ends. It does not concern stones, corpses, etc. The quantity of an active cognizer’s beliefs arguably depends contingently on her particular cognitive life-style and cannot concern epistemology proper. Thus I shall be satisfied to simply submit that the truth-goal concerns active cognizers and that such an agent has ultimately reached the truth-goal

when all of her beliefs are true. On this conception of the truth-goal, which I take to be the most natural, it is vehemently *not* supposed that piling up many true beliefs or even as many true beliefs as possible is somehow in itself desirable from an epistemic perspective, an idiosyncratic understanding of the truth-goal seemingly opted for by some authors hostile to the idea of awarding it a central role in their conception of epistemic normativity.¹⁶⁷ Epistemology, as it is conceived here, is not concerned with how *many* beliefs an agent holds or should hold, rather it is interested in the *way* agents manage their beliefs, insofar as they manage a mass of beliefs at all.

Recently epistemological “value-monism” hypostatizing the truth-goal as definitive of epistemic normativity has received a substantial amount of fire in the literature. Specifically some authors have rejected so-called “epistemic instrumentalism”¹⁶⁸ in arguing that the desideratum that one’s beliefs constitute *knowledge* (the “knowledge-goal”) is not derivable from the desideratum that they be true.¹⁶⁹ As already indicated in Section 2.1 above, I shall not be concerned with the complex notion of knowledge here. Thus I shall regrettably have to leave aside the important goal of knowledge, as well as the issue of its relation to the truth-goal. I hope this will not cripple my discussion. At least it should be clear how the knowledge-goal, if basic, could be integrated into my analysis of the notion of epistemic blameworthiness, although personally I must confess to being blank of intuitions concerning what should warrant blaming an agent for failing to *know* the propositional content of a particular belief, once epistemic instrumentalism is rejected.

I also think there is a reason for my lack of intuitions in this respect, which I shall shortly give: I find it obvious that blameworthiness of any kind must *prima facie* legitimise appropriate “rectifying or corrective measures”¹⁷⁰ against the blameworthy agent. Now, certainly it would be odd to rectify or punish an agent for failing to satisfy a desideratum that is essentially only desirable from her own perspective: Holding an agent blameworthy for failing to satisfy a desideratum must imply that the desideratum is of wider social importance. If these sketchy considerations are correct, the following point can be made:

The desideratum that an agent’s beliefs be true is obviously significant from a social perspective. An agent’s holding true rather than false beliefs highly increases the likelihood of her positive contribution to the satisfaction of her community’s wider cognitive goals. However, the knowledge desideratum, at least in its classical conception as epitomized by Descartes’ *Meditations*, narrowly concerns the agent herself. Aspiring to Cartesian-style knowledge is a solitary quest for individual certainty. Whether an agent’s belief counts as genuine knowledge on such an account in addition to being “merely” true, arguably cannot matter much to the social role played by that agent.¹⁷¹ This seems to explain why blaming an agent for failing to *know*, rather than merely believe, a proposition, seems an odd thing to do.

Some may wonder, why, when hypostatizing the truth-goal as the epistemic master desideratum, I do not forego the other epistemic desiderata and simply submit *falsehood* as the only significant epistemic indesideratum in my analysis of the notion of epistemic blameworthiness. The answer to this question I shall postpone to Section 5.5, as I cannot satisfactorily address it without first submitting the individual epistemic desiderata to a more thorough examination.

5.3. GROUNDS, REASONS AND EVIDENCE

In this section I shall attempt to explicate the relation between the notion of epistemic desirability and the notions of reasons, ground and evidence. This task is complicated since the notion of a *reason* may be employed in at least four different senses in an epistemological context:

First, the notion may be invoked in an entirely *explanatory* sense as in the phrase “the reason why A holds the belief that p was that she was hit with a shovel on her head yesterday.” Pointing to such *causally explanatory reasons* (or simply: explanatory reasons¹⁷²) may provide the causal explanation of the occurrence that A holds the belief that p. As shall later emerge the desideratum of truth-conducivity or reliability, which I have embraced above, may be formulated in terms of explanatory reasons: It is epistemically desirable that the explanatory reason why A holds the belief that p be the operation of a truth-conducive (reliable) process.

Explanatory reasons as the *explanans* featuring in a causal explanation, do not need to make any reference to the mental life of a cognizer at all: As stipulated, the explanatory reason why I hold a certain belief may, e.g. be the very non-mental fact that I was hit by a shovel on the top of my head. Typically, though, a viable causal explanation of the occurrence of a belief will appeal to mental states and events. In particular such explanations will typically quote the agent’s *reasons* for holding the belief that p in the sense that the agent’s awareness of certain pieces of evidence may well be the best *causal explanation* why she holds the belief. We may then say that her belief that p is *based* on that evidence, or that that evidence is the agent’s *basing reasons* for holding the belief. The notion of awareness here is of course vague. In the present context I will stay content with remarking that awareness intuitively includes the intentional states of perceiving or believing that certain evidence obtains.

It should be noted that an agent’s basing her belief on certain evidence in the present sense does not entail that this evidence was the basis on which she *originally* formed the belief. I may, e.g. originally have formed the belief that lions have yellow fur by studying a cartoon on television. However, in a more mature age, I came to discard cartoons as zoological authorities. Nevertheless I still believe that lions have yellow fur, due to watching specimens in zoo. In this example, as in many others, it clearly seems to be the case that, even though my awareness of some evidence originally induced in me a certain belief, my past awareness of this evidence no longer features in the best causal explanation of the fact that I presently hold the belief.

It should also be noted that many of our beliefs are not easily construed as *based* on reasons in the present sense. Take, e.g. my belief that I am a human being. Which mental states may a causal explanation of this belief appeal to? As shall be noted below, even if this question is ultimately unanswerable, I may still, in a suitable sense, have good reasons for believing that I am a human being.

This means that the basing relation desideratum is best construed as a desideratum that a belief be based on good reasons, *insofar as* it is based on reasons at all. Thus, I shall not argue that basing one’s beliefs on reasons is in itself epistemically desirable. The opposite conclusion would, e.g. have us declare a belief epistemically

non-desirable (i.e. undesirable, see Section 5.5) if not based on reasons, even if it was produced by a fully reliable method and held for perfectly good reasons. This seems an odd thing to do: relative to the truth-goal there simply does not seem to be anything wrong with such a belief: If produced by a reliable non-reason-related mechanism such as a certain drug, why should we desire from the perspective of the truth-goal that it was also produced or held due to an awareness of reasons?

Now, the notion of a reason may also be put to a slightly different use from the one it played in the notions of an explanatory or a basing reason as presented above. Regard, e.g. such phrases as “she had every (good) reason to believe that he was honest” and “he had more reason to believe that she was guilty than not.” Here reasons are called upon to *rationalize* a belief held by an agent: If an agent is claimed to have such *rationalizing reasons* for holding a belief, she can be made to come out rational or *reasonable* in holding it, at least in a minimal sense.¹⁷³

As indicated above, an agent can well have good *rationalizing* reasons for holding a belief without having any *basing* reasons for holding it at all: Suppose, e.g. that the formation of the belief was caused by an epileptic fit or the agent’s head being hit by a shovel. Now, surely in that case the agent did not base her belief on any evidence or reasons all. However, the belief can well be one she has perfectly good (rationalizing) reasons to hold.

Arguably, it is even possible that the agent has good rationalizing reasons to hold a certain belief, is aware of this, holds the belief; yet does not *base* the belief on these rationalizing reasons.¹⁷⁴ A typical case could be that of a man who recognizes ample evidence indicating that his wife is unfaithful: She stays out late and offers him implausible excuses for this, their sex-life has changed dramatically etc. Still, out of his deep-felt trust in her character, he simply cannot *believe* that she is cheating. However, this man has one severe cognitive fault: He is highly susceptible to gossip. Now, suppose that finally he adopts the belief that his wife is cheating on him, not because of his awareness of the good rationalizing reasons he actually has, but because he picks up some slander at a café, actually not referring to his wife at all, but to another woman with superficial similarities to her. In this situation he would have good rationalizing reasons to hold the belief that his wife is cheating on him, but terribly bad basing reasons.

Another crucial difference between rationalizing and basing reasons is the following. An agent cannot *base* her belief on certain basing reasons without at least at some point being aware of those reasons. In contrast, it seems perfectly possible that an agent can have good rationalizing reasons for holding a belief, even if she is totally unaware of these reasons. Consider, e.g. the unhappy husband from above, who may well be totally unaware of the ample evidence pointing to the unfaithfulness of his wife. Now, at the time he forms the belief that she is cheating on him based on the gossip, it still seems natural, at least to this author, to uphold the claim that this belief is one he has good rationalizing reasons to hold, even if he is basing it on bad reasons.

An important *terminological* similarity between the notions of basing and rationalizing reasons is worth noticing, though: Unlike in the case of explanatory reasons, the terms “reason,” “ground,” and “evidence” are freely interchangeable when talking

about reasons of these two kinds.¹⁷⁵ However, it takes some footwork to bring out the exact sense of the term “evidence” employed here.

First of all, what forming beliefs *on* evidence amounts to, stands in acute need of clarification. The term evidence has been employed in a wide number of senses in the epistemological literature.¹⁷⁶ I shall take it here that evidence is anything that may be taken epistemically to support a belief. This however leaves open a wide number of possibilities. For example P. Achinstein has distinguished three concepts of evidence, which applied to beliefs may be roughly paraphrased as follows¹⁷⁷: (1) Potential evidence: The occurrence of the piece of evidence is *objectively connected* to the truth-value of the belief, e.g. by way of regularity or association. (2) Veridical evidence: the piece of *evidence* is potential evidence for the truth of the belief, and the belief is actually true. (3) Subjective evidence: the piece of evidence is taken (believed) to be veridical evidence by an agent.

Arguably all of these senses are involved in our ordinary use of the term “evidence.”¹⁷⁸ Thus, as the above list is hardly exhaustive, prospects of fixing a sense for the term that satisfies all relevant intuitions are dim indeed. However, it would seem that in the present context, “believing *on* basing reasons (or grounds)” can only be synonymous to “believing *on* evidence” if by “evidence” is meant something like *subjective evidence* in the sense presented above. However, if the agent’s *awareness* of her subjective evidence is to *explain* why she holds the relevant belief, i.e. if that evidence is to act as her basing reason for holding it, it cannot very well be essential to the relevant notion of evidence that the agent is *already* aware of it. Thus Achinstein’s notion of subjective evidence will need a modification to be of use in the present context.

This modification, however, is no easy matter; Actually the relevant sense of evidence can hardly be constructed from the notions of objective truth-conducivity “objective connection to truth values” and subjective awareness at all. This is because in imaginable scenarios something may well be evidence for the truth of a proposition without as a matter of fact being truth-conducive (see the examples below). If this is correct, and the notion of truth-conduciveness must still be allowed a central importance relative to the notion of evidence, it seems it will have to be put into a proper intensional embedding. Once this observation is made, I think there is no reason to stop short of an *agent-relative* notion of evidence. Consider, e.g. the following recent formulation of the “Cartesian demon” argument due to Bruce Russell:

The grounds for saying that people in the demon and the non-demon worlds are, from the standpoint of [objective] justification, in exactly the same situation, is not ... (...) ... that if we did not say this we would be allowing luck to play a role in justification that it should not ... (...) ... Rather, it is the intuition that each has equally good evidence for what he believes – from which it follows, of course, that the goodness of evidence is independent of its tendency to produce true beliefs.¹⁷⁹

Invoking a Cartesian demon world to make this point may well be a severe metaphysical overkill: It is far from clear that such a world is possible in any sense. Further, if in such a world all beliefs apart from, e.g. a *cogito* belief are false, it is not even clear how epistemology could meaningfully be done in this world. However, Russell’s point can easily be given a more modest presentation, avoiding the worst problems

faced by the demon-world hypothesis concerning, e.g. the constituency of linguistic meaning: Suppose that the demon world is just as we take the actual world to be, only that the demon tricks inhabitants of the demon-world about a single class of beliefs (e.g. those concerning the position of tables) for a limited period of time, e.g. 10 years every century. Suppose further that the demon achieves this without giving anyone in the demon-world the slightest reason for thinking themselves tricked. Now it would certainly seem that the demon-world inhabitants have equally good reasons for their beliefs about the location of tables as has inhabitants of the actual world, even though in the demon-world those beliefs are all false.

However, once this conclusion is conceded, there appears to be no reason, why we should not also endorse the following line of reasoning: Suppose that the modest demon achieves to pull the same type as trick as above, only this time around it only tricks a single individual. Now, why shouldn't we say that this individual has good evidence supporting her beliefs about the location of tables, even though her relevant perceptions are all misleading? I believe this strongly supports making *agent-relative* the notion of reasons or evidence for beliefs employed in the notions of basing and rationalizing reasons. I will thus stay satisfied with the following analysis of the notion of good evidence or reasons (as in "basing reason" and "rationalizing reason") in the present context:

A matter of fact or proposition E constitutes good *evidence* for the truth of proposition p for an agent A *if, and only if*, the presence of E is correlated with the truth of p as seen *from A's epistemic perspective* (whether A is aware of this or not).

Despite the inherent vagueness of the notion of an epistemic *agent-perspective*, I hold this analysis clear enough for the present purposes. Intuitively, an agent's epistemic perspective comprises at least her doxastic states, her cognitive abilities, as well as her cognitive history. The main purpose of stating the analysis is negative: It has to be made clear that basing one's belief on good reasons *may* be a truth-conducive mechanism of belief-formation, but is not *essentially* so. Thus the "basing relation" desideratum may not replace the "truth-conducivity" desideratum (see Section 5.4), even if we keep bizarre, non-reason-driven methods of belief-formation out of the picture.

It is important to notice that, even though the judgment of goodness here must be made relative to the agent's perspective, and is thus not a simple judgment of objective truth-indicativeness, it is not essentially the agent's call. To employ the example of "the educated racist" (see Section 1.2) I may, e.g. well judge that, given her cognitive perspective including such things as her educational background and the cognitive opportunities offered to her, the pronouncements of her favourite demagogue are not sufficiently truth-indicative concerning racial matters as seen from her epistemic perspective, even if the educated racist may think they are. On the present conception, one's basing or rationalizing reasons for holding a belief are not good just because one thinks them so.

Having noticed the differences between explanatory, basing, and rationalizing reasons we may introduce the fourth and final type of reason relevant here: At least one version of *doxastic voluntarism* (see Section 8.1) is committed to the view that

beliefs may be formed voluntarily in roughly the same sense that one may voluntarily perform an action like raising one's arm. Now, given this commitment, just as one may raise one's arm *in order to* pass the salt, one may adopt a belief *in order to* bring about a certain state of affairs. For example if one is a WWII British spy in Germany and holds it an advantage to believe that one was actually born in Düsseldorf, one may adopt this belief for that *instrumental* reason, entirely independently of one's opinion concerning the truth of its content.

It should be remarked that we can still have instrumental reasons to hold beliefs, even if doxastic voluntarism fails. For example if a dissident in a totalitarian country I may well be said to have an instrumental reason for adopting the beliefs accepted by the governing party, because, unless I adopt those beliefs, I face the prospects of an unpleasant death. Without the possibility of some kind of voluntary influence on the content of my beliefs, though, my recognition of this instrumental reason cannot immediately help me achieve any ends whatsoever. Later I shall concede that we might have a limited ability to act on instrumental reasons for holding beliefs with certain contents (see Section 10.5), but that this possibility is irrelevant to standard ascriptions of epistemic blameworthiness.

With all four notions of a reason in place the following conceptual complications arise: Given the possibility of a *doxastic action* like forming a belief at will and given the plausible claim that situations like that of our unhappy husband before picking up the slander are indeed possible, the situation arises that an agent's *instrumental* reasons for adopting a belief may well be her *awareness* that she has good *rationalizing* reasons for believing that p, but does not believe that p. (i.e. she recognizes that despite her awareness of good reasons for believing that p, she fails to believe that p due to some counter-rational causal mechanism, and takes this as an *instrumental* reason to form the belief that p.)

Further the agent's *acting* on these instrumental reason now becomes the *explanatory* reason, why she believes that p. However, her awareness of these rationalizing reasons do not therefore constitute her *basing* reason for believing that p after she has performed the doxastic action of adopting it: It was exactly because she was aware of her inability to base a suitable belief on the evidence of which she was aware, that she had to perform a doxastic action in the first place.¹⁸⁰

With all these interrelated senses of the term "reason" in play in the context of epistemic deontology it is not surprising that confusions may arise. As shall later emerge, such confusion has indeed occurred, primarily in the discussion over doxastic voluntarism, where the differences between explanatory, basing, rationalizing, and instrumental reasons have arguably been obscured by some authors.¹⁸¹ I will now proceed to discuss the relevance of these notions to the issue of epistemic desirability.

We may begin by noticing that the notion of an instrumental reason is of little importance to the issue of epistemic desirability: If it is ever epistemically desirable to bring about a belief because one has good instrumental reasons to bring it about, those instrumental reasons can only be the fact that holding the belief is desirable along one or more of the *other* epistemic desiderata. The notions of

rationalizing and basing reasons, however, have a decisive role to play in the present context:

Alston's initial formulation of the "basing relation" desideratum really employs the notion of a *ground* in two different senses. It goes as follows:

In order that my belief that *p* be justified it is not enough that I merely have a ground of the appropriate sort. It is also necessary that my belief be based on that ground, that this be my reason (evidence ...) for believing that *p*.¹⁸²

Here the first occurrence of "ground" clearly denotes a *rationalizing* reason, the second occurrence a *basing* reason in my preferred terminology. As seen above, an agent may well have good rationalizing reasons for believing *p*, without having good basing reasons and vice versa. However, having *both* kinds of good reasons also seems desirable from an epistemic perspective. This suggests that the "basing relation" desideratum as presented above, should really be split up into two separate epistemic desiderata. Alston is aware of this, and suggests that it is epistemically desirable that:

A. The belief is based on adequate grounds (reasons, evidence).

As well as that

B.S [the agent] *has* adequate grounds (reasons, evidence) for the belief.¹⁸³

We then have the following pair of desiderata, replacing Alston's single "basing relation" desideratum from above:

(1)A: (adequate basing): It is epistemically desirable that a belief be *based* on good basing reasons (adequate evidence, adequate grounds).

(1)B: (reasonableness¹⁸⁴): It is epistemically desirable that an agent holding a belief has good rationalizing reasons for holding that belief.

As should be clear by now, the desideratum (1)A can only appropriately apply to beliefs *actually* based on some reason or other in the sense preferred here: Otherwise even a belief produced by a highly reliable non-reason-driven process would fail to qualify as epistemically desirable, since it is not based on good reasons in the relevant sense. This is clearly unacceptable to an approach to epistemic desirability in terms of the truth-goal. With this qualification I shall go with Alston's desiderata as formulated above.

One important issue, however, remains: It must be considered whether the desirability of fulfilling these desiderata may in fact be accounted for relative to the truth-goal, i.e. whether the above desiderata are really *epistemic* desiderata in the present sense.

There is a concern here that the understanding of (rationalizing and basing) reason goodness adopted above will make it impossible to uphold the epistemological value-monism opted for in Section 5.2: How, one may wonder, can the desirability of having good rationalizing reasons for one's beliefs or basing them on good reasons, be defended relative to the truth-goal, once it is conceded that such good reasons are not necessarily truth-conducive?

My answer to this challenge is simple: The connection between the goodness of basing and rationalizing reasons and the truth-goal is so obvious that it may easily be overlooked: good reasons of either sort are essentially good reasons of either sort to hold the relevant proposition *true*! Thus even without exploring the notion of rationalizing and basing *reason goodness* further, the conclusion can be made, that unless one assumes the value of pursuing the truth-goal, this notion does not make any sense: It might be that, as a matter of fact, an agent has good rationalizing reasons (evidence, grounds) for believing something or is basing that belief on good basing reasons (evidence, grounds) without that belief being true, and this may even be systemically so to a certain degree (constrained by metaphysical worries about the feasibility of radical demon worlds). However, if we did not believe these reasons to be actually truth-conducive as seen from the agent's perspective, we would hardly believe them to be good reasons in the first place.

This suggests that, when doing epistemology within our *own* perspective, we must suppose that the goodness of our reasons hinges on their truth-conducivity. Epistemology can hardly proceed on the assumption that we live in a radical demon-world, in which case it would be difficult to see how, short of dubious Cartesian meditative strategies, we could express our care for the truth-goal in the first place. In fact, I think, there is a common concern that we actually share epistemic perspectives to a wide degree: In evaluating doxastic states from an epistemic perspective we cannot proceed on the assumption that normal individuals are victims to Cartesian demons, even of the modest varieties faced above. Arguably, such concerns only apply to the severely insane.

On the contrary, we must normally presume that, both being normal agents, *my* good reasons for holding a belief could also under the relevant circumstances be *your* good reasons for holding this belief. Otherwise we immediately face the total end of rational conversation, which might well mean losing the purpose of doing epistemology of the type represented by the present study: It is not easy to see how we could fruitfully discuss a subject matter with someone under the spell of a full-blown Cartesian demon. Our assessments of reasons must then be under the following *naturalistic* constraint, lest we lose out on the purpose of our practice of discussing such assessments at all: Bruce Russell's conclusion from above that "the goodness of evidence is independent of its tendency to produce true beliefs," is a principle we can hardly ever afford to apply to particular cases. In other words: it is not a conclusion that should substantially worry us in germane assessments of epistemic blameworthiness.

5.4. TRUTH-CONDUCTIVENESS OR RELIABLE FORMATION

As we shall see in Chapter 6, Bruce Russell has opted for an account of epistemic blameworthiness solely based on a single epistemic desideratum: The *reasonableness desideratum* as construed above. He thus holds the reliability (actual truth-conduciveness) of the way in which an agent formed a belief entirely irrelevant to the question of its epistemic blameworthiness.

However, this stand arguably makes Russell unable to deal with even his own examples of blameworthy belief in a natural way: As I hope to demonstrate in this paragraph, there is no good reason why an ascription of epistemic blameworthiness cannot appeal to “externalist” notions such as reliable formation.

Consider again the infamous *Morgan* case from Section 1.3: Here the rapists formed their belief that Mrs. Morgan consented in a highly unreliable way, namely by trusting the testimony of her malicious husband. Suppose now that they were themselves at fault for their credulity, e.g. due to having ingested some credulity-inducing drug while perfectly foreseeing that this action might result in beliefs being formed in a very unreliable way. Now, I will submit, the rapists may plausibly still be blamed for their “honest mistake” in raping Mrs. Morgan, *even* if the drug should also alter their mental make-up in a way, so that their belief that Mrs. Morgan consented was a belief that each of the rapist had good rationalizing and good basing reasons to hold after ingesting the drug and listening to Mr. Morgan’s testimony. Intuitively, for each of the rapists, the drug-related way that his belief in Mrs. Morgan’s consent was formed was strongly representative of his disregard for the truth-goal and therefore one he should be condemned for, no matter what other effects the drug might have had. A conception of epistemic blameworthiness not taking into account the reliability of the way in which the rapist’s belief was formed, would not be able to account for his striking epistemic blameworthiness in the last-mentioned scenario. Consequently, a satisfactory conception of epistemic blameworthiness must allow that a belief is epistemically blameworthy, if it is formed in an unreliable way, and the believer has no adequate excuse for the unreliability of its formation.

This point runs counter to a supposition recently made by Josefa Toribio that “culpable error” (i.e. false and blameworthy belief) requires that the relevant belief be subject to “critical pop-out,” i.e. is a belief that the believer automatically recognises as one requiring her to step back and exercise critical appraisal of “tuneability by reasons.”¹⁸⁵ In the above example, surely the rapist did not recognise his belief that Mrs. Morgan consented as one requiring him to critically examine it.

However, I am unable to see why one should hold “mindfulness”¹⁸⁶ in this sense requisite for the appropriateness of deontic evaluations. As readily admitted by Toribio this restriction, e.g. excludes almost all perceptual beliefs from deontic evaluation,¹⁸⁷ which does not seem plausible (I shall offer quite mundane examples of blameworthy perceptual beliefs in Chapter 9). Moreover, empirical research seems to indicate that “critical pop-out” very often does not improve an agent’s beliefs from an epistemic perspective, as we tend to let ourselves influence to an unfortunate degree by the random evidence most readily accessibly in our memory¹⁸⁸: Thus, agents allowing themselves to be consciously “tuned by reasons” often end up with beliefs based on reasons not truth-indicative from any relevant perspective. It is hard to see why we should only allow ourselves to submit to deontic evaluations beliefs already disadvantaged in this fashion.

Still, the exact way in which the notion of reliability must figure in the context of deontic epistemic normativity is far from easy to spell out in a detailed fashion. Epistemologists of a “reliabilist” inclination have differed strongly over the role,

which the notion of reliability should play in epistemic evaluations of beliefs as well as over how to understand the notion itself. For example there is a divide between *reliable indicator* – and *reliable process* theories of epistemic justification, where, roughly speaking, the former stress the actual truth-indicativeness of an agent’s basing reasons for a belief and the latter stress the general reliability of the way in which the belief was produced, reason-based or not.¹⁸⁹

Reliable indicator theories shall not concern me further here. I take it to be pretty clear that an agent’s basing one of her beliefs on reasons that are not only truth-indicative from her own epistemic perspective, but also actually so, is only valuable from the perspective of the truth-goal, if her basing of her belief on these reasons is in fact a *reliable way* of forming such a belief. In other words, it is not easy to see what epistemic good the actual truth-indicativeness of her basing reasons does, if her belief remains unreliably formed despite being based on those reasons. The value of indicator reliability must thus depend on its role in reliable doxastic processes. Consequently, I take it that reliable processes are what we should be concerned about in the present context.

Specifying what a reliable process is, however, is no easy matter. For example it is worth noticing that the notion of a reliable process relevant to the present deontic concerns cuts across the categories of a “process” and a “method” devised by the prominent reliabilist Alvin Goldman:

“Processes” are basic psychological processes, roughly, wired-in features of our native cognitive architecture. “Methods” are learnable algorithms, heuristics, or procedures for forming beliefs, such as procedures that appeal to instrument readings, or statistical analysis. All beliefs are formed partly by processes. We cannot do anything in the cognitive realm without using basic psychological operations. Learned methods, by contrast, are not universally required.¹⁹⁰

Goldman stresses that “methods,” unlike “processes” must be “represented in the cognizer’s head,”¹⁹¹

Now, it seems pretty clear that an agent cannot be blamed for the output of her unreliable psychological “processes” in Goldman’s sense, as these are simple hard-wired mechanisms, whose operations are entirely beyond the control of the individual (see also Section 7.1). Thus, it would be tempting to conclude that only “methods” in Goldman’s sense could be significant from a deontic perspective. However, we may easily set up an example of a relevant unreliable process that does not immediately qualify as a Goldmanian *method*: Above we imagined that the *Morgan* rapist’s beliefs were produced by a process of listening to Mr. Morgan’s testimony while under the influence of a certain credulity-inducing drug. Now, we may assume, one of the *Morgan* rapists may also have taken a memory-erasing drug wiping out any internal representation concerning the first drug and Mr. Morgan (he may believe that Mr. Morgan’s voice belongs to a radio speaker). This, intuitively, does not seem to seriously influence the matter of his epistemic blameworthiness. However, the unreliable process by which his beliefs were produced now fails to qualify as a Goldmanian “method”: It is not represented in the rapist’s head. Neither, of course, does it qualify as a Goldmanian “process”: the cognitive mechanism set up by the drug is not a “native psychological process” in any natural sense.

Goldmanian “methods” thus hardly exhaust the kind of belief-producing/sustaining processes that may give rise to blameworthy beliefs. As shall emerge by the end of this study, the kind of processes or mechanisms relevant here comprise exactly those, whose unreliability may be the result of a violation of an *intellectual obligation* by the relevant cognizer. Intuitively, native hard-wired psychological mechanisms (Goldmanian “processes”) fail to meet this criterion, since they may not be influenced no matter what the agent does. However, the relevant mechanisms are not therefore restricted to internalised “methods.” In terms of the *Morgan* example, the rapists may well have violated an intellectual obligation (as well as a few other obligations) by ingesting the credulity-inducing drug. Still, of course, typical violations of intellectual obligations will concern Goldmanian methods in a quite straightforward sense: For example the failure to subject oneself to certain educational opportunities, which might have improved the reliability of one’s belief-forming habits.

Even with this *caveat* in place, we are still far from a satisfactory account of what is meant by a “reliable process” in the current context. The notion is exceedingly difficult to strap down. As pointed out by Goldman above, hard-wired psychological mechanisms always play a role in the causal history of a belief-holding (whereas Goldmanian “methods” need not be involved). Thus a satisfactory description of the actual process leading to a belief must necessarily refer to such hard-wired psychological mechanisms. The procedure for a deontic evaluation of a belief based on its causal history must then be:

1. Adequately describe the process responsible for the belief-holding including “hard-wired” psychological mechanisms, Goldmanian “methods” and external causal influences.
2. Assess whether the process so described is unreliable.
3. If, so, assess whether the believing agent is *at fault* for the operation of this process in this particular instance (through her influence on those features of the process, which are not hard-wired).

However, these steps are easier expressed than undertaken. First it is immensely difficult to specify the criteria for an adequate description of a particular process. One particularly pernicious problem is that one may opt for a definite description of the process so narrowly indexed that it fits only the particular process responsible for the particular belief under evaluation. Such a description would, e.g. have the form: “The particular process responsible for the fact that agent A holds the belief that p at time t in the actual world.” Since, necessarily, this “single-output”¹⁹² process can only produce the one belief under evaluation, if this belief happens to be true, the process will qualify as reliable even if the belief was produced by the operation of a type of mechanism, which most often produces false beliefs (e.g. a trust in clairvoyant powers). Goldman responds to this problem by pointing out that single-output “methods” are not available to the cognizer in the right sense; they are not the kind of methods the cognizer “could naturally be led to employ.”¹⁹³ Whatever its merits, however, this response does not seem to carry over, once we are no longer narrowly concerned with internalised “methods.”

Further, the “single-output” problem is only a particular version of a general problem concerning the adequacy of process descriptions. Assume, e.g. that a man has a habit of visually identifying the colours of cars at all times of the day. In broad daylight he almost never forms false beliefs concerning car colours, whereas it often happens at night. His trust in his own abilities, however, oddly remains constant under all lighting conditions. Now, assume that in broad daylight this man correctly identifies the colour of an Alfa Romeo as pearl blue. Should we say that he arrived at the belief that the car is pearl blue by the operation of his colour vision (which is not generally reliable)? Or by the operation of his daylight colour vision (which is very reliable)? What if the car was correctly identified at dusk on a dark day? Or on a particularly bright night?

The description problem briefly presented above is a problem that any belief-evaluation in terms of reliable formation must face. It is not particular to the present context. I shall then only trust, that my examples of unreliable processes will be of such a character that the description given intuitively qualifies as adequate.

Pernicious problems also adhere to the term “reliability” itself. On a first blush the notion appears crudely simple: A reliable process is simply a process that yields a sufficiently high ratio of true beliefs relative to its total belief-output.¹⁹⁴ However, there is a strong temptation to *modalize* this notion, i.e. consider in an evaluation of a process not only its truth-ratio in the actual world, but also in some set of possible worlds. This temptation arises, because arguably a process might have a high truth ratio in the actual world due to sheer *accidence*, and should not therefore be evaluated favourably from an epistemic perspective. For example this line of reasoning has it, some self-proclaimed psychic medium might actually have an astonishing success in predicting future events, even if a process of belief-formation involving trust in clairvoyant powers, being mere guess-work, should never qualify as reliable. However, success in *all* possible worlds is clearly too extreme a demand for reliability: Even an intuitively very reliable process, such as visual inspection of the position of middle-sized objects in the visual field, may arguably produce almost only false beliefs in an infinite amount of bizarre possible worlds.

Various attempts at characterizing the proper domain of possible worlds relevant to epistemic evaluations have been made. For example at some point Alvin Goldman suggested that only the performance of a process in so-called “normal worlds” was relevant from an epistemic perspective. He construed a normal world as a world “consistent with our general beliefs about the actual world.”¹⁹⁵ However, he soon recognized the difficulties of judging whether a particular possible world qualifies as “normal” on this criterion: Who are “we” in the formulation above? Which beliefs qualify as “general”? Which general beliefs (out of an unmanageable multitude) should be taken into account?¹⁹⁶ Not surprisingly, already in 1993 Goldman dismantled any talk of “normalcy” and opted for a “normative scientific epistemology” continuous with “our epistemic folkways,” which, so he argued, takes only into account the performance of a process in the world *believed* to be the actual world.¹⁹⁷

Regrettably I am not able to accurately specify the set of possible worlds that should be taken into account in particular evaluations of process reliability. Again, I shall

only trust that, in the cases I shall discuss, it will be immediately clear whether the processes described qualifies as reliable or not. We may, however, still answer in outline the question, *how* reliable a process must be in order to qualify as reliable in the present sense:

The desideratum of belief-formation by a truth-conducive or reliable process cannot very well desire beliefs to be formed by processes that yield true beliefs with absolute certainty: Given plausible fallibilist assumptions, such processes are simply not available to human beings, and if such a process were in fact what is desired by the desideratum, Blame_{epist} would yield that no belief is in fact blameworthy along this dimension. This is so because agents might then always pass the simple, yet seemingly appropriate excuse that they are *human*, and thus unable to attain absolutely reliable processes. In the present context I can only submit that the level of process reliability relevant to deontic concerns is a somewhat fuzzy matter. However, I am not, e.g. bound by any analysis of justification in terms of formation by a reliable process. Thus no “naturalistic” constraint requiring a large portion of beliefs to be in fact justified, hence reliably formed, applies. The question “How reliable must a belief-forming process be for maximal epistemic desirability?” may then receive the vague but pompous answer: As reliable as is possible, given the particular circumstances!

I trust my reader to see that the vagueness of this answer is not immediately problematic to my over-all concerns. What is relevant here is whether a certain belief qualifies as reliably formed to *some* degree, and if not, whether the believer is adequately excused for its non-reliable formation. It does not matter, e.g. whether the belief is formed by a *maximally* reliable process.

Above I denied the relevance of an epistemic value-dualism hypostasing *knowledge* as an epistemic goal independent of the truth-goal. However, I have not yet considered a second kind of value-dualism hostile to the present approach to epistemic desirability: It might be argued that the desirability of reliable formation for beliefs cannot, in a way relevant to deontic epistemic normativity, be derived from the desideratum that they be simply true. If, namely, a desideratum must be somehow *socially significant* in order to be relevant to the issue of epistemic blameworthiness, the desideratum of reliable formation faces the following challenge: That an agent has true beliefs matters uncontroversially to her standing as a good informer, which is socially significant. However, once a belief is true, it cannot matter much in this respect that it is also reliably formed: That a reliable informant with true beliefs can also be relied upon “on other occasions” is strictly speaking irrelevant to the desirability of her present beliefs.¹⁹⁸

It would seem then that the desideratum of reliable or truth-conducive formation is *independent* of the truth-goal and thus that a value-monism tying the notion of epistemic desirability closely to the truth-goal cannot be retained.

Luckily, I think there is way out of this predicament: It is undoubtedly correct that for a presently held belief the desideratum of truth-conducive formation cannot play a socially significant role once the belief is true. When taken in isolation, it is even hard to see why, if true, we should value its reliable formation at all. However, even if blameworthiness applies to presently held beliefs, the desideratum of reliable

formation also plays a distinctly *diachronic* role in the concept of epistemic blameworthiness: As shall be argued below, if a belief is blameworthy, some violation of an intellectual obligation must have occurred in the past (see Section 12.1). Now, an obvious way to guide one's future beliefs towards satisfying the desideratum of truth is in fact to opt for a satisfaction of the desideratum of reliable formation. In a straightforward sense then, an agent who has inexcusably formed a belief in an unreliable way may be said to be epistemically blameworthy for holding this belief simply because it is representative of her past inappropriate service to the truth-goal. In a deontic context the truth-goal remains basic after all. This is not to say, however, that a belief may be epistemically undesirable in the sense relevant to epistemic deontology simply by being false. A discussion of this point will follow in Section 5.5.

5.5. THE NOTION OF EPISTEMIC UNDESIRABILITY

Above we have singled out a number of epistemic desiderata or dimensions of epistemic desirability for beliefs. Special attention has been given to the following three desiderata:

1. Truth-conducivity: It is epistemically desirable that a belief is formed and (causally) sustained by a reliable process.
2. Adequate basing: It is epistemically desirable that, insofar as a belief is based on reasons, it is based on *good* reasons (evidence, grounds).
3. Reasonableness: It is epistemically desirable that an agent holding a belief has *good* rationalizing reasons (grounds, evidence) for holding that belief.

I have argued that the desirability of a belief's fulfilling any of these desiderata may be accounted for relative to the value of agent's pursuit of the epistemic goal of holding only true beliefs (the truth-goal). Thus, I have so far resisted the idea that the truth-goal need to be supplemented by other epistemic goals in order to account for the notion of epistemic *undesirability* needed for my analysis of epistemic blameworthiness. I shall now add some more substance to this claim and submit that a belief is simply epistemically *undesirable* in the sense that matters here, if, and only if, it is *not* epistemically desirable in the above sense. The notion of epistemic undesirability thus, via the notion of epistemic desirability, hinges on the truth-goal.

The equivocation of epistemic undesirability with non-desirability in the present context may be defended in the following way: The notion of epistemic blameworthiness we were after should be able to account for ascriptions of moral blameworthiness in "honest mistake" cases like those encountered in Section 1.3. Here it seems plausible that we regard as *undesirable* a belief that is not held for good reasons, not reliably formed or not based on adequate evidence. If an agent holds such beliefs, and they happen to be false, she better have an excuse for it, lest we condemn her actions based on such beliefs. Our actions and the beliefs that prompt them are simply too important matters to allow that beliefs that are merely "whimsical" (neither based on – nor held for – good or bad reasons) or neither reliably nor unreliably formed (if this is possible) should pass without an adequate excuse.

I shall thus say that the following properties of a belief are *undesirable* from an epistemic perspective:

1. Lack of formation by a truth-conducive process: It is epistemically undesirable that a belief is not formed and (causally) sustained by a reliable process.
2. Inadequate basing: It is epistemically undesirable that a belief based on reasons is not based on *good* basing reasons (adequate evidence, adequate grounds).
3. Unreasonableness: It is epistemically undesirable that an agent holding a belief does not have good rationalizing reasons (adequate grounds, adequate evidence) for holding that belief.

I acknowledge that a belief may in a broader sense be epistemically undesirable in virtue of violating *any* epistemic desideratum. However, I shall limit myself to considerations of undesirable beliefs satisfying either of the three epistemic *indesiderata* listed above: I shall thus submit the following analysis as adequate for the present purposes:

Undesire_{epist}: An agent's holding of the belief that *p* is epistemically undesirable if, and only if, it is not formed and sustained by a reliable process, not based on adequate reasons, or unreasonable.¹⁹⁹

At this stage I need to address a fundamental objection, already alluded to above: Some may wonder why, since after all the epistemic undesirability of the above properties is accounted for relative to the truth-goal, I have not simply taken *falsity* (non-truth, if one eschews bivalence) as my basic notion of epistemic undesirability and offered the following, seemingly simpler, basic analysis of epistemic blameworthiness:

Blame_{epist, falsehood}: An agent is epistemically blameworthy for her holding the belief that *p* *if, and only if,*

1. She believes that *p*.
2. Her belief that *p* is *false* (untrue).
3. She is not appropriately excusably for the falsity (untruth) of her belief that *p*.

On such an analysis the notion of epistemic desirability evolved above would be tied to the non-excusability clause (2): If a belief is false, even though it satisfies all of the epistemic desiderata listed above (none of which infallibly yields true beliefs if satisfied), a proponent of Blame_{epist, falsehood} would need to say that the agent is then appropriately excused for the *falsehood* of this belief. This would be in order to escape the highly implausible conclusion that the agent would be blameworthy for holding a false belief that is otherwise epistemically desirable in any way.

I shall offer two important reasons for preferring Blame_{epist} over Blame_{epist, falsehood}: First, the above stand that a proponent of Blame_{epist, falsehood} is forced to make on pains of immediate implausibility, simply strikes me as stretching the notion of excusability in a very unnatural way: If a belief satisfies the epistemic desiderata listed above, yet still accidentally happens to be *false*, the agent holding this belief surely has nothing to *excuse*: From the perspective of the truth-goal there is nothing wrong with the *way* she holds this belief, its falsehood is due to circumstances she could not be required to

control in any relevant sense. Now it simply seems highly strange to excuse someone for something she could never be required to control²⁰⁰: This would, e.g. open up the possibility of “excusing” any normal agent for the bad weather condition (since this is something undesirable she could not possibly control). I take these considerations to sufficiently establish that a more “economic” conception of the notion of epistemic undesirability simply puts too heavy a strain on the notion of excusability.

Second, and perhaps even more important, the notion of epistemic undesirability preferred here allows for the possibility that a belief may be epistemically undesirable, hence potentially blameworthy, although not false. I take this to fit with our intuitions. Consider, e.g. the *Morgan*-style (see Section 1.3) case of a man, Jones, who believes that a woman consents to having sex with him. This happens to be true. In fact this woman is a desperate prostitute looking for any customer she can find. However, Jones’s belief is based entirely on the observation that this woman is wearing a very short dress; in fact he would have believed that she consented even if she had protested wildly and given him any good reason to believe that she did not consent. Now, Jones’s belief that this woman consented to have sex with him, even if true, is hardly any better from a deontic perspective than the *Morgan* rapist’s false belief that Mrs. Morgan consented to have sex with him. Thus, unless Jones has an appropriate excuse for basing his belief on the skimpy evidence, it seems highly appropriate to blame him for holding it under the above circumstances.

Some may feel at this stage that I have erred in the opposite direction by preferring a notion of epistemic undesirability too lean for the present purposes. In particular it might be argued that I have seriously amputated my notion of epistemic undesirability by failing to take into account the supposedly central epistemic goals of *understanding* and *wisdom*. For example Linda Zagzebski has argued:

The important point is that understanding ought to an important concept for us as well, it has clearly been neglected, and this neglect cannot be remedied if epistemology persists in making the locus of evaluation individual propositions or states of believing single propositions, as is the case with justification. Understanding is not a state directed towards a single propositional content at all.²⁰¹

Now, Zagzebski does not do very much to explicate what exactly *understanding* in her sense of the term amounts to, apart from some scattered remarks that, e.g. understanding essentially concerns entire “chunks of reality.”²⁰²

On the epistemic goal of wisdom Zagzebski has the following to offer:

The nature of wisdom is elusive, but it is clear that whatever it is, wisdom is an epistemic value qualitatively different from the piling up of beliefs that have the property of justification, warrant or certainty. Wisdom is neither a matter of the properties of propositional beliefs, nor is it a matter of the relations among such beliefs; it is a matter of grasping the whole of reality.²⁰³

I shall agree with Zagzebski that the import of phrases like the last of the above passage is indeed elusive: Here we are no longer satisfied with grasping just “chunks” of reality, but opt for the full package! Even if to me it is unclear, what “grasping the whole of reality” amounts to, I shall grant here that this is indeed a valuable goal from a suitably wide epistemic perspective.

Above (Section 5.1) I shrug off the relevance of the Cartesian goal of knowledge conceived as a solitary quest for certainty, as this goal concerns only the isolated individual subject. However this move is hardly warranted in the cases of wisdom and understanding: Wisdom and understanding are plausibly construed as “traits that enable individuals to live well in communities.”²⁰⁴ It would seem, then, that a notion of epistemic desirability as that evolved above, which evaluates beliefs in a piece-meal fashion, is simply too narrow.

However, even if wisdom and understanding are aims too lofty to directly concern singular beliefs, I take it that an agent’s neglect of the more humble truth-goal does not exactly promote her pursuit of the goals of understanding and wisdom either: How can she aim to grasp the whole or even mere chunks of reality, if she does not aim to hold true beliefs about it? It would seem, then, that on a virtue-driven approach to epistemic normativity like Zagzebski’s, the truth-goal must still be recognised as central epistemic goal after all. Her above statement that wisdom is not a matter of the properties of beliefs or their relations, is hardly accurate: If an agent aims to “grasp the whole of reality” she must surely care about the truth-value of her beliefs, even if this should perhaps be the least of her lofty concerns.

If this is correct, the present account of epistemic undesirability is not seriously endangered by the recognition of understanding and wisdom as over-riding epistemic goals: Violating the three epistemic desiderata from above is then also undesirable from this broadened perspective.

However, a proponent of the epistemic goals of understanding and wisdom may still object to my preferred conception of epistemic undesirability by arguing that really the appropriate objects of epistemic evaluation should not be an individual’s singular holdings of beliefs at all. For example Linda Zagzebski has submitted that it is simply “a mistake to evaluate beliefs singly.”²⁰⁵

Here I am forced to take a fundamental stand: Throughout I shall offer examples, mostly taken from the recent literature, of agents epistemically blameworthy for holding an undesirable belief. To me and to other authors, these examples have seemed intuitively compelling. I thus confess myself unable to understand the objection that accounting for a notion of epistemic blameworthiness applying to singular beliefs is an illegitimate project. This is not to deny that an alternative notion of epistemic blameworthiness may perhaps be developed, which does not take singular holdings of beliefs as its object. Such a notion could, e.g. be imagined to turn on the epistemic goals of understanding and wisdom in a substantial way; goals intuitively relevant from a suitably broadened epistemic perspective. However, as long as the nature of these goals is not stated more accurately than in, e.g. Zagzebski’s above-quoted passages, I find it hard to see how such a notion could be developed in a fruitful manner.

BRUCE RUSSELL'S BASIC ANALYSIS OF THE NOTION OF EPISTEMIC BLAMEWORTHINESS

Abstract. In this chapter I confront an alternative basic analysis of the notion of epistemic blameworthiness due to Bruce Russell. Russell's analysis takes account only of one epistemic *indesiratum*: the undesiratum of unreasonableness. Further, Russell holds the unreasonableness of higher-order beliefs relevant to the epistemic blameworthiness of first-order beliefs. Most importantly, he argues that an agent is epistemically blameworthy for holding a belief, even if she has a good reason to hold this belief, if only she also has a good reason to believe that she lacks a good reason to hold this belief. I argue that this claim is ill guided. More generally I argue that there is no imminent reason to hold second-order beliefs relevant to issues of epistemic blameworthiness at all. In particular, the cases of epistemic blameworthiness lined up by Russell are not very useful in this respect.

6.1. A CRITIQUE OF RUSSELL'S ANALYSIS

I shall now be able to confront a competing basic analysis of the notion of epistemic blameworthiness, which has recently been advanced by Bruce Russell. As shall emerge, this analysis shares significant features with Blame_{epist}, but nevertheless deviates from it in a number of important respects. Russell's most explicit statement of his analysis can be found in the following passage:

I propose that we adopt [Richard] Feldman's account of an objectively justified belief, namely, a belief that a person has good reason to believe, and let a subjectively justified belief be one that the person is epistemically blameless in holding. She will be blameless in holding it if [and only if] the belief is objectively justified for her and she has no adequate reason to think otherwise, or . . . the belief is not objectively justified for her but she has an excuse [which is legitimate] for holding it . . .²⁰⁶

The last two brackets supplied above drastically alter the import of Russell's text. However, I am convinced that I am true to Russell's intentions here: Consider the following formulation of the corresponding moral principle presented by Russell not long before the above quote:

She [the person] will be blameless *just in case* she has *no adequate reason to believe* that what she did was objectively wrong or has a *legitimate* excuse for doing it.²⁰⁷

Russell gives us no reason why he should be content with an "if" rather than a "just in case" or "if, and only if" clause in his epistemic version of the principle. Rather it would seem odd that he should offer an analysis of "subjective justification" and not hold that an agent is subjectively unjustified if these conditions are not met. Concerning the last bracket added: I suppose that a *legitimacy* requirement for excuses is implicit in Russell's analyses. I will thus proceed on the assumption that I have understood his view correctly by bringing it out.

On the reading adopted, the following analysis of epistemic blameworthiness follows²⁰⁸ from the first passage quoted:

$\text{Blame}_{\text{epist, Russell}}$: An agent is epistemically blameworthy (i.e. not blameless) for holding the belief that p *if, and only if,*

1. She believes that p .
2. She does not have good “adequate” reasons for holding the belief that p , or she has good reasons for holding the belief that she does not have good reasons for holding the belief that p and
3. She has good reasons for holding the belief that p , or she has no legitimate excuse for holding the belief that p .

$\text{Blame}_{\text{epist, Russell}}$ deviates from $\text{Blame}_{\text{epist}}$ on a number of counts. Most conspicuously, Russell’s analysis is built on a single epistemic indesideratum, the unreasonableness indesideratum as presented in Section 5.5 above. I shall argue that this presents the analysis with some serious difficulties in dealing with compelling examples of epistemic blameworthiness.

However, I shall begin by engaging the analysis on its own terms: Unlike my preferred analysis, $\text{Blame}_{\text{epist, Russell}}$ mentions second-order (rationalizing) reasons. To appreciate whether Russell has made a case for taking account of such second-order reasons, the interesting cases to consider are those in which an agent’s blameworthiness or lack thereof allegedly hinges on her second-order reasons.

Russell offers two examples of epistemically blameworthy agents, supposedly involving second-order beliefs in an essential way. The first is the following:

On the basis of his professor’s arguments, a student might have good reason to believe that he has good reason to believe, say, that God does not exist and yet still in fact believe that he does not have such reasons. Suppose as the result of the emotional pressure of his atheist friends, the student ends up disbelieving in God. Then he will be epistemically blameworthy for believing contrary to his epistemic conscience . . .²⁰⁹

Remarkably, Russell does not make it clear, whether the student actually has good reasons for believing that God does not exist. For this reason, it is not easy to see, how this example adds support to Russell’s analysis or even fits into it. However, if there is indeed a case of epistemic blameworthiness here, it is hard to see how this blameworthiness turns on the student’s second-order reasons (his good reason to believe that he has good reasons to believe that God does not exist). Rather, if anything seems immediately relevant, it is the fact that the student gave in to the emotional pressure of his friends, rather than to the rational arguments of his professor. If so, we may here have a violation of the adequate basing desideratum of Section 5.3: Even if aware of good reasons (offered by the professor) for believing that God does not exist, the student’s belief is best explained by his awareness of terribly bad reasons, namely those offered by his atheist peers. We may also have a violation of the truth-conducive mechanism desideratum from Section 5.4: Emotional pressure caused by college students surely is no good candidate for a truth-conducive belief-producing mechanism (perhaps in contrast to the persuasive force exerted by certain philosophy professors) and on $\text{Blame}_{\text{epist}}$ beliefs caused by such a mechanism must be blameworthy unless the agent is appropriately excused for falling victim to it.

Since $\text{Blame}_{\text{epist, Russell}}$ takes account only of the unreasonability indesideratum it is unable to account for the example in the above fashion. Russell’s first example

thus hardly presents a compelling case for the relevance of an agent's second-order reasons to her epistemic blameworthiness. However, the main charge is still to come. Russell presents the following line of reasoning:

Consider Keith Lehrer's example of Mr. Raco, who originally, and out of prejudice, believes that only members of a certain race contract a certain disease but who then learns in medical school that this is true. Suppose that at some point Mr. Raco has the medical evidence but does not hold his belief *on the basis of* that evidence. . . (. . .) . . . Here we might say that Mr. Raco acts in accordance with his epistemic duty but not from it and so is like Kant's grocer who gives the correct change because doing so is good for business. However, earlier I said that someone who has sufficient evidence to believe he will win some coveted prize but bases his belief on wishful thinking is epistemically blameworthy in what he believes. . . (. . .) . . . I said the prize-winner is blameworthy in what he believes but Mr. Raco is not. What is the difference between the two? . . . (. . .) . . . I think the difference is that the person who bases his belief on wishful thinking must think that he does *not* have sufficient evidence to support his belief and so believes against reason. However, after Mr. Raco attends medical school, he does not believe he lacks sufficient evidence and so does not believe against reason. The first person goes against his epistemic conscience but Mr. Raco does not.²¹⁰

I completely endorse Russell's comparison of Mr. Raco with Kant's scheming grocer. The only problem is that Kant actually thought the grocer morally blameworthy or bad-willed, since his actions did not have the right *source*: The grocer did not act out of duty or respect for the Moral Law. Now, it is clear that the above example only supports the relevance of second-order reasons (or, as Russell suggestively puts it, "epistemic conscience") to evaluations of epistemic blameworthiness if there is in fact an interesting difference between Mr. Raco and the wishful thinker. However, it is hard to see that any such difference obtains: Just as Kant considered the actions of the scheming grocer morally undesirable due to their tainted source, Mr. Raco's prejudiced belief, even if true and reasonable relative to an impressive hierarchy of good reasons, still violates the epistemic desideratum of adequate basing: Surely it must be epistemically undesirable to base a belief on a severely falsity-conducive reason like racial prejudice, no matter the luck one has in its turning out true, reasonable or even "conscientious" on a later stage of life. Thus unless Mr. Raco is adequately excused for basing his belief on such bad reasons, he must be epistemically blameworthy for holding it. Exactly the same point applies to the wishful thinker.

Arguably, Russell, in trying to make sense of such cases of epistemic blameworthiness as that of the wishful thinker or the atheist student, is only forced up the hierarchy of good reasons due to his failure to take account of first-order epistemic indesiderata other than unreasonableness. From the perspective of Blame_{epist} he simply misses out on the more natural analysis in these cases, and in the case of Mr. Raco simply fails to come up with the right result. Further, much of the appeal of "Mr. Raco" as an example for Russell's purposes arguably lies in the fact that the case is highly psychologically implausible. Definitely you would expect a real-life Mr. Raco to have shifted²¹¹ his basis for the relevant belief at some point, such that that belief was no longer based on bad reasons, i.e. no longer epistemically undesirable at all. He would then fail to be epistemically blameworthy. I strongly suspect that this intuition accounts for some of the persuasive force the example may in fact command. Still,

as presented by Russell the example makes no compelling case for taking account of second-order reasons in evaluations of epistemic blameworthiness.

This goes to show two things: First, Russell has offered no reason to include in an analysis of the notion of epistemic blameworthiness any mention of second-order reasons. Second, my above consideration of Russell's examples has added further support to the claim that a satisfactory conception of epistemic blameworthiness cannot ignore the epistemic *indesiderata* of inadequate basing and non-reliable formation.

One issue between Russell's analysis and $\text{Blame}_{\text{epist}}$ still needs to be considered: Russell has it that legitimate excuses must excuse the agent's *holding* of a belief, whereas $\text{Blame}_{\text{epist}}$ has it that such excuses must excuse the *epistemic undesirability* of that belief. My reasons for formulating the inexcusability clause thus will be clearer once the complex issue of doxastic control has been considered further below. Suffice it to say here that cases may exist, where, in a suitably out-lined situation, one agent forms a desirable belief and another agent forms an *undesirable* belief with exactly the same content, although neither agent could have *prevented* forming the belief under the relevant circumstances (see Section 10.4). Thus when possibly excusing the latter agent, it would seem odd to excuse her holding of the relevant belief, rather she should be excused for her belief's not being desirable like that of the former agent.

In conclusion, I have offered four reasons to prefer my analysis of epistemic blameworthiness over $\text{Blame}_{\text{epist, Russell}}$:

1. Russell presents no compelling reasons to take account of second-order reasons in an analysis of the notion of epistemic blameworthiness.
2. There are examples of epistemic blameworthiness (like the atheist student and the wishful thinker) that $\text{Blame}_{\text{epist, Russell}}$ cannot make sense of in a natural way.
3. Some epistemically blameworthy agents (like Mr. Raco) come out blameless on $\text{Blame}_{\text{epist, Russell}}$.
4. $\text{Blame}_{\text{epist, Russell}}$ has an inadequate inexcusability clause.

Thus Russell has hardly offered a plausible statement of the necessary conditions for epistemic blameworthiness. It remains now to consider, whether he has at least stated sufficient, albeit not *necessary* conditions for epistemic blameworthiness, i.e. whether his claims go through on the more modest, literal reading of the crucial passage from Russell (2001). If so, at least the reverse implication of $\text{Blame}_{\text{epist, Russell}}$ holds.

However, I believe that not even this more modest claim can be upheld, as the following demonstrates: First, on Russell's analysis second-order reasons only come into play in the following type of case: The agent must have good reasons for holding a belief, yet also have good reasons for believing that she does not have such good reasons.²¹² Now actually stating such a case is not an easy matter. Further, if no such case can be stated, second-order reasons immediately lose all relevance to the notion of epistemic blameworthiness, since there will then be no case at all where an agent is blameworthy due to the "wrong" kind of second-order beliefs. Now even stating such a case is hardly enough for Russell. He must in fact present an example satisfying the following even more exotic requirements: The agent must

1. Hold the belief that p.
2. Have good reasons for holding the belief that p.
3. Have good reasons for believing that she does not have good reasons for holding the belief that p.
4. Be blameworthy for holding the belief that p due to her having these second-order reasons.

Perhaps what Russell should offer here is an example of the following type: Suppose that another student of the philosophy professor from Russell's first example firmly believes that God exists. Suppose further that this student actually *has* good reasons for believing that God exists (she has had a genuine and impressive revelation). However, after attending a lecture by the sharp-witted philosophy professor, the student also acquires good reasons for believing that she does not have good reasons for believing that God exists: Some of the professor's arguments gave good reasons for serious doubts concerning the existence of genuine revelations.

Should we hold this student epistemically blameworthy without further ado for staying with her theist belief? Russell supposedly would. Now, at least to me, this is implausible. First, we might wish to take into account whether the student is perhaps *excused* for sticking to her theist belief, insofar as we take the circumstances to support her epistemic blameworthiness otherwise. Relevant possible excuses may involve the presence of an unmerited emotional pressure from religious peers, etc. Second, insofar as the revelation really gave her good reasons for believing that God exists, and this is *still* the case after she has apprehended the professor's arguments, can we really fault her for not letting those arguments influence her theist belief? At least to me, it seems that we should only fault her for her "doxastic passivity," insofar as the professor's arguments at least provided her with good reasons *not* to believe that God exists, i.e. good reasons for either agnosticism or atheism. Otherwise it escapes me why we should require her to give up a belief, which, after all, she still had good reasons to hold.

However, in this case we have again left behind the scenario Russell needs: If the professor provides the student with good reasons for agnosticism or atheism, she no longer has good reasons for theism, contrary to the necessary suppositions. Now, as we saw above, Russell's own examples have hardly strengthened the case for the claim that second-order reasons are at all relevant to the notion of epistemic blameworthiness. I may therefore safely confess my inability to see, how a convincing example can be set up to meet Russell's requirements. It seems that in fact second-order reasons *are* irrelevant to evaluations of epistemic blameworthiness.

In conclusion: No matter if Russell is interpreted as purporting to offer necessary and sufficient conditions or only unnecessary, but sufficient conditions for epistemic blameworthiness, his claims offer no threat to $\text{Blame}_{\text{epist}}$, the analysis of epistemic blameworthiness preferred here.

DOXASTIC CONTROL

Abstract. In this chapter I point to the crucial link between the notions of doxastic control and epistemic blameworthiness. In Section 7.1 I shall argue against some recent authors that a legitimate ascription of epistemic blameworthiness demands the relevant agent to be in control of the relevant belief in some sense. In Section 7.2 I proceed to discuss the general requirements that a mode of doxastic control must satisfy in order to be legitimately invoked in such ascriptions. In Section 7.3, then, I present a taxonomy of modes of doxastic control. Previous authors, William Alston in particular, have offering similar taxonomies, but the present differs from its predecessors on several counts. I bring out the most important differences and defend my preferred solution. In Section 7.4 I provide a brief overview of the chapters ahead (Chapters 8–11), where I shall discuss under which conditions the diverse modes of doxastic control presented in Section 7.3 may in fact be invoked to underwrite legitimate ascriptions of epistemic blameworthiness.

7.1. THE IMPORTANCE OF THE ISSUE OF DOXASTIC CONTROL

The issue of doxastic control, the control an agent enjoys over her holdings of beliefs, is absolutely central to epistemic deontology, at least if based on a deontic concept as strong as epistemic blameworthiness. If doxastic control is generally impossible or entirely irrelevant to evaluations of epistemic blameworthiness, a grim dilemma arises: Either we will have to accept an account of the notion of epistemic blameworthiness denying that the blameworthiness of a belief invariably stems from the believer's failure to exercise her doxastic control appropriately. Or we will have to give up the concept of epistemic blameworthiness altogether. The first possibility has recently been advocated by David Owens, the latter by William Alston.²¹³

None of the horns of this dilemma looks particularly inviting on a closer inspection. Owens boldly grabs the first horn by arguing that an agent's epistemic blameworthiness is linked, not to her possibilities of doxastic control, but to her fundamental standing as a person:

I'll argue that responsibility is what is registered when blame is applied to persons *as such* . . . (. . .) . . . My hypothesis about responsibility is that one is held to account for one's merit as a person; one gets blamed for those things (both moral and non-moral) which are thought to make one a bad person.²¹⁴ (emphasis in original).

Owens' central argument for the above thesis may be paraphrased thus: He first observes with Hume that "*sentiments* are every day experienced of blame and praise, which have objects beyond the dominion of the will and choice . . ." ²¹⁵ He then goes on to claim that such sentiments are often entirely appropriate, whether in the form of resentment and indignation directed towards fellow agents, or in the form of remorse and guilt directed towards one self. Only, the blameworthy state or condition needs to be "subject to rational assessment" ²¹⁶ and in some sense vicious, i.e. a state or condition that makes one "a bad person."

Owens takes this to mean that any kind of blameworthiness is entirely independent of the possibility of control. He uses as a illustration of this point the example of road rage, which to him remains a blameworthy emotional condition even if a driver, after unsuccessful efforts to temper herself through therapy, has finally discovered that her rage towards fellow road users is entirely beyond her control:

Suppose two of us suffer from road rage. We each assiduously attend anger-management classes and do our relaxation exercises in the car simulator with the same level of commitment, but the therapy works for you and not for me . . . (. . .) . . . Here one's reproaches would be tempered by a knowledge of my efforts at self-improvement, but one should hardly cease to blame me, as one would cease to blame you: continued guilt on my part would be not at all inappropriate.²¹⁷

Thus, to Owens' mind, an agent's blameworthiness for her states does not even require the possibility of a modest mode of indirect control over such states as, e.g. exerted through wilfully accepted education. Translated into the epistemic realm, Owens thus maintains that whether or not an agent was ever in any position to improve her cognitive dispositions, she may remain at fault for the results of her deficiencies.

I think we should concede without hesitation that a person suffering from road rage or a person holding certain undesirable beliefs, is in some sense a worse person than a person not suffering from road rage or holding such beliefs. Also a person's beliefs as well as her emotional states are in some sense subject to rational assessment. However, it is not at all easy to see why the supposition that a state of anger may be judged irrational or rational should matter to its deontic status, if this state remains entirely beyond the angry agent's control. It may of course be true that people keep on blaming agents, including themselves, for the occurrence of certain undesirable states entirely outside of their control. This might even typically be the case for states subject to rational assessment. The crucial question remains, whether such attitudes are merited by states of blameworthiness.

Owens, I take it, has not presented a convincing argument that they are in fact so merited. Rather, a natural response to his road rage example is to acknowledge that, if the angry driver's condition is truly outside of her past and present control, it is simply inappropriate to blame her for it, even if she is in some sense a bad person and may need to be contained due to the harmfulness of her actions. The intuitive force that Owens' road rage example might in fact command might well rest in its sheer psychological implausibility: Certainly we would expect an angry driver to enjoy some past or present measure of control over her condition, either because she has brought it upon herself by her actions, or because she has or has had access to some means, which she could effectively apply in order to temper her condition. If her rage is truly outside of her past and present control, however, just like a person born irredeemably ugly, the angry driver seems to deserve our pity rather than our resentment. This is quite consistent with the fact that, in the case of an actual driver suffering from uncontrollable anger, it might be almost impossible for us to realize her blamelessness and even harder for us to treat her accordingly.

In contrast, Owens thesis seems to resemble a doctrine of hereditary sin, maintaining that persons remain guilty for their badness, even if they are bad entirely due to forces outside of their control. In the present context I shall be satisfied to submit

that such a position is better avoided as long as a consistent alternative is available. Certainly Owens' example does not force his conclusions upon us.

It deserves mention that Richard Feldman, like Owens, has recently offered an account of deontological epistemic normativity that explicitly does not presuppose any possibility of doxastic control. Feldman insists that the "ought" involved in deontic epistemic evaluations "judgments" is really a "role-ought" based on "good performance"²¹⁸:

Furthermore, it is plausible to say that the role of a believer is not one that we have any real choice about taking on... (...)... It is our plight to be believers. We ought to do it right. It doesn't matter that in some cases we are unable to do so.²¹⁹

He then explicitly argues that his account of epistemic deontology can underwrite the notion of a belief being epistemically prohibited or permitted. I seriously doubt whether this is so, unless these terms are used in a highly unorthodox manner.²²⁰ In fact it is not at all clear that "role-oughts" are deontic terms in the present sense (see Section 3.1): An actor may well be said not to play her role in a play as she *ought* to play it, even if playing it right exceeds her abilities. However, we would hardly hold her blameworthy simply for that reason.

This point is conceded by Feldman, who submits that, given our supposed lack of doxastic control, "it is less clear what to make of judgments in which people are praised or blamed for their beliefs."²²¹ Thus, even if superficially similar to Owens's analysis, Feldman's is really much weaker in the sense that he does not try to account for the central notion of epistemic blameworthiness. For that reason it is also irrelevant to the current proceedings.

The second horn of the dilemma set up in the beginning of the present paragraph faces problems of another kind. As Alston readily recognizes, deontic epistemic evaluations of beliefs are so firmly entrenched in our ordinary discourse that discarding them easily leads to an "uneasy linguistic conscience."²²² Further, as should be clear from Chapter 1, we need the notion of epistemic blameworthiness in order to account for certain compelling cases of moral blameworthiness as well as to justify certain common reactive attitudes towards one another as believers. Abandoning as unviable ascriptions of epistemic blameworthiness would arguably mean a serious impoverishment of our general normative framework.

In order to avoid the teeth of the dilemma, we must then make the following basic supposition: An agent can only be blameworthy for holding a belief, if she has exercised, or could have exercised, some kind of doxastic control over that belief.²²³

This supposition can now be substantiated within the framework of the basic analysis of epistemic blameworthiness ($\text{Blame}_{\text{epist}}$) offered in Section 4.1: $\text{Blame}_{\text{epist}}$ hinges on the notion of appropriate excusability. Now, if no account of the doxastic control underwriting ascriptions of epistemic blameworthiness could be given, we would be at a total loss concerning what agents should appropriately be excused for: It is highly dubious whether the notion of excusability could then apply to belief-holdings at all, insofar as it simply seems inappropriate to excuse an agent for some undesirable state of affairs she could not have controlled in any sense. For example it

does not make much sense to say about an average agent, except as a joke, that she is appropriately excused for the bad weather situation, even if this situation is of course not “her fault”: The bad weather is simply something she could never have controlled in any sense, hence something for which she can never stand in need of an excuse. If asked to excuse her for the occurrence of the bad weather, rather than offering an excuse to exempt her from guilt, the most natural response would be to dismiss the demand as genuinely misguided.

However, if it makes sense that an agent could be excused for some undesirable occurrence, there are clearly two basic ways in which her control over that occurrence could enter into the excuse: Either it could be claimed that she did in fact bring about the undesirable occurrence, but that in doing so she did nothing inappropriate, for which reason the occurrence was none of her fault. Or it could be claimed that, even if at some time she omitted to exercise her control over that undesirable occurrence and it occurred because of this omission, she did nothing inappropriate by omitting to exercise her control. Upon these basic observations I will present the following analysis of the notion of epistemic blameworthiness, which refines the inexcusability clause of the initial analysis provided in Section 4.1 above:

$\text{Blame}_{\text{epist, control}}$: An agent A is *epistemically blameworthy* for holding the belief that p *if, and only if*

1. She believes that p.
2. Her holding of the belief that p is *epistemically undesirable*.
3. She enjoys a mode of *doxastic control* M belonging to the set \mathcal{M}^* over her holding of the belief that p and it is either the case that
 - a. The epistemic undesirability of her holding of the belief that p²²⁴ is caused by an exercise of M and she has *no appropriate excuse*²²⁵ for that exercise. Or
 - b. The epistemic undesirability of her holding of the belief that p is caused by an omission of an exercise of M and she has **no appropriate excuse** for that omission.²²⁶

Here \mathcal{M}^* is a suitable set of modes of doxastic control, such that $\text{Blame}_{\text{epist, control}}$ provides necessary conditions of epistemic blameworthiness, i.e. such that the analysis may adequately cover all possible cases of epistemically blameworthy belief.

Obviously, now, the substantive import of $\text{Blame}_{\text{epist, control}}$ hinges upon which modes of doxastic control must, as a minimum,²²⁷ be demanded to be members of \mathcal{M}^* , i.e. which modes of doxastic control the variable M must minimally range over in order to make $\text{Blame}_{\text{epist, control}}$ adequately cover all possible cases of epistemically blameworthy belief. Specifying this range will occupy me below, where, after having introduced a suitable taxonomy of modes of doxastic control in Section 7.4, I shall go about the task by simple enumeration. To make this process more manageable I introduce the following definition:

(Deontological Adequacy): A mode of doxastic control is **deontologically adequate** with regard to an agent’s *epistemically undesirable* belief that p *if, and only if*, if the variable ‘M’ is interpreted as denoting this mode of doxastic control, either the condition (a) or the condition (b) of $\text{Blame}_{\text{epist, control}}$ yields *sufficient* conditions for the epistemic blameworthiness of that belief.

In less technical, but also rough, terms: A mode of doxastic control is deontologically adequate with regard to an epistemically undesirable belief if, and only if, it may legitimately be invoked in ascriptions of epistemic blameworthiness to that belief. In what follows, for ease of expression, I shall often say that a mode of doxastic control is deontologically adequate, if it is deontologically adequate with regard to some non-empty domain of undesirable beliefs.

It should now be clear that a mode of doxastic control must belong to \mathcal{M}^* , i.e. be within the minimal range of M , if, and only if, it is deontologically adequate with regard to some non-empty domain of undesirable beliefs: It is exactly the modes of doxastic control that play a role in the deontic evaluation of particular undesirable beliefs, which cannot be dispensed with, if $\text{Blame}_{\text{epist, control}}$ is to adequately cover all possible cases of epistemic blameworthiness.

The primary task ahead, then, becomes to specify exactly which modes of doxastic control are in fact deontologically adequate under which circumstances. This project shall occupy the Chapters 8–11. As shall fully emerge in the run of these chapters, during recent philosophical history the issue of doxastic control and its relevance to epistemic deontologism has been a subject of much controversy. This controversy is hardly surprising. As I shall hope to make clear, no neat and tidy theory can account for all legitimate ascriptions of epistemic blameworthiness.

7.2. SOME BASIC REQUIREMENTS FOR DEONTOLOGICAL ADEQUACY

Due to the many differing versions of doxastic control suggested in the literature and the differing opinions over which kinds of doxastic control are relevant to evaluations of epistemic blameworthiness, imposing some general requirements for deontological adequacy will greatly help my discussion.

Most importantly, I shall submit that a mode of doxastic control capable of underwriting an ascription of epistemic blameworthiness must *at least* satisfy the following three requirements:

1. It must be possible for a normal agent to exercise the relevant kind of doxastic control under normal circumstances. That is, the possibility of exercising this control must not be limited to, e.g. a few esoterically trained agents (Tibetan lamas or other “mental masters”) under highly exotic circumstances. Call this *the naturalistic requirement*.
2. It must be possible for a normal agent to guide her exercise of the relevant kind of doxastic control according to the norm of epistemic blameworthiness (a basic outline of such a norm ($\text{Blame}_{\text{epist, control}}$) has been offered above), at least to the degree that her failure to do so may, under appropriate circumstances, warrant an accusation of inappropriate conduct. Call this *the activity-guiding requirement*.
3. The exercise of the relevant kind of doxastic control must be related to the ensuing holding of a belief in an epistemically relevant way. Call this *the epistemic relevance requirement*.

These requirements now stand badly in need of some explicatory comments. First, notice that in order to achieve a maximum of generality, I employ the notion of

doxastic control in a normatively neutral way (see Section 3.2). This allows me to consider various suggestions for doxastic control found in the literature, without assuming beforehand that these modes of control can underwrite ascriptions of epistemic blameworthiness. If any reader should find “control” a term too normatively invested, the more neutral term “influence” may be a fair replacement. This term, however, has connotations too passive to cover adequately the wide array of modes of doxastic control considered in the present chapter. For this reason, I shall in most contexts prefer the former term.

I take the import of the first two requirements to be fairly obvious: If *the naturalistic requirement* is not met, I submit, basic intuitions are violated: For example the intuitive force of the “educated racist” example from Section 1.2 arguably rests on the tacit supposition that this agent is not equipped with any extra-ordinary cognitive powers.

The need to enforce *the activity-guiding requirement* should be equally obvious. If this requirement is not met, intuitively an agent, even if having the possibility of exercising a kind of doxastic control, cannot be blameworthy for holding a certain belief, since she can never be appropriately accused of any inappropriate conduct whatsoever: The contrary would be like blaming a construction worker for failing to follow a manual written in a language she could not possibly understand. The agent cannot then on $\text{Blame}_{\text{epist,control}}$ lack an appropriate excuse for any epistemic undesirability her beliefs might have. Most importantly, the activity-guiding requirement must demand that the relevant kind of doxastic control can be exercised with an appropriate amount of *foresight*. The crucial link between the notions of control, blameworthiness, and foresight has been explicitly recognized by several authors.²²⁸ It should be noted that of course the required possibility of foresight does not demand that in order for an exercise of doxastic control to establish blameworthiness, the agent must *actually* have foreseen the consequences of that exercise. Rather it is enough that she *should* have foreseen them in a certain sense. I shall investigate the complicated nexus between control, blameworthiness and foresight in Chapter 13.

The epistemic relevance requirement, now, demands a more elaborate explanation. I shall offer the following analysis of the notion of epistemic relevance:

A mode of doxastic control is *epistemically relevant* if, and only if, it may influence the desirability level of a belief along one or more non-deontic dimensions of epistemic desirability.²²⁹

However such an analysis may be challenged, as the following consideration will show. If, namely, our beliefs are controllable as are our actions (I shall argue that in fact they are not), we may well face the possibility that we are still responsible for a belief, even if we could not have influenced its occurrence at all, had in fact we tried. Recently, the possible relevance of such examples to the present context has been brought out by Sharon Ryan:

Once I've decided to believe that p, I will believe p, and I can't then just take on an alternative doxastic attitude. Does this show that my doxastic attitudes are not free? . . . (.) . . . Suppose I am in a room listening to music and enjoying myself thoroughly. I do not wish to leave the room. In fact, there is

really no place I'd rather be right now. Unbeknownst to me, my friends have locked the door to my room . . . (. . .) . . . I stay in the room freely, even though I could not do otherwise and leave the room. As long as I'm believing what I meant to believe, I'm free.²³⁰

I shall have more to say on this compatibilist approach to epistemic control below (Section 8.3.1.1). Ultimately, I shall argue, Ryan does not succeed in vindicating a mode of deontologically adequate mode of doxastic control. Here it suffices to notice, that Ryan's locked-room example presents no real challenge to the proposed analysis of the notion of epistemic relevance, as the following consideration will show: The music-listener could hardly be blamed for staying in the room, if this was not somehow undesirable.²³¹ Now, even if she cannot leave the room, she may well influence the moral desirability level of her stay in the room, in as far as it may matter morally whether she stays in the room for sheer pleasure, or because she has recognized her captivity. Likewise, even though an agent may be unable to prevent herself from believing a certain propositional content under given circumstances, she may well be able to exercise some control over whether she holds this belief on inadequate evidence or not, i.e. whether it violates the desideratum of adequate basing. Such "epistemic Frankfurt-examples" shall play a crucial role in the present argument. Explicit examples shall be offered in the relevant context (Section 10.4).

I shall now briefly consider some examples of doxastic control from the literature, which may be immediately excluded from the following discussion due to a lack of epistemic relevance:

Richard Feldman has acutely pointed to a rather banal way in which normal agents may control some of their belief-formations: Many states of affairs I can bring about by performing an action. Further, if my cognitive system registers such states reliably, and I bring about such a state, I will form a belief that that particular state obtains. A particularly banal example is the following: Given that the electrical systems of the room that I occupy function correctly, I can control whether I believe that the lights are on or not: Simply switch on the lights to form the belief that it is on, and switch it off to form the belief that it is off.²³² I will term doxastic control exercised in this manner "Feldman-control."

Now, an agent's power to exercise Feldman-control can hardly underwrite standard ascriptions of epistemic blameworthiness. First, there are simply too many kinds of belief that are not subject to this kind of control. In Feldman's words:

The existence of the sort of voluntary control over our beliefs described here [i.e. Feldman-control] does nothing to alleviate the worries about the viability of deontological epistemic judgments. There are many beliefs that are not about states of the world that we can control, so no argument like mine can support the conclusion that we have control over them. Yet we make deontological epistemic judgments about beliefs that we can't control [via Feldman-control], and these judgments are as routine and commonplace as are judgments about beliefs that we can control.²³³

Second, Feldman-control also falls short of underwriting epistemic blameworthiness for a more decisive reason:

The control that we do have [i.e. Feldman-control] comes at a wrong point in the belief-forming process. Favourable epistemic judgments are made when a person believes the right thing given the evidence the

person has, and unfavourable judgments are made when a person believes something not supported by the person's evidence. What my argument shows is that we can control our evidence and thereby control our beliefs. The argument does not contest the proposition that we are at the mercy of our evidence (or perhaps our evidence combined with whatever other nonevidential facts cause belief). But it is this fact that is, I think, the real heart of the worry about epistemic deontologism.²³⁴

Now, Feldman can hardly take refuge in the claim that an *epistemically relevant* mode of doxastic control must matter to the *content* of ensuing beliefs. It would seem that in the above passage, he comes fairly close to this: He argues, correctly, that no matter how I exercise my Feldman-control, my beliefs are equally "at the mercy of my evidence," i.e. their content is equally beyond my control, but this in itself is not enough to rule out a notion of epistemic blameworthiness built on Feldman-control: As noted by Sharon Ryan above, there are even cases of action, where we are entirely "at the mercy" of causal factors beyond our control, yet may still be appropriate objects of deontic evaluation.

Rather, Feldman-control is irrelevant to ascriptions of epistemic blameworthiness for a related, but significantly different, reason: Beliefs formed through Feldman-control are always epistemically desirable (it was assumed that they accurately registered what was actually the case). Further, the *failure* to exercise Feldman-control can never be undesirable from an epistemic point of view, even if, in bizarre cases, such failure may still have *moral* significance.²³⁵ That is because, on Feldman's assumptions, no matter how I exercise my Feldman-control my relevant beliefs *remain* epistemically desirable and even *equally* desirable: If the light are off, I reliably register this and *vice versa*. Consequently, Feldman-control can never underwrite an ascription of epistemic blameworthiness on a conception like $\text{Blame}_{\text{epist, control}}$: It simply never yields epistemically undesirable beliefs. Whether an agent exercises it or not remains entirely irrelevant to ascriptions of epistemic blameworthiness.

A second kind of doxastic "freedom" or "control" which is clearly epistemically irrelevant, should be mentioned. Ryan's above-quoted invocation of the doxastic freedom of "believing what one meant to believe," may easily be given an interpretation that renders such "freedom" entirely irrelevant to the issue of epistemic blameworthiness: There is indeed a *stoic* sense of freedom according to which I am *freely* a slave as long as I do not "mean" to be otherwise. Now, even if a stern stoic "means" herself or even her daughter to suffer from a fatal cancer or "means" the clouds to rain on her annoying neighbour, we would hardly ever hold her blameworthy for any of this for that reason. Thus Ryan will have to provide a more substantial account of doxastic control than merely "believing what one means to believe" in order to underwrite ascriptions of epistemic blameworthiness. As I shall argue below (Section 8.3.1), she has not provided such an account.

7.3. A TAXONOMY OF MODES OF DOXASTIC CONTROL

No serious attempt to investigate the relation between doxastic control and epistemic blameworthiness can be made without an adequate taxonomy of modes of doxastic control. In this section I shall offer such a taxonomy, which, I submit, will allow me

to classify in a rewarding manner all serious suggestions in the recent literature for modes of doxastic control relevant to evaluations of epistemic blameworthiness. For that purpose I will have to risk a substantial increase in technicality relative to what is encountered elsewhere. I trust it will be clear that this is no mere sophistry but necessitated by the entangled nature of the topic and its literature. As the resulting taxonomy differs significantly from influential taxonomies in the recent literature (in particular that of William Alston), after presenting my preferred taxonomy, these deviations will have to be addressed before further advances can be made.

I shall consider as significant four distinctions between modes of doxastic control: Doxastic control may be:

1. Positive or negative.
2. Genetic or abortive.
3. Direct or indirect.
4. Content directed or property directed.

To enjoy *positive genetic* control over a belief is to have the ability to bring oneself to *form* a belief or instantiate its normatively significant properties²³⁶ in a given situation. *Negative genetic* control relates to the ability to *prevent* oneself from forming a belief or instantiating its normatively significant properties in a given situation. Corresponding to the above, *positive abortive* control relates to the ability to bring oneself to *suspend* a belief or its normatively significant properties in a given situation, whereas *negative abortive* control relates to the ability to bring oneself *not* to suspend, i.e. *sustain*, a belief or its normatively significant properties in a given situation.

These modes of control are decisively different, as some comparisons with similar modes of control over actions and bodily functions will show: Over my *actions* I enjoy positive genetic control,²³⁷ but not always abortive control of any kind: If I quickly move my arm down on my knee, I may well lack the ability to stop it in mid-air, once the action is instantiated. Over my *patterns of action* (e.g. swimming), however, I typically enjoy negative as well as positive abortive and genetic control. Over some of my bodily reflexes, as, e.g. breathing, I enjoy a limited amount of positive abortive control and positive and negative genetic control (once I have stopped breathing) and (life-time) unlimited negative abortive control (when above water). Over others, as, e.g. sneezing, I enjoy a limited negative genetic control, but neither positive genetic control (without the use of, e.g. sneezing powders) nor positive abortive control, though I might train myself to acquire a small amount of negative abortive control (sustaining the sneeze, once it begins). Over falling asleep, I (frustratingly) do neither enjoy much positive genetic – nor positive and negative abortive control. However I may well enjoy a good deal of negative genetic control, i.e. I may be able to prevent myself from falling asleep.

When trying to gain a fruitful perspective on the position known as doxastic voluntarism below, the distinctions discussed above will prove highly important.

The last two distinctions between modes of doxastic control are a bit harder to handle. The distinction between direct – (or “immediate”) – and indirect control I take

to be the difference between the control we enjoy over our actions and the control we enjoy over their consequences. We may exercise *indirect* control over something by doing something else as a means to bring about that something as an end, e.g. tying down an arm to prevent it from twitching.²³⁸ However, in many cases it is far from clear which events are part of an action, and which are genuine consequences of that action.

To bring the distinction between direct and indirect control into a clearer light, it is necessary to risk a short digression into the issue of responsibility for *actions*. The concept of agency is itself a difficult one to handle. In the present context I shall follow Donald Davidson in making the following supposition, which seems to capture the core of the notion of agency:

Action does require that what the agent does is intentional under some description . . .²³⁹

I take this to mean that an event can only be an action if, under some (appropriate) description, the agent was able to bring it about simply by *intending/willing* to bring it about, i.e. without having to do something else in order to bring it about.²⁴⁰

I shall assume that an agent is always responsible for her actions, so conceived. I make this assumption, which I consider fairly uncontroversial, in order to avoid unnecessary complications with the issue of responsibility for actions.²⁴¹

The qualification “under some description” in Davidson’s above-quoted analysis is needed, because many actions are clearly not intentional under all legitimate descriptions (intentions are semantically opaque). A particularly clear example is that of committing a mistake, such as, e.g. misreading a sign. This is an action, that under the description just given cannot possibly be intended; if it were, it would not be a mistake proper. Now, a misreading easily allows of a re-description under which it is intended: the misreading is also a particular act of reading (only one that falls short of certain standards), and supposedly the very reading of the sign was intended.²⁴²

The present analysis of the notion of agency may gain weight from considering another example due to Davidson: Suppose that a naval officer aboard a World War II British submarine presses a button that fires a torpedo towards the German battleship *Bismarck* and sinks it. Now clearly, if the officer intended to sink the *Bismarck* by pressing the button, sinking the *Bismarck* was his action. Suppose now that the officer pressed the button with the intention of summoning a steward to bring him a cup of tea, accidentally sinking the *Bismarck* in the process. Still sinking the *Bismarck* qualifies as the officer’s action²⁴³ (who sank the *Bismarck*? He did!), and Davidson’s analysis accounts for this. The action *sinking the Bismarck* may in the present case be re-described as *pressing the button at time t* and that he did intend. Only if, e.g. he fell against the button because a wave upset his balance, sinking the *Bismarck* fails to be something he did in any substantial sense, at least in such a case he can hardly be either praised or blamed for this military feat with any appropriation.

The fact that almost any action allows of a re-description under which it fails to be intentional has been readily exploited by some doxastic voluntarists wanting to draw an analogy between belief-formation and supposedly non-intended actions. This is a clever version of the voluntarist argument I shall label “the special analogy

argument” and treat at length below (Section 8.3.1.1). As any other version of this argument it fails: As should already be clear by now, offering a description of certain actions under which they are not intended goes no way towards rendering belief-formation a kind of agency. Actions fail to be actions unless produced by intentions under *some* description. Hence the voluntarist, rather than providing descriptions of certain actions, under which they are not intended, must provide a legitimate description under which a belief-formation *is* intended, and this I shall argue, she cannot convincingly do.²⁴⁴

With the above digression in mind, an explication of the distinction between direct and indirect doxastic control may at last be given: I shall say that an agent *directly* controls the doxastic occurrences that she may, under some appropriate description of them, bring about simply by intending to bring them about. In other words, doxastic occurrences under direct control are those occurrences, which amount to *belief-influencing actions*. *Indirect* doxastic control is then naturally construed as doxastic control, which is not direct, e.g. the control we enjoy over those *doxastic consequences* of our belief-influencing actions, which are not themselves such actions.

A few examples might clarify the above distinction. I shall stress below (Chapter 9) that an agent may exercise a mode of *direct* control over her beliefs by way of her intentional control over her perceptual apparatus. For example she might exercise such control by keeping her eyes peeled. This is because, should she form some beliefs due to her visual identification of some object, the state of her visual apparatus at that moment may matter to the epistemic quality of the ensuing beliefs, in particular the reliability of the process leading to their formation. Her action of keeping her eyes peeled may then appropriately be re-described as an instantaneous improvement of the epistemic quality of her visual beliefs, or a “refinement” of those beliefs, to use a phrase borrowed from John Heil.²⁴⁵ Describing it as the *formation* of a certain belief, however, is hardly appropriate. As I shall argue in Chapter 8, our intentions have no direct power over particular belief-contents.

On the other hand, I shall stress (see in particular Chapter 11) that an agent may exercise an *indirect* mode of doxastic control over some of her beliefs by way of her influence on her cognitive dispositions. We might, e.g. imagine an agent, who improves her critical abilities concerning some subject matter by reading a book. This improvement may affect the epistemic quality of some of her beliefs long after she has finished reading it.

It is vital to notice that direct control is in a crucial sense subordinate to indirect control from a deontological perspective: If namely an agent enjoys a deontologically adequate mode of *direct* doxastic control over a belief, her possible corresponding mode of *indirect* doxastic control over that belief immediately becomes deontologically inadequate. This may quickly be seen from considering an example from the practical realm: If I am in direct positive genetic control over the event of my arm moving to hit you with a hammer, my possible previous efforts to indirectly bring about that event cannot matter at all to the blame I may face for bringing it about: If I am to blame for hitting you with the hammer, I am to blame because, in the

moment I swung the hammer, I was in direct control of this. Thus, a mode of indirect control gains deontological relevance only once the deontological adequacy of the corresponding mode of direct control is excluded.

A final note on the direct-indirect distinction is in place: By dividing modes of doxastic control into direct and indirect modes in the above manner, I am effect assuming that an agent may only be blameworthy for an undesirable state or event, if that state or event is either one of her actions (or omissions of action) or a consequence of her actions (or omissions of action). I shall defend this assumption, which to me seems almost a truism, only indirectly. In particular I shall aim to debunk an attempt to circumvent the relevance of intentional action to ascriptions of blameworthiness in favour of a control by “character” in Section 8.3.1.2. The claim that blameworthiness does not demand the presence of control of any kind, I have already discarded in Section 7.1.

Let us move on to consider the basic distinction between content-directed and property-directed doxastic control. Content-directed doxastic control (content control) can take at least two forms, only one of which is to be taken seriously here. The form that shall be taken very seriously is the control we enjoy over adopting a doxastic attitude towards a certain propositional *content*, in the present context the attitude of *belief*. This is the form of control, diverse direct modes of which has been invoked by doxastic voluntarists to account for the legitimacy of ascriptions of epistemic blameworthiness.

The form of content control, which shall not concern me much, is the control an agent could supposedly wield over the very propositional contents of her beliefs. To use a formulation unacceptable to many contemporary philosophers of language, this would be an agent’s control over the *meaning* of a content, which she believes: An agent having this kind of control over her belief that grass is green would, e.g. be able to bring it about that “that grass is green” (insofar as she believes this proposition) comes to mean, e.g. that the sun is red. At least since Wittgenstein’s late writings became known to wider circles, the possibility of such control has not been in mainstream philosophical favour. In the current context, I shall not venture further into the complex considerations of private languages and rule-followings. I shall simply submit that the possibility of such control is too farfetched to be taken seriously in the present context.

Rather, I contrast doxastic content control of the kind taken seriously here (from now on, simply: doxastic content control) with property-directed doxastic control (or simply: doxastic property control). Doxastic property control I take to be a control over the properties of a belief relevant to the belief’s level of epistemic desirability.

The deep-seated difference between doxastic property control and doxastic content control may perhaps be clarified by the following analogy: Consider a worker on the assembly line of a car factory, whose job it is to spray-paint blue the cars that roll by him. Imagine that all cars that eventually roll out of the factory must first pass by his position on the assembly line. However, it is not this worker’s job to determine which *particular* cars are fit to roll out of the factory. This decision is taken further down the assembly line, where colour is only one among several features of the

car taken into account. Thus, the spray-painter has a significant amount of *property control* over the finished products of the factory since he controls properties of the finished cars that matter to their aesthetic evaluation. For example by painting all the cars blue, this worker can make it the case that no non-blue car roll out from the factory, without having any say concerning the fate of any particular blue car. In other words, he does not have content control over the finished product.

It should be noted that this analogy is dangerously misleading on a number of counts, since an agent's exercise of her doxastic property control may in fact, as shall become clear later below, often influence the content of her beliefs. In order to get a closer analogy to the doxastic case we should really imagine a science fiction car factory, where sometimes a car disappears if particular colours are applied to it in a certain way or mutates into a different model. Still, I trust that the import of the analogy is sufficiently clear to add intuitive support to the following claims:

Doxastic content control entails doxastic property control, as long as the agent has available to her a choice between contents, which would lead to beliefs with different levels of epistemic desirability. On the above analogy it is clear that if I control which particular cars are allowed to roll out of the factory, and there are among these candidates both blue and non-blue cars, I can control whether blue cars rolls out as well.

However, doxastic property control does not entail doxastic content control: An agent can enjoy a mode of doxastic property control over a belief without enjoying the corresponding mode of doxastic content control over that belief. In terms of the analogy from above, this corresponds to the situation where the spray painter may control that only blue cars roll out of the factory, without thereby controlling which particular blue cars roll out. Consequently, if a mode of property control is deontologically adequate, the possibility arises that an agent can be blameworthy for instantiating the undesirable properties of a belief, without enjoying the least degree of deontologically adequate control over the *content* of that belief. Indeed, as I shall later argue, this is a quite normal situation.

It is important to notice at this stage that in some sense it is doxastic *property control* that wears "the deontological trousers": A belief is blameworthy in virtue of having epistemically undesirable properties, not in virtue of simply being held. Exactly the same point arguably applies to morally blameworthy actions: An action is morally blameworthy in virtue of being bad from a moral perspective, not in virtue of simply being performed. This point is easily missed, since we enjoy an obvious analogue of doxastic content control over our actions: we simply instantiate entire actions at will. Thus, "practical property control," which is arguably the most important mode of control over actions from a deontological perspective, is easily overlooked, overshadowed as it is by our ubiquitous "practical content control."²⁴⁶ This may account for the fact that the deontological primacy of property control has been overlooked by all doxastic voluntarists, occupied as they are with conceiving of doxastic events like belief-formations as some kind of actions. If successful, of course some doxastic property control (and even a direct version of it) would probably

be included in the bargain. However, from a deontological perspective, doxastic voluntarism is a conspicuous case of detour driving.

With the distinctions between indirect versus direct and content- versus property-directed modes of doxastic control now clarified I am in a position to show how these distinctions cut against William Alston's influential taxonomy of modes of doxastic control.²⁴⁷ First, Alston conceives of "direct control" as encompassing only a control similar to the one we enjoy over "basic acts":

Let's call the kind of control we have over states of affairs we typically bring about by basic actions, "basic voluntary control."²⁴⁸

Such "basic voluntary control" Alston also terms "basic *direct* control"²⁴⁹ and contrasts it with "nonbasic immediate control." Here things get fairly complicated, since Alston takes "nonbasic immediate control" to comprise at least some types of control, which other authors have labelled "indirect."²⁵⁰

Dissolving this terminological knot requires a little footwork. First consider how Alston conceives of "nonbasic control":

[L]et's first note that we take many familiar nonbasic overt actions to be voluntary (and their upshots to be under voluntary control) in a way that is sufficient for their being required, permitted and prohibited [i.e. subject to deontic norms]. Consider opening a door, informing someone that *p*, and turning on a light. To succeed in any of these requires more than a volition on the part of the agent; in each case I must perform one or more bodily movements and these movements must have certain consequences, causal and conventional, in order that I can be said to have performed the nonbasic action in question . . . (. . .) . . . Here we might say that the action, and its upshot is under the immediate voluntary control of the agent (more strictly, *nonbasic* immediate voluntary control), even though more than an act of will is required of the agent.²⁵¹

The confusion arises, because several authors have used the term "indirect control" to cover all modes of doxastic control that are not analogous to the control an agent enjoys over her bodily movements (her so-called "basic actions").²⁵² I perfectly agree with Alston that there is nothing "indirect" over the control we typically enjoy over, e.g. opening a door. In fact, I will submit, the control we enjoy over this action is so direct as to cast in doubt the usefulness of Alston's distinction between basic and non-basic direct control: There is simply no interesting difference from a normative perspective between so-called "basic" actions such as moving my arm and "non-basic" actions such as opening a door:

As noticed by Alston, in order for an agent to succeed in opening a door "extra conditions of success" must be satisfied apart from her simply willing it: The door must not be too heavy for her physical strength, etc. However, this is equally the case for even the simplest bodily movements: Even if willing to move my arm I cannot do it, unless it is not tied down, not paralysed, my nerves are not cut, etc. Thus, contrary to what Alston seems to suppose, more than a "simple volition" or intention is required for performing *any* action occurring in the physical universe (including, of course, mental actions. Metaphysical dualism is a position I shall not take seriously here). Consequently, I see no reason to uphold Alston's distinction between basic and non-basic modes of direct doxastic control in the present context.

The present account of *indirect* control also significantly deviates from Alston's 1989 taxonomy. In addition to his dual modes of "direct control" presented above, Alston operates with two further modes of doxastic control: "Long-range control" and "indirect influence." I shall briefly consider in which ways these modes fit into my preferred taxonomy.

Alston introduces the notion of *long-range control* in the following passage:

Long-range control is simply the foil of immediate control. It is the capacity to bring about a state of affairs, C, by doing something (usually a number of different things) repeatedly over a considerable period of time, interrupted by activity directed to other goals. One has this sort of control, to a greater or lesser degree, over many things: one's weight, cholesterol concentration, blood pressure and disposition.²⁵³

The doxastic relevance of this mode of control is brought out ensuingly:

It does seem that we have some degree of long-range voluntary control over at least some of our beliefs. As just noted, people do set out on long-range projects to get themselves to believe a certain proposition, and sometimes they succeed in this. Devices employed include selective exposure to evidence, selective attention to supporting considerations, seeking the company of believers and avoiding nonbelievers, self-suggestion and more bizarre methods like hypnotism.²⁵⁴

From these passages it seems safe to conclude that Alston's mode of "long-range control" encompasses mostly what is here termed "indirect doxastic content control." If the results of hypnosis are instantaneous, perhaps *direct* doxastic content control is suggested by this example, but I find it hard to believe that this could be the case in practice. However, the allusion to *repeated* doings in the first of the above passages is puzzling. I think it is safely ignored in the doxastic context, since Alston's examples do not suggest repetition: Hypnotism might (and perhaps should) be effective on a first try.

Let us now consider Alston's notion of "indirect influence," which he introduces in the following passage:

Now it does seem that we have voluntary control over many things that influence belief. These can be divided into (1) activities that bring influences to bear, or withhold influences from, a particular situation involving a particular candidate, or a particular field of candidates, for belief, and (2) activities that affect our general belief-forming habits or tendencies.²⁵⁵

The greatest problem with this account is the difficulty of seeing any relevant difference from Alston's above-mentioned account of "long-range control." Surely the former account mentioned "repeated doings," of which there is no trace in the latter account, but repetition seemed inessential to "long-range" control in the first place. Thus the "indirect influence" passage seems more like an account of *how* to exercise "long-range control" than an account of a distinct mode of doxastic control: Surely it seems impossible to generate a particular belief from long range, without somehow manipulating the evidence-stream one encounters or change one's general belief-forming habits. Alston's "sect example" from the "long-range control" passage seems a paradigmatic example of strategy (1), his hypnosis example seems a paradigmatic example of a local application of strategy (2). Thus nothing justifies Alston's distinction between two separate modes of indirect doxastic control in the

context of epistemic deontology. Both are effectively modes of indirect doxastic content control in my preferred terminology.

Alston's taxonomy represents a very valuable pioneering effort. It is however unfortunate that it focuses entirely on modes of doxastic *content* control and their deontological adequacy, even despite the fact that his mode of "long-range control" might easily be given a property-directed reading. However, Alston effectively blocks this interpretation: First, he maintains that no mode of direct doxastic (content) control may be deontologically adequate. I shall agree to this point entirely (see Chapter 8). However, lead on by his taxonomy he then concludes that an account of epistemic blameworthiness based on the mode of "indirect influence," which, as we saw, is a mode of indirect *content* control in my terminology, is in fact the "only viable"²⁵⁶ option:

This suggests that even if propositional attitudes are not under our effective voluntary control, we might still be held responsible for them, provided we could and should have prevented them; provided there is something we should and could have done such that if we had done it we would not have had the attitude in question.²⁵⁷

In Section 10.1 I shall consider Alston's detailed version of this approach. I shall argue that not only is this approach *not* the only viable option for epistemic deontology, it is hardly even a viable option at all. Since, as argued above, doxastic property control is really primary to doxastic content control from a deontological perspective, it is highly unfortunate that Alston did not consider the option of a deontologically adequate mode of doxastic *property* control. In Chapters 9 and 11 below I will fill out this gap.

It should be remarked that I do not engage much with modes of *abortive* doxastic control below. I will briefly account for my reasons for this: As shall be remarked in the relevant contexts I take my negative points on modes of *genetic* content-directed doxastic control to be equally valid for modes of *abortive* content-directed doxastic control, although the latter have not loomed large in the literature. With respect to modes of *property-directed* doxastic control, however, abortive and genetic control are *asymmetrical* in the following sense: There simply seems to be no way of exercising a mode of abortive *property-directed* control without enjoying the corresponding mode of abortive *content-directed* doxastic control.

The correctness of this abstract claim may be recognized after some more concrete considerations: Consider some of the central sources of epistemic undesirability: Formation (or preservation) by an unreliable mechanism and basing on inadequate evidence. Obviously, once an unreliably formed belief is formed, one cannot control that one does not have a belief that instantiates the undesirable property of being *formed* by an unreliable mechanism, except if one has the control to *suspend* that belief entirely. But the same seems to go for inadequate basing. How, apart from giving up an inadequately based belief, should one bring it about that one *no longer* has a belief that instantiates the property of being based on inadequate evidence?

The only option seems to be the following: While still holding the belief, I must recognize that I hold it on inadequate evidence. That is: Evidence, which before

seemed adequate to me, must now seem inadequate to me. Further, I must then *begin* basing the belief on adequate evidence or sustain it by a reliable method. That is: I must now be aware of new evidence, which not only seems,²⁵⁸ but really *is* adequate, and this awareness must cause in me a belief with the very same content. I shall not argue that this procedure is impossible. It is however, highly exotic and very unlikely to satisfy the naturalistic requirement for deontological adequacy.

At this point we may attempt to take refuge with the epistemic indesideratum of unreasonableness and consider a case in which an undesirable belief was formed by a reliable mechanism, based on good reasons, yet is epistemically undesirable because it is not supported by good rationalizing reasons. Such a case may arise, if after the formation of the belief the agent's evidential situation changes dramatically. If the agent can now, while retaining her belief, make it the case that this belief is again supported by good rationalizing reasons, she may have exercised a mode of abortive *property*-directed doxastic control, while not relying on a corresponding mode of abortive *content*-directed doxastic control.

Such a case, however, is rather bizarre: Suppose that an agent believes that the lights are on in a certain room and that this belief was reliably based on empirical evidence. However, since the belief was formed, the lights in the room have been turned off. The agent has checked the room with her well-functioning eyes wide open, but still sticks to her original belief. In that case her belief is hardly supported by good rationalizing reasons any longer. Now, in order to exercise the abortive doxastic genetic property control, she must make it the case, without simply suspending her belief, that it is again supported by good rationalizing reason, e.g. by re-observing the room, now with the lights back on. It is clear that this is a bizarre procedure to say the least. I take it then that the principled possibility of such cases does not affect any interesting particular ascriptions of epistemic blameworthiness.

7.4. PREVIEW OF THE ARGUMENT AHEAD

The above taxonomy allows a total of 16 modes of doxastic control. With this taxonomy in place I am now able to accurately state the structure of the argument, which I shall offer in the following chapters. For the purpose of strictness I shall here stick to the full-blown idiom as worked out above.

My most substantial conclusion shall be the following: Normal doxastic agents may exercise either a positive or a negative mode of deontologically adequate genetic indirect property-directed doxastic control with regard to many of their epistemically undesirable beliefs. That is: For many of the epistemically undesirable beliefs generated by her cognitive system, a normal agent may indirectly influence her belief-formation so as to control the instantiation or non-instantiation of properties of her beliefs relevant to their level of epistemic desirability. Further, her power to exercise such doxastic control over those undesirable beliefs may, given appropriate further circumstances, account for her epistemic blameworthiness for holding such beliefs.

To drive home the main conclusion I will need to exclude the deontological adequacy of specific other modes of doxastic control. As argued in Section 7.3 an

indirect mode of doxastic control becomes insignificant from a deontological perspective if the corresponding *direct* mode of doxastic control is deontologically adequate. Thus, I will have to deny the global deontological adequacy of the modes of negative and positive genetic *direct* property-directed doxastic control. In other words, I shall deny that an agent's power to exercise one of these two modes of doxastic control over an undesirable belief may, *in all cases*, be invoked to account for legitimate ascriptions of epistemic blameworthiness to her for holding that belief. I shall however argue that, in a restricted set of cases, these *direct property*-directed modes of doxastic control may account for the legitimacy of such ascriptions.

I will however reject the deontological adequacy of any mode of *direct content*-directed doxastic control. This is a showdown with an approach to deontic epistemic normativity commonly known as "doxastic voluntarism."

As doxastic voluntarism has appeared in a number of guises, rejecting it will be rather complicated. First, in Section 8.2.1 I shall discuss and reject the thesis advanced by a number of authors that modes of *direct content*-directed doxastic control are impossible for conceptual reasons, thus opening up for a genuine discussion of their deontological adequacy. Then, in Section 8.2.2, I shall reject doxastic voluntarism due to a number of plausible empirical observations about the abilities of the human psyche. Most importantly, I shall argue that the relevant modes of doxastic control fail to satisfy *the naturalistic requirement* (see Section 7.2): Agents capable of exercising doxastic control of the relevant types are probably as rare as agents with telekinetic powers, if existent at all. Lastly in Section 8.3 I shall consider and reject a string of arguments to the opposite conclusion. Some of the arguments advanced to support doxastic voluntarism give rise to important considerations pertaining, e.g. to the nature of reasons for belief and the nature of intentions. However, I shall argue that, despite their occasional great ingenuity, all of these arguments fail.

Having put aside modes of *content*-directed doxastic control as serious candidates for deontological adequacy, in Chapter 9 I shall examine the domain of deontological adequacy for modes of *direct property*-directed doxastic control. I shall submit that James A. Montmarquet is alone in the literature in having explicitly discussed the deontological adequacy of such modes of doxastic control at length, although John Heil has advanced sketchy remarks in a similar vein. I conclude that although modes of *direct property*-directed doxastic control may be deontologically adequate in some cases, many interesting cases of epistemic blameworthiness are still left unaccounted for.

Having at this stage established the possible deontological importance of modes of *indirect* doxastic control, I can proceed to the next stage of the argument.

First, in Chapter 10 I shall examine the domain of deontological adequacy for modes of *indirect content*-directed doxastic control. As already remarked in Section 7.3, William Alston in Alston (1989) suggests such a mode as the *only* viable candidate for deontological adequacy. I shall argue that his specific proposal runs into

insurmountable difficulties, although certain modes of indirect content-control may aspire to deontological adequacy under rare circumstances.

However, in Chapter 11 I shall aim to establish that a deontologically adequate mode of indirect *property*-directed doxastic control is in fact not an uncommon phenomenon. Ensuingly, I discuss how normal agents in realistic settings may in fact exercise such modes of doxastic control.

DIRECT CONTENT-DIRECTED DOXASTIC CONTROL
OR DOXASTIC VOLUNTARISM

Man pflegt sich oft der Ausdrücke zu bedienen: *Seinem Urtheile beipflichten, sein Urtheil zurückhalten, ausschieben oder aufgeben*. Diese und ähnliche Redensarten scheinen anzudeuten, dass in unsern Urtheilen etwas willkürliches sei, indem wir etwas für wahr halten, weil wir es für wahr halten wollen. Es fragt sich demnach hier: *Ob das Wollen einen Einfluss auf unsre Urtheile habe?*

Unmittelbar hat der Wille keinen Einfluss auf das Fürwahrhalten; dies wäre auch sehr ungereimt.

Immanuel Kant²⁵⁹

Ich denk' was ich will
Und was mich beglückt

Popular German folksong

Abstract. In this chapter I reject the popular stance on deontic epistemic normativity commonly known as “doxastic voluntarism.” In Section 8.1 I present a generalized understanding of this position, designed to cover all interesting self-acclaimed proponents of the position in the recent literature. It emerges that doxastic voluntarism relies on the possibility of doxastic action (direct doxastic content control) in normal agents. In Section 8.2 I examine the case against doxastic voluntarism. Initially, in Section 8.2.1 I consider and reject the claim due to Bernard Williams and others that doxastic actions are impossible due to conceptual reasons. However, in Section 8.2.2 I proceed to endorse the Psychological Impossibility Claim that such actions are simply not achievable by normal human agents. In Section 8.3 I estimate the opposition to this claim, and consider four popular strategies for countering it. I argue that all of these strategies fail.

8.1. THE IMPORT OF DOXASTIC VOLUNTARISM

By “doxastic voluntarism” I shall here understand any position that argues for the deontological adequacy of one or more modes of direct content-directed doxastic control with regard to a non-empty domain of epistemically undesirable beliefs. In less technical terms: The doxastic voluntarist claims that at least sometimes an agent is epistemically blameworthy for holding a belief, because she has *directly* brought about her holding of that belief by way of a doxastic action (or an omission thereof).

The doxastic voluntarist I shall consider is consequently also an epistemic deontologist.²⁶⁰ I shall not consider positions that argue for modes of doxastic “freedom” without assuming that such “free” beliefs are also proper objects of deontic evaluation. A position left out of consideration on this count could, e.g. be a brand of stoicism, arguing that beliefs can be free only in the sense of being regarded with “stoic calm” by the relevant believer. As already proclaimed above (Section 7.2), the possibility of such calm is entirely irrelevant from the present perspective.²⁶¹

Also, Bas C. van Fraassen has advocated a self-proclaimed “voluntarist”²⁶² position without explicit deontological ambitions. Van Fraassen argues that beliefs or

“epistemic judgments” bear certain similarities to fairly straightforward actions like making a promise:

It seems then that, of the alternatives examined, epistemic judgments are most like expressions of intention. I may express an intention either by simply stating the outcome of what I have decided upon (“You will be my successor”) or by choosing a form of words traditionally suited to such expression (“You shall be my successor”). In either case, it is conveyed that I have made a decision, have formed an intention, am committed to a certain stance or program or course of action ... (...) ... I have, as it were, entered a contract with myself. If I express this intention to an audience, then, just as in the case of a promise, I invite them to rely on my integrity ...²⁶³

Undeniably van Fraassen points to a significant aspect of believing in the above passage. It seems to be his point that holding a belief bears with it a certain degree of “cognitive engagement,”²⁶⁴ which may utter itself *as if* the believer had in fact made some kind of promise to herself. This claim in itself seems quite uncontroversial. However, just as with the “stoic calm” above, such “commitment” in itself hardly makes for any epistemic blameworthiness unless a result of some exercise of doxastic control, and van Fraassen does not talk of doxastic control at all, at most he hints to some terminological similarities between “promise-talk” and “belief-talk” in the passage above. As I shall argue further below (Section 8.3.2.1), these terminological similarities mask substantial differences between the formation of a belief and actions like the making of a promise: Decisions to believe, ordinarily conceived, do not involve anything like “the forming of an intention.” In fact belief-formations in a normal agent do not directly involve intentions at all. Hence it would be uncharitable to attribute to van Fraassen doxastic voluntarism in anything like the sense discussed here, especially given the fact that his self-acclaimed “voluntarism” appears in a discussion without obvious affinities to the issue of epistemic deontology.²⁶⁵

Historically, doxastic voluntarism has come in a number of significantly different varieties. First of all voluntarists have differed over whether to consider occasionally under direct content control *doxastic events* like the formation of a belief or simply *believing* itself. Following Robert Audi we may term the former version of doxastic voluntarism “the genetic version,” the latter “the behavioural version.”²⁶⁶

In the recent literature the genetic version of doxastic voluntarism has been the overwhelmingly dominant one. One possible exception is Matthias Steup who in a passage of Steup (1988) writes:

Moreover, it seems to me, there are positive reasons for the view that **having beliefs** is a form of agency.²⁶⁷

However, Steup’s 1988 commitment to the behavioural version of doxastic voluntarism is far from univocal. In another context in the very same paper he ventures the following claim:

Whenever a belief is an appropriate object for the attribution of epistemic blame, it must have been within the agent’s power to do something that would have led him to refraining from this belief.²⁶⁸

If a doxastic voluntarist in the present sense, Steup here seems to endorse the *genetic* version of the position: The deontologically adequate control referred to here must be of the negative genetic content-directed variety. Not surprisingly, his commitment

to a genetic version of doxastic voluntarism has been more outspoken in later papers, e.g. in Steup (2000) he takes the stance that often beliefs may at least be *generated* by a decision in exactly the same sense as the action of raising my arm under typical circumstances.²⁶⁹

Historically, the only completely unqualified advocate of behavioural doxastic voluntarism seems to be William James in *The Principles of Psychology*. Here James defended the peculiar thesis that belief is identical to the will:

Will and Belief, in short, meaning a certain relation between objects and the Self, are two names for one and the same PSYCHOLOGICAL phenomenon.²⁷⁰

Now, in the same study, James makes the following claim:

Effort of attention is...the essential phenomenon of will.²⁷¹

At this point it seems safe to conclude with Richard M. Gale:

His [i.e. James's] overall argument for belief being an *intentional action*, though he did not use this contemporary terminology, is based on his identification of belief with the will, and the will, at least in one of its senses, with effortful attention to an idea.²⁷²

One might well suspect that the notion of "belief" encountered here is hardly similar to the contemporary notion. However, the following passage makes it clear that at least at some point James thought of belief as a state of taking the world to be in a certain way:

Each of us literally chooses, by his ways of attending to things, what sort of a universe he shall appear to himself to inhabit.²⁷³

In other passages, however, James leaves his reader to wonder whether he is conceiving of wilfulness or choice in anything like the modern sense, thus whether he is really a voluntarist at all:

The accommodation and the resultant feeling *are* the attention. We don't bestow it, the object draws it from us. The object has the initiative, not the mind.²⁷⁴

Here it seems like the control over beliefs is suddenly handed over from the will to the world of objects after all: What started as bright voluntarist bang seemingly ends as at best some compatibilist whimper.

The historical and recent dominance of genetic over behavioural doxastic voluntarism is hardly surprising. Neither are perhaps James's seeming inconsistencies as seen from a modern perspective.²⁷⁵

Behavioural doxastic voluntarism simply appears to be a very unattractive position. As shall emerge, many arguments in favour of genetic doxastic voluntarism may be offered, but none of them offer the slightest support to behavioural doxastic voluntarism.²⁷⁶ I will therefore not consider the viability of *behavioural* doxastic voluntarism further. Rather in the remainder of this chapter I shall try to get into a clearer focus the import of *genetic* doxastic voluntarism. As introduced above, genetic doxastic voluntarism (from now on simply "doxastic voluntarism"), involves the claim that certain doxastic *events* stand under the agent's direct content-directed

doxastic control. A more precise account of these events may now be given using the distinctions between modes of doxastic control introduced in Section 7.3 above:

We face the following four basic modes of direct content-directed doxastic control:

1. If an agent has positive genetic direct content control over a belief (that she does not yet hold) she may, under some appropriate description, bring it about that she *forms* that belief simply by intending to form it.
2. If an agent has negative genetic direct content control over a belief (that she does not yet hold) she may, under some appropriate description, bring it about that she *prevents* (refrains from) forming that belief, simply by intending to prevent it.
3. If an agent has positive abortive direct content control over a belief (that she holds) she may, under some appropriate description, bring it about that she *suspends* that belief, simply by intending to suspend it.
4. If an agent has negative abortive direct content control over a belief (that she holds) she may, at under some appropriate description, bring it about that she *sustains* that belief, simply by intending to sustain it.

Now, given my conception of direct doxastic control as presented in Section 7.3 above, an agent's exercise of any of the four modes of direct content control, amounts to the performance of an action. Thus, to the above four modes of direct content control corresponds four basic types of *doxastic action*:

1. Forming belief.
2. Preventing belief.
3. Suspending belief.
4. Sustaining belief.

In order to get deontologically adequate modes of doxastic control out of any of the above types of doxastic action, the following strong empirical supposition must now be made: The consequences of the performance of any of the four basic types of doxastic action must exhibit a certain *cognitive inertia*: for example it would be disastrous for a version of doxastic voluntarism based on the doxastic action of forming beliefs, if the beliefs formed by the performance of that action always passed away a few seconds after their formation. If the same happened to be the case for all types of doxastic action, this would quite simply render very unlikely the possibility that the occurrence of a long-lasting epistemically undesirable belief as, e.g. the racist belief of "the educated racist" from Section 1.2 was related to a doxastic action at all. However, I believe this "cognitive inertia hypothesis"²⁷⁷ to be comparatively innocuous compared to the far more serious problems, which adherents of doxastic voluntarism will have to face in what follows.

Another issue should be confronted at this stage: It may be thought that suspending belief and sustaining belief are really only species of forming belief and preventing belief respectively. The argument to this conclusion may use "simple" actions as analogies and go as follows: An agent may "suspend" her action of moving her arm by "forming" the alternative action of holding her arm still. Likewise she may "sustain" her action of moving her arm in circles by "preventing" the alternative

action of bringing her arm to a halt. By analogy, it may be claimed, an agent may suspend her belief that *p* by forming the alternative belief that not-*p* and sustain her belief that *q* by preventing the belief that not-*q*.

This analogy argument fails due to an important difference between doxastic states and actions: There is of course no way, in which I can move my right arm and hold it still at the same time. However, there is no reason why I should not be able to hold inconsistent beliefs (I take it here that proponents of this line of reasoning would take contradictory beliefs as the proper candidates for wielding the relevant psychological powers). In fact, I will submit, discovering inconsistencies in one's set of beliefs is not that uncommon a phenomenon: Due to fallacious inferences, I may well find myself to have at some point believed inconsistent mathematical propositions. Thus, the set of four basic types of doxastic action is irreducible.

An implicit appropriation of the above mistaken analogy argument may perhaps explain the account of "Categorical Voluntary Control over Doxastic Attitudes" found in Steup (2000), an account explicitly tailored to mirror a similar account of "Categorical Voluntary Control" over actions like arm-raising. Steup, an avid voluntarist, writes:

I have categorical voluntary control over my doxastic attitude towards *p* if, and only if, (i) I can decide to take an alternative doxastic attitude toward *p*; (ii) if I decide to take an alternative doxastic attitude toward *p*, I can take that attitude.²⁷⁸

Steup here operates with *three* basic doxastic attitudes: belief, disbelief, and suspension of judgment.²⁷⁹ "Suspension of judgment" may plausibly be identified with "suspension of belief" in the present terminology. However, "disbelief" is harder. If "disbelief" means simply "preventing belief" no problem occurs. However, this reading seems unnatural. Given Steup's formulation in the passage quoted above, "disbelief" must supposedly be an attitude, which I can adopt *after* forming a belief.²⁸⁰ Thus, "disbelieving *p*" can hardly mean anything else than "believing that not-*p*." Given this reading, Steupian "categorical voluntary control" is hardly deontologically adequate, for in order to be in "categorical voluntary control" over my belief that *p*, I need only be able to also believe that not-*p*. This peculiar psychological power does not seem immediately relevant to the aims of doxastic voluntarism, since possibly an agent can exercise it without influencing her belief that *p* at all. Consequently, Steup's account would be better off with an explicit acknowledgement of all four types of doxastic actions (or "takings of doxastic attitudes" in Steup's preferred terminology).

With these worries put aside, it is clear that no less than four strategies are open to doxastic voluntarism in accounting for cases of epistemic blameworthiness: Given Blame_{epist, control}, doxastic voluntarism must argue that at least one of the followings claims is true for some non-empty domain of epistemically undesirable beliefs:

1. An agent is epistemically blameworthy for holding the epistemically undesirable belief that *p*, if she holds it because she performed the doxastic action of *forming* the belief that *p* and has no appropriate excuse for this.

2. An agent is epistemically blameworthy for holding the epistemically undesirable belief that *p*, if she holds it, because she omitted a doxastic action of *preventing*²⁸¹ it and has no appropriate excuse for that omission.
3. An agent is epistemically blameworthy for holding the epistemically undesirable belief that *p*, if she holds it because she omitted a doxastic action of *suspending*²⁸² it and has no appropriate excuse for that omission.
4. An agent is epistemically blameworthy for holding the belief that *p*, if she holds it because she performed the doxastic action of *sustaining* it and has no appropriate excuse for this.

The voluntarist literature has centered almost exclusively on the first mode of doxastic action (belief-formation),²⁸³ with only scattered remarks on the other modes of doxastic action.²⁸⁴ This is hardly coincidental. Examples are ready at hand, that seemingly lend themselves to a first mode interpretation. In the example of “the educated racist” from Section 1.2, e.g., it seems natural to say that the racist is epistemically blameworthy for holding her undesirable racist belief simply because she *formed* the racist belief, when she should not have formed that belief and has no appropriate excuse for this.

Ultimately, I shall argue, any version of doxastic voluntarism fails: There is simply no such thing as a mode of direct content control over beliefs which satisfies the requirements for deontological adequacy laid down in Section 7.2 above. It is, however, remarkable that voluntarists have systematically failed to take advantage of modes (2) and (4) which offer genuine possibilities of attributing epistemic blameworthiness and are moreover perfectly analogous to the limited control we normally enjoy over such bodily functions as breathing and falling asleep, which at least to some might seem much better analogues of doxastic states than, e.g. arm-raising.

Historically, various versions of doxastic voluntarism have enjoyed a significant degree of popularity, especially among religious philosophers believing that an atheist belief is something that an agent may be blamed for and that theist beliefs may be achieved by some kind of mental action such as a sudden “leap of faith.” Thus, according to Sharon Ryan, figures no less than Saint Augustine, Thomas Aquinas, Blaise Pascal, Søren Kierkegaard and Leo Tolstoy have lined in the voluntarist ranks.²⁸⁵ The attraction of doxastic voluntarism is not hard to account for: It provides for a relatively simple account of the legitimacy of ascriptions of epistemic blameworthiness, especially when compared to the more complicated accounts I shall present later: As may be seen from the arguments offered in Section 7.3 above, modes of direct content control, if deontologically adequate for a domain of undesirable beliefs, in a sense excludes all corresponding modes of indirect content control or indirect or direct property control from deontological adequacy within that domain: Appeals to the agent’s direct content control over her undesirable beliefs becomes all that matters to the issue of her epistemic blameworthiness for holding those beliefs.

Nevertheless, despite its attractions, doxastic voluntarism utterly fails due to empirical facts about the human psyche. However, before advancing my argument for this

“psychological impossibility claim” I will have to deal with the fairly popular thesis that doxastic voluntarism in its present genetic version is impossible due to *conceptual* reasons.

8.2. THE CASE AGAINST DOXASTIC VOLUNTARISM

8.2.1. *The Conceptual Impossibility Claim*

8.2.1.1. *The nature of the conceptual impossibility claim.* I shall now examine the claim that *genetic* doxastic voluntarism fails due to *conceptual* reasons; the claim that “nothing could be a belief and be willed directly”²⁸⁶ or, to use another phrasing encountered in the literature, that beliefs are simply “logically unwillable.”²⁸⁷ Following Dion Scott-Kakures who has offered the most extensive treatment of this subject so far, I shall term this thesis “the conceptual impossibility claim”²⁸⁸ (CIC).

In the fairly recent literature, CIC has been defended openly by at least Bernard Williams,²⁸⁹ Louis P. Pojman,²⁹⁰ Brian O’Shaughnessy²⁹¹ and Dion Scott-Kakures.^{292,293} These authors agree in proposing the *conceptual impossibility of forming* or *acquiring* a belief simply by *intending* to form or acquire it. Scott-Kakures shows particular care in stating this claim. I shall therefore take his formulation of the claim as canonical:

(CIC) The claim is that there are no transitions of the following kind: At *t*, an agent does not believe that *p* and at *t*+1 the agent believes that *p*, and the transition from the one state to the other is accomplished by a direct and unmediated willing to believe that *p*.²⁹⁴

At all times in the present context, when I talk about, e.g. “producing a belief at will,” “acquiring a belief at will,” “believing at will,” or “believing by fiat” I shall be alluding to a transition of the type described in the above passage. I take it to be implicit in CIC that it allows beliefs to be sometimes generated *accidentally* by an intention to believe. For example proponents of the claim are unlikely to be impressed by an actual case where some intention to form a belief managed to cause the relevant belief by “luckily” triggering some exotic causal mechanism. What is at stake here is arguably really whether it is (conceptually) possible for an agent to perform the *action* of forming a belief, just like she raises her arm.²⁹⁵

The proponents of CIC have cared solely about the possibility of *forming* beliefs at will. Consequently no author has discussed the conceptual possibility of voluntarily preventing, suspending or sustaining²⁹⁶ beliefs. It should, however, be clear that the arguments in favour of CIC discussed below are easily adopted to cover these other types of supposed doxastic action. As I believe these arguments to be fallacious for quite fundamental reasons, given I am right it should be clear that similar arguments aiming to establish the conceptual impossibility of the other types of doxastic action must equally fail.

8.2.1.2. *Three bad arguments in favour of the conceptual impossibility claim.* Various arguments to the conclusion that intentions to believe are somehow “self-defeating” shall be treated at length below. However, in order to clear the territory

I shall begin by considering the three arguments encountered in literature, which do not fall under this heading. I believe these arguments to be fairly easily countered. Brian O'Shaughnessy has advanced the first and second arguments, whereas the third is due to Bernard Williams.

O'Shaughnessy's first argument occurs in the following passage, following his remarks on believing as a *non-processive continuant* quoted in Section 8.1 above:

[S]ince believing is a non-processive continuant, it follows that b-believing [i.e. believing at will] could only be the act of bringing about a momentary and discontinuous event: *the onset* of belief B. Yet if the bringing about of this latter something were an activity, then each phase of that activity would have to relate merely *additively* and hence *non-mechanistically* to the next. (As happens in the phases of the comparable activity: trying to remember a name). But there has to be [a?] mechanism if B-onset is to be *immediately* and *actively installed*. It follows that b-believing cannot be an *activity*. So it has to be the momentary act of bringing about the simultaneous and momentary event: B-onset. So it must be an act that is not an activity: an active event that is not constituted out of an active process. Now are such entities even possible? Do we know of any acts that are not realised through the on-going of an activity? The act of touching occurs at the end of and completes an activity of reaching, but is not itself a distinct act ... (...) ... I cannot believe in these entities, nor a fortiori in a b-believing matching these specifications.²⁹⁷

This passage employs a large inventory of metaphysical assumptions. However, to cut brutally through the intricacies, what goes on here is probably just an illicit *mystification* of mental action similar to the one made in the following version of the *psychological* impossibility claim due to Jonathan Bennett:

There is no reason in principle why some virtuoso of bio-feedback should not be able to focus his attention directly on the neural basis for acquiring the belief that P; but none of us can do this ... (...) ... and probably none of us ever will.²⁹⁸

The common mistake here²⁹⁹ is the supposition that, in order to perform a mental action, e.g. believing at will, one really needs to *do something else* like focusing on one's bio-feedback (Bennett), or adding up phases of some activity (O'Shaughnessy). However, even if this is perhaps so in some cases (as in O'Shaughnessy's example of *trying* to remember a name), I think general experience tells us that it is not universally the case. Consider, e.g. the mental action of calling to mind the image of a red triangle. I believe this is something most people can do immediately and at will, without engaging in any recognizable processive activity (or being for that matter a "virtuoso of bio-feedback"). Thus, O'Shaughnessy's above argument is simply unconvincing: If we are indeed able to believe at will, he has offered us no reason why this mental act should be part of an activity (like trying to remember a name) rather than something immediately achievable (like imagining a red triangle).

However, O'Shaughnessy is quick to offer another argument:

B-believing ... (...) ... must be one of a *limited menagerie* of mental acts. That is, it must be a bedrock *sui generis* type of mental act in the sense in which attending or ratiocinating or trying-to-remember or sustaining-an-image are all of them *types of mental act*. And there are very few of these. For in the sense of "mental act type" here being employed, those mental act headings that are *necessarily instrumental*, such as "arousing desire," do not rate as mental act types ... (...) ... To repeat: while there is a limitless number of mental act headings, there is an extreme shortage of mental act headings that can be used non-instrumentally. Then as a matter of blunt, and I think highly significant fact, no such act type as b-believing *actually exists*.³⁰⁰

The reader may well wonder if I have left something significant out here. If there is even an intelligible argument here, it must run on the premise that the uncanniness of including “b-believing” (believing at will³⁰¹) in the limited-access category of “non-instrumental” mental acts counts as a good reason for denying its existence, as if b-believing were simply unworthy of such exquisite metaphysical company; O’Shaughnessy takes it upon himself to act as a “metaphysical bouncer” guarding the entrance of the elitist club of *sui generis* mental act types. I take it that he simply begs the question against doxastic voluntarism here: If performing the mental act of believing at will is in fact possible, of course it should be assigned to its proper metaphysical category, no matter how exclusive this category might be.

Leaving O’Shaughnessy’s arguments behind, we may now consider the following line of reasoning offered by Bernard Williams:

[A] very central idea with regard to empirical belief is that of coming to believe that p because it is so, that is, the relation between a man’s perceptual environment, his perceptions, and the beliefs that result. Unless a concept satisfies the demand of that notion, namely that we can understand the idea that he comes to believe that p because it is so and because his perceptual organs are working, it will not be the concept of empirical belief... (...) ... But a state that could be produced at will would not satisfy these demands, because there would be no regular connexion between the environment, the perceptions and what the man came out with, which is a necessary condition of a belief...³⁰²

Trudy Govier, Jonathan Bennett and Dion Scott-Kakures have in turn responded to this argument and have all denied its worth in defending CIC.³⁰³ It must be noticed that the above passage really contains *two* arguments: One directed against the possibility of producing *perceptual* beliefs at will, the other directed against the *general* possibility of producing beliefs at will.

The first argument is certainly sound, as long as a perceptual belief is simply *defined* as a belief causally produced by the relevant agent’s “perceptual environment.” One may even concede that perceptual beliefs, so understood, do in fact constitute the bulk of our body of beliefs. Still CIC is not given much assistance. Trudy Govier has forcefully brought out this point:

The empirical way of coming to believe p could remain the normal one, even though there were a few instances of belief by *fiat*. Thus Williams’ second argument also fails to yield a general conclusion about the logical possibility of belief by *fiat*: it fails to do so even if we grant him his basic claim about empirical belief.³⁰⁴

Once the narrower argument concerning empirical belief is put aside, it becomes painstakingly clear that Williams’ general second argument in the above passage crudely begs the question against doxastic voluntarism³⁰⁵: If voluntarism is true, it simply cannot be necessary for a belief to satisfy Williams’ above constraints on its mechanism of formation. Thus CIC clearly stand in need of a more substantial argument.

8.2.1.3. *Bernard Williams’ “self-defeat” arguments.* The most intriguing type argument in favour of CIC appears to be, what we may term “the self-defeat argument.” This argument comes in a prospective and a retrospective version.

The *prospective self-defeat argument* argues that the very *notion* of an intention to believe is in fact incoherent. Thus, agents cannot even have such intentions and genetic doxastic voluntarism utterly fails.

The slightly more modest *retrospective self-defeat argument* argues that even if such intentions are in fact possible, they are *necessarily impotent* insofar as, necessarily, they cannot result in actual beliefs; such beliefs would simply “fade before our very eyes.”³⁰⁶ The more modest version of this argument has it that at least it is conceptually impossible that an agent could hold a belief while aware that it was brought about by will. Below I shall present reasons to reject the idea that even this type of argument adds any support to CIC.

The *locus classicus* for both varieties of the self-defeat argument is the following passage from Williams (1973):

[I]t is not a contingent fact that I cannot bring it about, just like that, that I believe something ... (...) ... Why is this? One reason is connected with the characteristic of belief that they aim at truth. If I could acquire a belief at will, I could acquire it whether it was true or not; moreover I would know that I could acquire it whether it was true or not. If in full consciousness I could will to acquire a “belief” irrespective of its truth, it is unclear that before the event I could seriously think of it as a belief, i.e. as something purporting to represent reality. At the very least, there must be a restriction on what is the case after the event; since I could not then, in full consciousness regard this as a belief of mine, i.e. something I take to be true, and also know that I acquired it at will. With regard to no belief could I know – or, if all this is to be done in full consciousness, even suspect – that I had acquired it at will. But if I can acquire beliefs at will, I must know that I am able to do this; and could I know that I was capable of this feat, if with regard to every feat of this kind which I had performed I necessarily had to believe that it had not taken place?³⁰⁷

In this dense passage no less than three arguments (or claims) occur: The *first* argument is conspicuously prospective, arguing that, since beliefs “aim at truth,” the content of a supposed intention to form a belief cannot be coherent. The *second* “argument” (from “At the very least, there must be a restriction ...”) is a retrospective claim that it is somehow incompatible to hold the belief that *p* and also “know” or even “suspect” (believe?) that the belief that *p* was brought about by will. The *third* argument (from “But if I can acquire beliefs at will, I must know ...”) is stated imprecisely enough to allow of either interpretation.

I suggest that we take this third argument as a prospective argument *ad absurdum*, arguing in less than clear terms that genetic voluntarism is committed to the following pair of contradictory claims: (1) In order to intend something, I must at least believe (know)³⁰⁸ that this intention is potent in bringing about the intended result. (2) On the other hand, given the correctness of the above retrospectivist claim, I cannot believe (know) that an intention to believe is potent in bringing about the intended result. This inconsistency seems to force upon the voluntarist the conclusion that no such intention can possibly exist.

William’s retrospectivist claim has commanded by far the most interest in the literature. I shall, however, begin by engaging with his two prospective arguments. As we saw, his first prospective argument hinges crucially on the idea that beliefs “aim at truth.” In the quoted passage, this supposed feature of belief is given yet another *pseudo-teleological* spin: Beliefs “purport to represent reality.” Surely this must raise legitimate concerns about thwarting beliefs in pursuing such noble goals?

However, once the *literal* meaning of the “truth-aim” is brought out, matters look differently: Earlier in the text, Williams has stated that: “When I say that beliefs aim at truth, I have particularly in mind three things.”³⁰⁹ These things appear to be the following:

1. “[W]hen somebody believes something, then he believes something which can be assessed as true or false, and his belief, in terms of the content of what he believes, is true or false.”³¹⁰
2. “[T]o believe that p is to believe that p is true. To believe that so and so is one and the same as to believe that that thing is true. This is the second point under the heading of ‘beliefs aim at truth’.”³¹¹
3. “The third point, closely connected with these is: to say ‘I believe that p’ itself carries, in general, a claim that p is true.”³¹²

Apart from the bivalence principle endorsed in (1), the three above claims are hardly controversial. However, it is hard to see that we cannot simply concede these points and still maintain that we *can* in principle intend to form a doxastic state satisfying these specifications. The only concern Williams manages to raise against this supposition, is the fact that I would then “know” that I could acquire a belief “whether it was true or not.” However, if this fact poses a threat to voluntarism, it is equally dangerous to any *consciously fallible* mechanism of belief-formation as Trudy Govier has acutely pointed out:

The problem with this argument [of Williams] is that it seems to rule out much more than believing by *fiat*. It seems to rule out all believing, for anyone with sufficient self-knowledge not to regard himself as infallible. Most people know that they have held beliefs, which have turned out to be false and “not representative of reality”; they nevertheless, in believing those things they do believe, believe these things to be true. Now if this is possible, then it would appear that Williams’ argument contains some crucial flaw; for he is maintaining that *because* a person believing by *fiat* would know that he could believe irrespective of truth, believing by *fiat* is impossible.³¹³

Govier’s well-put point shows that Williams’ argument simply has consequences too counter-intuitive to merit serious attention. I think we may safely conclude that Williams’ first prospective argument is not something doxastic voluntarism should worry about. A fair diagnosis seems to be that Williams is simply in the grip of his teleological metaphors here.

Williams’ *second* prospective argument does not fare much better. As we saw, this argument ad absurdum appealed to the soundness of the retrospective argument. However, even granted this premise, the argument faces serious difficulties: Voluntarism can simply press the charge that it is not committed to the second of the supposedly inconsistent claims. As pointed out above, this claim takes the correctness of the retrospective point to imply that I cannot believe that my intentions to believe are potent. The retrospective point is expressed thus in the relevant passage: “With regard to no belief could I ... even suspect that I acquired it at will.”³¹⁴ This quotation, I submit, is most naturally read as expressing the following proposition:

$\forall A \forall p \forall t: A \text{ believes } p \text{ at } t \Rightarrow \square [\forall r: r \text{ is later than } t \Rightarrow \neg(A \text{ suspects at } r \text{ that she acquired the belief that } p \text{ at will})]$, where “A” ranges over the class of all agents, “p”

ranges over the class of all propositions, “t” and “r” range over the class of all points in time, and “□” denotes the conceptual necessity operator.

Now, arguably the above proposition does not cause any trouble for a belief in the potency of intentions to believe: It seems highly implausible that my ability to form a potent intention in a particular case should *necessarily* rely on my beliefs (or “suspicions”) about the *past* potency of such intentions. This, e.g. would rule out any case, where an unconfident agent manages to do something intentionally for the first time, e.g. wiggle her ear. *At the very most*, it would seem that in order to form an intention of a certain type, I have to believe in the *future* potency of intentions of that type, and this even seems a strong concession to make. However, the aforementioned proposition says nothing about *future* intentions at all. Thus, there is hardly any *conceptual* obstacle to the idea that I may optimistically believe that my future intentions to believe at will can be potent and so form such intentions, even if I also believe that no such intentions have been potent so far. Doxastic voluntarism may simply deny the second premise of Williams’ argument and escape absurdity in style.

We may then conclude that Williams’ prospectivist strategy utterly fails. None of his two prospectivist arguments support CIC by way of presenting a convincing argument to the conclusion that intentions to form beliefs are in the least incoherent or impossible to form. Below (Section 8.2.1.6) we shall encounter a more recent version of the prospective argument due to Dion Scott-Kakures that expressly claims to avoid Williams’ errors. However, as shall emerge, this version also presupposes the validity of Williams’ *retrospective* self-defeat claim and so stands and falls with it.

8.2.1.4. Williams’ retrospective incompatibility claim. As we saw above, Williams claims that, if both are held *consciously* by an agent, the following two beliefs are somehow *incompatible*: The belief that p and the belief that the belief that p was acquired at will.³¹⁵ There are several difficulties with this claim as shall be seen below. The greatest difficulty lies in its very nature: Suppose that we grant for the sake of argument that such beliefs are really incompatible to the highest possible degree, say that (contrary to appearances) the conjunction of their propositional contents is necessarily false. Now, perhaps it is somehow impossible consciously to believe a conjunction of propositions while also consciously believing that this conjunction is necessarily false. However, it is surely *not* impossible consciously to hold two beliefs, just because the conjunction of their contents is necessarily false. Expressed in a more formal vein the first claim has it that:

$\Box \neg [\exists A \exists p \exists q: A \text{ consciously believes } (p \ \& \ q) \ \& \ \Box \neg (p \ \& \ q)]$

whereas the second has it that

$\Box \neg [\exists A \exists p \exists q: \Box \neg (p \ \& \ q) \ \& \ A \text{ consciously believes } p \ \& \ A \text{ consciously believes } q],$

where in both cases “A” ranges over the class of all agents, “p” and “q” range over the class of all propositions, and “□” denotes the conceptual impossibility operator.

These are clearly very different claims. The second claim does not seem correct even if the box operator is interpreted as a *psychological* impossibility operator: Surely there are many actual cases where an agent is discovered to have held beliefs whose contents logically contradict each other. Further, *two beliefs cannot really be*

conceptually or logically incompatible in any ordinary sense of these terms. At most they can be *psychologically* incompatible: It is an empirical fact about the human psyche whether it is capable of holding two particular beliefs at the same time, at least as long as the holding of neither belief is conceptually impossible in isolation.

Now, one might think that Williams escapes this problem if relying only on the following “weaker” premise:

$\Box \neg [\exists A \exists p \exists q: \Box \neg (p \ \& \ q) \ \& \ A \text{ consciously believes } (p \ \& \ q)].$

However, this claim runs into similar problems: Believing a necessarily false conjunction does not even seem psychologically impossible: If the class of necessarily false conjunctions contains, e.g. false mathematical statements, the above proposition arguably entails that any agent is necessarily infallible in her beliefs concerning mathematical matters,³¹⁶ an optimistic statement to say the least.

Thus, whatever merit Williams’ retrospective incompatibility claim might have, it can *at most* show the *psychological* impossibility of acquiring beliefs at will. To attribute to Williams’ the view that his retrospective incompatibility claim supports CIC, seems very uncharitable.

8.2.1.5. Doxastic voluntarism and self-deception. All may not be lost for CIC, however. Its proponents, even if forced into modal retreat, might well opt for a weaker impossibility claim than CIC, upholding Williams’ basic point that our inability to acquire beliefs at will is in no way “a contingent fact.”³¹⁷ The strongest possible claim along these lines seems to be the following: $\forall A \forall p \forall t \Box A \text{ believes that } p \text{ at } t \Rightarrow \neg A \text{ acquired the belief that } p \text{ at will}$, where the variables range over the same classes as before and “ \Box ” now denotes the *psychological* impossibility operator.

In other words: It may be argued that no agent with a human psyche may acquire a belief at will at any time.

However, even this weaker impossibility claim is not aided much by Williams’ retrospective incompatibility claim. First, Williams’ incompatibility claim hardly renders beliefs acquired at will psychologically impossible in the first place. As Barbara Winters nicely points out, conceding Williams’ claim would just make the ability to acquire beliefs at will a member of an exotic, but well-known, class of mental abilities:

To see this, consider another example of something I can do in a particular case only if I cannot believe that I have done it in this case. If I have forgotten forever that Bird Day is the last Friday in April, then I cannot believe that I have forgotten forever that Bird Day is the last Friday in April. Nevertheless it is still possible for me to show that I have the capacity to forget things for ever ...³¹⁸

Arguing along this line, Winters, Bennett and Scott-Kakures agree that conceding the truth of Williams’ retrospective claim does not rule out acquiring beliefs at will, this method of belief formation just needs to be *self-deceptive* in the sense that, in each particular instance of application, the agent cannot *after* the application *believe* that she applied the method at that instance.³¹⁹ In other words, rather than attacking directly Williams’ retrospective incompatibility claim, which in the suggested interpretation may be expressed thus:

$\Box\forall Ap$: A acquired the belief that p at will $\Rightarrow \neg(A$ consciously believes that A acquired the belief that p at will), " \Box " denoting global psychological necessity

These authors argue that the truth of this claim does not make it psychologically impossible to acquire beliefs at will. I take this point to be entirely correct. However, relying exclusively on this line of attack against Williams poses at least two severe problems.

First, Williams at most implicitly claims that his retrospective claim ultimately rules out believing at will. Rather he seems to count on his prospective arguments for dealing a knockout blow to doxastic voluntarism.

Second, as pointed out above, the primary appeal of the possibility of doxastic action lies in its promise of underwriting ascriptions of epistemic blameworthiness. Arguably some exotic self-deceptive ability is not of much use in this respect, at least insofar as deontic evaluations of beliefs are taken to be legitimate in fairly commonplace cases. Williams' retrospective point thus cannot be conceded even in its weakened version if an appeal to doxastic action is to play its principal part as a support for epistemic deontologism.

The really interesting question, then, is why the presence of a conscious belief that I formed a certain belief at will should necessarily be a factor influencing the stability of this belief, that is, whether Williams' retrospectivist incompatibility claim is true in its last suggested interpretation. I do not believe that any author has presented an even remotely convincing argument to the conclusion that it is. In fact Williams' claim is vulnerable to obvious counter-examples:

Suppose that I come consciously to believe that one of my most cherished beliefs, say the belief that giraffes have hoofs, I originally acquired at will out of non-truth-oriented motives, say the desire to please my biology teacher in grade school. Now, why should this necessarily make me give up my belief that giraffes have hoofs? Surely this seems preposterous: Even if I came to believe that I originally acquired the belief that giraffes have hoofs at will, I would certainly still believe myself to have lots of good reasons for believing that giraffes have hoofs; I have seen such animals in the zoo, seen nature programs on television, etc. Further these reasons seem entirely unrelated to the original mechanism forming the belief and my awareness of these reasons seems to be enough to sustain my giraffe belief under typical circumstances. Consequently, the claim that an agent's conscious belief that she acquired a belief at will should necessarily make her give up that belief, seems entirely misguided.

At this stage a proponent of Williams' claim could offer the following reply: But if you manage to hold on to your belief that giraffes have hoofs even after you come consciously to believe that you originally acquired this belief at will, certainly you must stick to your giraffe-belief *solely* as a result of your awareness that you have good reasons to hold it. In other words, the fact that (supposedly) you originally acquired the giraffe belief at will has no causal relevance to the fact that you now hold it. Surely, then, your giraffe belief cannot possibly qualify as a belief formed at will in the relevant sense, since you no longer hold it due to your doxastic action of forming it. Thus, even if your counter-example defeats Williams' claim in the reading you have offered, certainly you have not demonstrated that a *willed* belief that, say,

giraffes have hoofs, could ever be held simultaneously with a conscious belief that that belief was produced at will. The stability of *willed* beliefs was the real issue of the debate!

In the vein of this response Dion Scott-Kakures has submitted:

But if after a survey of the evidence, S's decisive epistemic judgment is that she has good and sufficient reason to believe that p, how could she fail *at the very time that she so judges* to believe that p? To say that the judgment is strictly epistemic is to agree that the agent believes that considerations having to do with the truth of p, render it the case that p is more likely to be true than not ... (...) ... [H]er judgment is not merely that the evidence makes it the case that p is more likely to be true than not, it is that p is true. And to affirm that p is true is to affirm that p.³²⁰

However, Scott-Kakures' claim that an agent, who takes herself to have good and sufficient reason to believe that p, necessarily therefore believes that p, is highly implausible to say the least. James Montmarquet has offered the following response:

Isn't it quite easy to imagine cases of persons who recognize that, given the evidence, they should believe that p (that they have sufficient reason to believe that p) but who still cling for epistemically disresponsible motives to the opposite belief? If this were not so, it would be hard to understand all of the recent attention paid by epistemologists to the notion of "epistemic responsibility."³²¹

I take it then that, *a fortiori*, the situation is far from unimaginable where an agent (1) believes herself to have good and sufficient reason to hold a certain belief (2) does not hold that belief (e.g. due to the causal efficacy of emotional factors) and (3) *therefore* intends to form the relevant belief.

In the context of my first counter-example to Williams, if my giraffe belief happens to be a belief which, despite my belief that I have a good and sufficient reason to hold it, I would not have held, had I not brought it about by will, this belief surely must qualify as a *willed* belief in any relevant sense. Remember that Williams' retrospective claim does not deny that forming beliefs at will is globally psychologically possible in the first place; it only imposes a global constraint on the successful performance of such doxastic actions. Thus, even if agents capable of performing such actions are very rare or even non-existent in the actual world, this is of no relevance to the present context, where we are allowed to suppose that such agents exist. In fact we may even suppose that I am such an agent.

Suppose now that I suffer from some childhood trauma involving giraffes, and therefore have a strong emotional barrier towards forming any first-order beliefs about giraffes at all. At some stage I realized that, even though I had any good reason to believe that giraffes have hoofs, I did not at that time believe that giraffes have hoofs. I therefore formed this belief at will. In this scenario, surely my conscious belief that I brought about my giraffe belief at will is not going to affect the stability of that belief. My giraffe belief, then, in this scenario must uncontroversially qualify as a *willed* belief, which, *contra* Williams, is held simultaneously with a conscious belief that it was brought about by will. The suggested response to my first counter-example cannot get a grip here. Williams' retrospective incompatibility claim comes out defeated.

After finally having saved voluntarism from Williams' retrospective claim, we may well wonder what retrospective restrictions on *willed* beliefs may *legitimately*

be imposed. The above considerations have strongly suggested that at least it *typically* needs to be the case that, after the formation of a belief, the agent believes herself to have good reasons for holding this belief. Now perhaps this constraint could be considerably relaxed. One might wonder whether there is really only need for a *negative* retrospective requirement, requiring only that, after the acquisition of the belief that *p*, the agent does not (consciously) believe that she does *not* have a good reason for believing *p*, or even weaker: that the agent does not (consciously) believe that she has a good reason for believing that not-*p*.³²²

To this question regarding the positive versus negative nature of legitimate retrospective requirements on willed beliefs, I can offer only a tentative answer. Arguably genetic voluntarism must be committed to a “cognitive inertia hypothesis” (see Section 8.1): Its project of underwriting deontic evaluations of beliefs falters, unless the belief-states resulting from the performance of doxastic actions are stable to a rather high degree.

Now, in order not to violate this inertia hypothesis, surely the weak version of the negative requirement seems both minimal and legitimate: In typical cases I cannot stably hold on to a belief while consciously believing that I have a good reason for believing its opposite: I can only achieve this in the presence of contingent extra-voluntary belief-influencing factors and surely voluntarism cannot afford to base its account of epistemic deontology on such *doxastic luck*. Imposing the stronger negative requirement on typical cases also seems unavoidable: In fact believing that I have no good reason for holding a belief is often enough to make me give up that belief.

Imposing any positive requirement is a more controversial issue: Is it not possible that I may hold a belief without having any second-order beliefs concerning its epistemic status at all? Even though such *capricious* beliefs are arguably rare in the actual world, I do not think we can rule out their possibility, even in normal agents.³²³ Still, for many (perhaps most) of our beliefs holding them capriciously would seem very odd. This includes, e.g. most beliefs about our recent history and our recent experiences.

The only answer fitting this context then seems to be the following: The voluntarist must concede that, typically, a belief acquired at will must satisfy at least the weaker and stronger negative requirements on the presence of second-order beliefs. Further, depending on the content of the belief and the nature of the agent holding the belief, some beliefs must also satisfy the *weaker* positive requirement that the agent consciously believes that she does not have good reasons to believe its contradictory opposite, or must even satisfy the *stronger* positive requirement that the agent consciously believes that she has good reasons to hold the belief.

If this is correct the voluntarist may well have to concede *some* self-deceptiveness in many cases of belief acquisitions by will: In order to successfully install a belief, if the relevant second-order requirements are not already fulfilled, the agent will need to make sure they are; either by relegating her problematic second-order beliefs to the subconscious or by simply acquiring the relevant second-order beliefs by will in the bargain. One may well fear a vicious regress here, but I think this is no imminent

worry: The psychological higher-order conditions for holding second-order beliefs are arguably of the negative variety: Consciously to hold the belief that I have good reasons for holding a belief, it is arguably enough that I do not believe that I have good reasons for not holding this second-order belief. The regress then stops at the third level: If I have pernicious conscious third-order beliefs (arguably, in normal agents such third-order beliefs are rarely conscious), these must be relegated to the subconscious.

8.2.1.6. *Dion Scott-Kakures' prospective argument.* I mentioned above Scott-Kakures' version of the *prospective* self-defeat argument, which I am now finally in a position to evaluate.

Scott-Kakures' strategy is subtle. He concedes that intentions to believe are coherent (thus avoiding the question-begging involved in Williams' prospective argument) and even concedes that such intentions may in fact *cause* belief-formations. What he denies is that such intentions may perform the causal role that intentions perform in bringing about intentional *actions*:

In fact, it is at just this point that the case for believing at will founders. I will argue that believing at will cannot in principle be distinguished from non-intentional anomalous belief state transitions. Thus, though I may desire to believe that *p*, and in causal consequence may come to believe that *p*, in no case will this count as succeeding in directly willing a belief.³²⁴

Scott-Kakures' reasons for this slightly surprising conclusion are made clear in the following. He here draws upon a wide array of sophisticated theses from the philosophy of action. His central premise is the claim that intentional action is essentially "monitored" and "guided":

Thus, when I intend to raise my arm and I succeed, I know when to stop trying. It is this guiding and monitoring of the agent's by appeal to the content of the intention that serves to distinguish the brute humean cases from the rationalizing causes essential to intentional psychological explanation.³²⁵

Now, Scott-Kakures' argues, would-be intentions to believe cannot assume this "guiding" role due to the necessity of a cognitive "blind spot" or "fissure" in the process of acquiring beliefs at will:

There must be a cognitive blind spot or fissure between *t* and *t*+1. I cannot from my cognitive perspective at *t*, see my way through to my altered cognitive perspective at *t*+1. If this is so then the intention that I formulate at *t* cannot be one which I govern or monitor my behavior through to *t*+1 ... (...) ... my arrival at this belief cannot count as something I succeed in willing directly, as I do when I succeed in directly willing an arm rise.³²⁶

Obviously, here everything hinges on the nature of this "cognitive fissure," which is clarified in the following passage:

In cases in which self-fulfilling belief³²⁷ is not an issue, the relevant aspects of my cognitive perspective at *t*+1 are likely to be quite simple:

I believe *p*

I do not believe anything that rules out my currently believing that *p*; e.g. I do not believe that my belief that *p* is solely sustained by the will; ...³²⁸

By the phrase “my belief is solely sustained by the will” Scott-Kakures seems to mean simply that *before* the formation of the belief I did not consciously believe that I had good reasons for holding the belief (in the contrary case, as seen in Section 8.2.1.5, Scott-Kakures (incorrectly) thinks that the first-order belief would necessarily have formed automatically irrespective of my intentions).

Here then Scott-Kakures is naturally interpreted as endorsing a radical version of Williams’ retrospective claim: In order to hold a belief acquired at will I cannot (even *unconsciously?*) believe that I acquired this belief at will. This interpretation is only strengthened by the observation that Scott-Kakures takes great care in ruling out the possibility that I may acquire at will a belief that already satisfied the relevant second-order requirements prior to its supposed acquisition. However, as seen above Williams’ retrospective claim is simply false: there is not even a serious psychological obstacle to the possibility that I may hold a belief while also consciously believing that I acquired it at will. A “cognitive fissure” in Scott-Kakures’ sense is certainly not required.

Thus, contra Scott-Kakures, if intentions to form beliefs can indeed be causally effective, I can easily “monitor and guide” my belief-forming actions in the sense that “I know when to stop trying”: There is no psychological pressure on me to forget that I had the relevant intention in the first place and I can therefore simply retrospectively check by introspection whether this intention has been effective, about as easily as I can check visually or introspectively whether my arm has risen or not. Consequently, Scott-Kakures’ subtler prospective argument fails and CIC is stripped of any adequate defence in the literature. We will thus have to resort to more modest means if we are to set up a compelling argument against doxastic voluntarism.

8.2.1.7. Heinrich Wansing against the conceptional impossibility claim. One author, Heinrich Wansing, has recently opted for a stronger argument against CIC, claiming to have provided a positive model for so-called “factual voluntarism”; i.e. a claim that at least one already acquired belief was acquired by way of a belief-forming action.³²⁹ His suggested semantic for such a claim ingeniously combines a so-called *dstit* (deliberately seeing to it that) logic for action talk as developed by a.o. F. von Kutschera and J. Horty with a conception of the *implicit* belief of entirely rational and logically omniscient agents as attributable by way of a necessity operator within a certain normal doxastic polymodal logic (polymodal KD45).

Very briefly and untechnically put, Wansing’s proposal turns on evaluating sentences concerning an agent’s doxastic perspective relative to a temporal framework consisting of the relevant moment in time as well as a certain history or time-line. This allows the setting up of a *Choice* function for the agent portioning histories running through a certain moment into sets of histories *choice—equivalent* for the agent at that moment. The intuitive guiding idea is dividing histories of the world according to an agent’s active intervention; by her choice the agent makes sure that the world enters into one of a certain bundle of states rather than another. The agent’s choice is not modelled as a simple choice between incompatible worlds, since there are certainly consequences, even necessary consequences, of an agent’s choices, which

she does not – and sometimes even cannot – choose: By choosing to hold my breath I necessarily confine a certain number of air molecules in my lungs, but this is not something I choose.

Despite their obvious limited applicability to actual human agents who are far from logically omniscient, Wansing's results provide reason for cautious celebration. If one aspires for the rigidity of finitely axiomatizable frameworks, surely Wansing has pointed to an interesting one relative to the discussion of CIC. Nevertheless, Wansing has hardly provided a fully conclusive response to CIC. The main weakness of his approach lies in its focus on the semantic value of *sentences* relative to certain frameworks: A proponent of CIC may well have to grant that given Wansing's idealizations it leads to no logical inconsistency to presume that along one or more routes, which the development of the world might have taken at a prior moment, the sentence "A implicitly believes that p" (or rather: its suitable technical equivalent) is false and that A's Choice function maps that prior moment onto a different bundle of options where "A implicitly believes that p" is true and which contains the history of evaluation. This is the knob Wansing hangs his hat on. Still the proponent of CIC would hold that, given its particular contrast in semantic value across modal contexts, the sentence "A implicitly believes p" would then fail to express anything concerning A's *belief* (be it "implicit" or "explicit"). Hence, given the interpretation in point, the sentence "A deliberately saw to it that A implicitly believes that p" will fail to express anything concerning belief at will. Why? Because on pains of a *contradictio in adjecto* a proponent of CIC simply will not allow that one can in any sense choose to believe p!

To see the force of this point consider that, unless one wants to claim implausibly that all sentences we normally take to express natural laws are true by logical necessity, certainly no inconsistency arises from taking one such sentence and assuming it was made true by some past human agent in the *dstit* sense. Nevertheless, many would still protest wildly if we presented this instance as a model for the choice of *a natural law*, since they would claim that our concept of a natural law is among other things that of a regularity not fixable by human convention or choice: If the regularity in question were really chosen, what we *thought* was a natural law really was no genuine natural law after all. CIC wants to hold something similar concerning beliefs: If it turns out that a certain state of an agent was chosen by her or even *could* have been chosen by her, surely it cannot be a belief, even if we used to think of it as one. The sentences we use to speak of that state hence, if true, cannot speak of her beliefs, they must speak of something else even if the term "belief" should occur in them.

The fact that CIC is committed to this stance, however, brings out its weakness in full. CIC seems committed to a perverse scepticism concerning belief-ascriptions: Each time we attempt to characterize an agent's doxastic state, the mere *logical possibility* that that state was chosen by the agent disqualifies it as a genuine belief. Thus, CIC seems under severe pressure to provide criteria by which we may discern whether an agent could possibly actively have brought herself into a certain state under current investigation. However, such criteria seem excessively difficult to provide. Remember that CIC cannot just point to psychological regularities or even

psychological laws, should such obtain. Such would fall short of establishing the strict kind of impossibility CIC is after, and could only support the weaker *psychological* impossibility claim, which we shall soon investigate below. Thus the proponent of CIC seems saddled with the uncomfortable position that it is fact quite impossible for her to determine, whether a certain belief-ascription is ultimately incorrect due to its being possible in the widest “conceptual” sense that the agent had acquired the seeming “belief” at will. If she wants to ascribe any beliefs at all, her radical resistance to belief by fiat leaves her with a radical commitment to belief-*ascription* by fiat!

This is the closest we will get to a refutation of CIC in the present context. Its proponent may still insist that her semantic insight leaves her no choice (sic!), but to insist that CIC is correct no matter the hardships it may entail. Also she may question the perspicuously ontic interpretation of CIC taken for granted just above, objecting that CIC says nothing about a certain relational property (having not been brought about by choice) which certain doxastic states (beliefs) have by a certain strong sort of necessity, but is rather a claim about the semantic status of certain *dicta*, e.g. “A wilfully brought about her belief that the sun is yellow,” which CIC deem false by logical necessity. In response to this last point we may only stress that if CIC does not concern the histories of our doxastic states, but does only concern certain statements, it is hard to see its relevance to the discussion of how we may in fact control our doxastic states. Stripped of their arguments, persistent proponents of CIC seemingly have nothing left but to clench their fists and stamp their feet. Here I will leave them at that.

8.2.2. *The Psychological Impossibility Claim*

8.2.2.1. *Self-deception revisited.* In this section I will offer my reasons for endorsing the Psychological Impossibility Claim (PIC); the claim that doxastic voluntarism fails due to the fact that actual normal agents with human psyches are incapable of performing doxastic actions.

As demonstrated in Section 8.2.1.5 above, at least some intentional belief acquisitions would need to be self-deceptive to a certain degree in the sense that they would need to be accompanied by acquisitions of second-order beliefs as well as relegations of second- and third-order beliefs to the subconscious. Now, in order accurately to gauge the strength of PIC, it is important to make clear whether doxastic voluntarism needs to appeal to such cases of self-deceptive belief acquisition in order to underwrite ascriptions of epistemic blameworthiness. Clearly, if it needs to rely on appeals to virtuoso self-deceptive patterns of doxastic action, PIC is close to being vindicated for that reason alone.

We may state the important question thus: Does a normal agent’s performance of any doxastic action require her to be self-deceptive, i.e. are there instances where the relevant psychological higher-order requirements for the stability of the resulting belief-state are not already met *prior* to the performance of the doxastic action? As we saw, four types of such higher-order requirements for psychological stability may

be imposed: In order stably to be in the state of believing that *p* an agent may be demanded to satisfy

1. The weaker negative requirement: The agent does not consciously believe that she has good reason to believe that not-*p*.
2. The stronger negative requirement: The agent does not consciously believe that she does not have good reasons to believe that *p*.
3. The weaker positive requirement: The agent consciously believes that she does not have good reasons to believe that not-*p*.
4. The stronger positive requirement: The agent consciously believes that she has good reasons to believe that *p*.

To these requirements may now be added similar requirements for stably *not* being in a state of believing that *p*:

5. The weaker negative requirement: The agent does not consciously believe that she has good reasons to believe that *p*.
6. The stronger negative requirement: The agent does not consciously believe that she does not have good reasons to believe that not-*p*.
7. The weaker positive requirement: The agent consciously believes that she does not have good reasons to believe that *p*.
8. The stronger positive requirement: The agent consciously believes that she has good reason to believe that not-*p*.

Here both sets of requirement are ordered relative to the psychological force of their violation. Violating the requirement 1 typically makes an agent suspend her belief that *p*, violating the requirement 5 typically makes an agent form the belief that *p* “automatically.”

Most likely voluntarism will have to concede at least the weaker negative requirements for stably believing versus not believing a proposition. Only luck may allow an agent to violate these requirements and still be in a stable belief-state versus a stable state of not believing. However, whether or not an agent needs to satisfy any of the last three requirements in order to stably believe versus not believe, plausibly depends on the content of the particular belief and the properties of the particular believing agent. Thus, offering a global answer to the question stated at the beginning of this section seems an immensely complex project.

We therefore need to simplify matters somewhat: I will concede, for the sake of argument, that voluntarism is only committed to the relatively weak general thesis that belief-states versus states of not believing resulting from the performance of doxastic actions must at least satisfy the weaker negative requirements. It remains then to consider whether, in any particular instance of doxastic action relevant to an ascription of epistemic blameworthiness, the weaker negative requirements are fulfilled only by self-deception, i.e. by simultaneously performing the doxastic actions necessary to meet the relevant requirement.

At this stage, voluntarism may attempt the following defensive strategy: Surely no such case may arise, for if the relevant higher-order requirement is not met prior to the

performance of the doxastic action, the agent is *appropriately excused* for performing this action, should it result in an undesirable belief. Thus she is not blameworthy for holding the resultant belief at all.

However this strategy will not cut, as can quickly be seen: Consider how a doxastic voluntarist strategy would account for the epistemic blameworthiness of the educated racist's undesirable belief (see Section 1.2). One strategy would be to claim that the educated racist is blameworthy, because she could and should have performed the doxastic action of preventing the racist belief or at least have suspended it, once it was formed, but regrettably and inexcusably did neither. Now, suppose that the educated racist, after forming the undesirable belief, also believes that she has good reasons to hold the belief: "Of course I have a good reason to believe that blacks are generally inferior to whites," she might say, "the demagogue told me." She thus lacks the psychological prerequisites for stably being in the state she should be in, namely the state of *not* holding the racist belief.

Now, should this be allowed to excuse her from epistemic blame? I think it suffices to recognize here that typically agents believe themselves to have good reasons for what they believe, and very often even consciously so. Thus, if doxastic voluntarism were committed to the claim that such higher-order beliefs appropriately excuse from epistemic blame, it would arguably lose out on the possibility of accounting for anything but very strange cases of epistemic blameworthiness.

In conclusion: Doxastic voluntarism seems committed to the claim that performing two or more doxastic actions (almost) simultaneously is in fact possible for normal human agents. I take this claim to be highly implausible. Already at this stage, doxastic voluntarism is in serious trouble.

8.2.2.2. The psychological impossibility of doxastic action. It may well be possible to deal an even more decisive blow to doxastic voluntarism than was administered above: I shall now argue that even performing a *singular* doxastic action in fact exceeds the powers of a normal agent. This argument may appear most convincing regarding types of belief that seem strongly to force themselves upon us, e.g. empirical beliefs concerning the present state of our perceptual environment. William Alston has offered perhaps the most influential formulation of this point:

Thus, even if I willingly, or not unwillingly, form, for instance perceptual beliefs in the way I do, it by no means follows that I form those beliefs at will, or that I have voluntary control over such belief formation, or that I can be held responsible or blameworthy for doing so. It would have to be true that I have effective voluntary control whether I do or do not believe that the tree has leaves on it when I see a tree with leaves on it just before me in broad daylight with my eyesight working perfectly. And it is perfectly clear that in this situation I have no power at all to refrain from that belief. And so with everything else that seems perfectly obvious to us. We have just as little voluntary control over ordinary beliefs formed by introspection, memory and simple uncontroversial inferences.³³⁰

Arguably, here Alston obscures the difference between the doxastic actions of forming and suspending a belief, but I hold it clear that his argument is not damaged by supplying the necessary corrections: In the above scenario, as little as I can intentionally *suspend* "refrain from" my belief that the tree has leaves on it, I can intentionally

form this belief when confronted by the tree with my eyes open: This belief simply forces itself on me or escapes me: I cannot form, prevent, suspend, or sustain this belief by way of my intentions.

In order to deny Alston's claim here, the voluntarist could reject the first claim of the above quotation. However, this just takes us back to the pseudo-stoic position that, in order to be responsible for holding a belief, it is enough that the agent forms this belief "willingly" or at least: that she does not form it "unwillingly." As seen in Section 7.2, Sharon Ryan has recently advocated a position of this type, arguing that I am responsible for my beliefs, as long as I believe what I "meant to believe."³³¹ However, as argued, this position is not convincing, as, on parity of reasons, it makes the agent responsible for anything that she "meant to happen," be it the state of her health or the current weather situation. In Section 8.3.2.2 below, we shall encounter a more ambitious recent voluntarist attempt to argue that even the most "pressing" of beliefs is in fact the agent's responsibility, since it is the result of the agent's "decision-making."³³² I shall offer reasons to reject this claim in the relevant context.

For now I shall simply say that the more sensible voluntarist line of response is partly to concede Alston's claim: Beliefs like the perceptual belief that a particular near-by tree has leaves, so a voluntarist may argue,³³³ are simply not the kind of beliefs that agents can be blameworthy for holding. Carl Ginet, in arguing for doxastic voluntarism, has expressed this concession thus:

It is fairly clear that in the large mass of beliefs held by any normal person at a given time, the overwhelming major part will have come about involuntarily and only a small portion will have been adopted voluntarily (by decision). Coming to believe by deciding to believe (or seeming to do so) is undoubtedly a rare phenomenon in that sense. But it is nevertheless a phenomenon that we are familiar with – it may happen every day in the life of some of us – and that is why we find it intelligible, when judging whether a particular belief was [deontologically] justified, to think of ourselves as judging whether the believer ought to have chosen to adopt the belief or ought to have done so if she could have done otherwise.³³⁴

Offering this line of response to Alston is a clever move: After all, the most intuitively compelling cases of epistemic blameworthiness (as, e.g. that of the "educated racist") do not concern simple empirical beliefs, but rather beliefs, which an agent, after a process of deliberation, has in some sense *decided* to believe. Still, this concession does not offer the voluntarist shelter for long. Alston continues his above-quoted passage:

The discussion at this point will suggest to the voluntarist that he can still make a stand on propositions that do not seem clearly true or false [to the agent] and hold that there one (often) has the capacity to adopt whatever propositional attitude one chooses. In religion, philosophy and high level scientific matters it is often the case that, so far as I can see, the relevant arguments do not definitively settle the matter one way or the other ... (...) ... So what am I to do? I could just abandon the quest. But alternatively I could, so it seems, simply decide to adopt one of the positions and/or decide to reject one or more of the contenders.³³⁵

Alston now proceeds to reject this last voluntarist suggestion. First, he submits, if during my deliberations I reach a point where one of the available positions seems clearly true or false, and thereafter forms a belief, we are right back to the case of

perceptual belief:

The most obvious suggestion is that although in these cases the supporting considerations are seen as less conclusive, here too the belief follows automatically, without intervention by the will, from the way things seem at the moment to the subject ... (...) ... Thus when our philosopher or religious seeker “decides” to embrace theism or the identity theory, what has happened is that at that moment this position seems more likely to be true, seems to have weightier considerations in its favor, than any envisaged alternative. Hence S is, at that moment, no more able to accept atheism or epiphenomenalism instead, than he would be if theism or the identity theory seemed obviously true.³³⁶

Further, if no such “forcing” situation occurs, the intentions of the agent are still powerless. At most the agent may then “resolve to act as if *p* is true, adopt it as a basis for action,”³³⁷ or engage in an “elaborate pretence of believing.”³³⁸ In my preferred terminology this claim may be expressed thus: At most I can intentionally make myself *accept* a proposition, I can never intentionally believe it.³³⁹ Alston’s most simple and also most convincing argument to this conclusion is a simple first-person challenge to the voluntarist:

I shall merely contend that we are not so constituted as to be able to take up propositional attitudes at will. My argument for this, if it can be called that, simply consists in asking you to consider whether you have any such powers.³⁴⁰

Now, of course this argument only works, insofar as Alston’s reader may take herself to be a normal agent. However, once it is made clear, as above, what forming beliefs at will amounts to (e.g. forming a belief in a case where one does not take oneself to have good rationalizing reasons for holding that belief), I find the argument entirely convincing. I do not think that I can do this, and (from my arm chair³⁴¹) I do not think that I have evidence that anybody else can in fact do this. If a normal agent (akratically) believes against the force of her conscious second-order beliefs, I shall certainly think that this is only possible due to “luck” of some kind; an “arational” factor such as an emotional state must be the cause of her belief.

Further, I believe the same points apply to all types of doxastic action: My beliefs come and go “at the mercy of my evidence,”³⁴² or perhaps more accurately: at the mercy of my mental states. I am not able to form them, prevent them, suspend them or sustain them at will. It does not detract from this point that a doxastic intention to form a belief that *p* might accidentally cause me to form that belief by some strange “deviant” causal mechanism: for example it might be that by some strange means a hypnotist has installed in me a psychological mechanism such that an intention to form the belief that a blue cat is in front of me will “detonate” the belief that a blue cat is in front of me. It is clearly incorrect to say that I performed any kind of doxastic action in this scenario.

I thus take my stand with Alston against doxastic voluntarism: The Psychological Impossibility Claim is true and defeats doxastic voluntarism as a serious attempt of accounting for the legitimacy of actual ascriptions of epistemic blameworthiness: If actual normal agents cannot perform doxastic actions, it is hard to see how an appeal to such actions could be ever be of any use in this respect.

However, I have not endorsed The Conceptual Impossibility Claim, and will remain agnostic on a global Psychological Impossibility Claim: I will not deny the possibility

that certain prodigies³⁴³ or highly trained “mental masters” may in fact perform the doxastic actions that I, as a normal agent, am sadly unable to pursue. However, given the naturalistic requirement imposed in Section 7.2 on any deontologically adequate mode of doxastic control, the existence of a few such prodigies do not disturb my conclusion: Doxastic voluntarism is unable to satisfactorily account for the legitimacy of our standard ascriptions of epistemic blameworthiness. If a society of prodigies or mental masters should eventually be realized, doxastic voluntarism would perhaps then be in a more favourable position. However, I find it unlikely that such a society was ever or should ever be realized, and further I have no idea as to which deontic epistemic norms would be current in such a society.

I shall now turn to considering some of the attempts encountered in the literature of saving doxastic voluntarism from the teeth of The Psychological Impossibility Claim. Some of these attempts have it that voluntarism may somehow concede this claim and still aspire to underwriting our ascriptions of epistemic blameworthiness. I shall argue that this *compliant* strategy fails. Finally I shall face a *defiant* strategy, which simply rejects PIC. The more direct version of this strategy simply purports to identify actual cases of doxastic action. I shall show that these cases are really misidentifications of intentional acceptances, or not cases of intentional action at all. At last I shall face a recent ambitious attempt by Matthias Steup to reject PIC by drawing an analogy with a compatibilist notion of responsibility for actions. I shall argue that, even when leaving the merits of this compatibilist notion out of the equation, Steup does not manage to threaten PIC.

8.3. THE CASE FOR DOXASTIC VOLUNTARISM

8.3.1. *The Compliant Strategy*

The compliant strategy for defending doxastic voluntarism grants to The Psychological Impossibility Claim that in fact normal agents are unable to either form, prevent, suspend or sustain their beliefs by way of their intentions. However, it insists that some form of direct doxastic content control may still underwrite ascriptions of epistemic blameworthiness. I shall consider two types of arguments encountered in the literature employing this strategy. The former, which I shall term “the special analogy argument,” argues that really our intentions to influence our beliefs need not be potent at all in order to uphold doxastic voluntarism. The second, known as “Hobartian Voluntarism,” has it that even though normal agents cannot perform doxastic actions, they may still directly control the content of their beliefs, only this control must be exercised by their “character,” rather than their will. Further, this direct control by way of character is held to be deontologically adequate. I shall offer reasons to reject both types of argument.

8.3.1.1. The special analogy argument. The special analogy argument has come in two versions. The first, more radical, version of the argument has it that our doxastic control is analogous to the control we have over actions that we cannot possibly omit due to our psychological make-up. The second, more moderate, version has it that

our doxastic control is analogous to our control over actions we *do* perform, but do not perform as a direct result of our intentions to perform them. Thus, according to both arguments, doxastic voluntarism may concede Alston's PIC, while still maintaining that the notion of doxastic action may be invoked to underwrite ascriptions of epistemic blameworthiness

Perhaps the most out-spoken defender of the radical version of the special analogy argument has been Sharon Ryan:

I cannot now, given my current evidence, just believe the population of the United States exceeds that of China. That's right. Does that fact show that my belief is not under my control or voluntary or free? Not at all, and this can be seen by considering other actions one would regard as free. There are plenty of things I cannot "just do." Suppose I am driving down the road and I see a mother and her child walking down the road. I see them and ask myself, "should I run them over?" I immediately and freely decide not to and then freely drive on. In fact, in a very important sense, I just can't do something like that ... (...) ... You could give me one hundred million dollars and I could not do it. That's just not something I could do. Does this show that I did not freely decide not to run over the mother and child? Of course not.³⁴⁴

Here, Ryan clearly takes it that, if a belief is "free" or "voluntary" in the sense in which the action of driving past the mother and child in the above example is free and voluntary, the agent may be blameworthy for this belief: She explicitly offers the above considerations as a response to the claim that "punishment for failure to believe seems completely out of line"³⁴⁵ and generally argues against the claim that "We do not have any epistemic obligations."³⁴⁶ Thus she may safely be taken to claim that, even if beliefs are at the mercy of our current evidence (As says roughly PIC), they are still "free" in the sense required for ascriptions of epistemic blameworthiness. A similar argument to the same conclusion has been endorsed by Mathias Steup³⁴⁷ and Bruce Russell.³⁴⁸

Once brought out, the short-comings of the above argument are fairly obvious: Perhaps, under the right circumstances the driver from Ryan's example may be *praiseworthy* for driving on without hitting the mother and child. However, Ryan has vehemently not established a case of *blameworthiness*, where it was psychologically impossible for the agent not to perform a blameworthy action: The driver in Ryan's example is clearly not blameworthy for anything. The example does not appear well chosen for her purposes.

However, this fault of Ryan's example may be remedied by adjusting the example to make the driver's "forced" action a morally undesirable one: Suppose, e.g. that the driver is a devoted nazi and that the mother and child have a distinctive Jewish appearance. Further suppose that, once the driver catches eye of the mother and child, she cannot make herself avoid them, not even if offered any price in the world to spare them. Does not this establish a compelling case of (morally) blameworthy action, performed "freely," even if the agent had no power to omit her action? If so, does not this result establish per analogy that we may be blameworthy for our epistemically undesirable beliefs, even if we had no power to prevent them?

Here it is important to scrutinize what exactly is involved in the claim that the driver could not do otherwise in the type of scenario outlined above. It seems fairly clear that the claim that the driver could not do otherwise, must rest on the claim

that she could not *intend* to do otherwise and that her intention to harm the mother and child caused her to harm them: No matter how much money or other goods the nazi was offered, we supposed, she remained bent on harming the mother and child, and thus proceeded to harm them. On the other hand, if in hitting them she reacted automatically (on a kind of reflex), in the sense that her intentions played no part in the event, it suddenly seems off the mark to say that she was in control of this at all.

Now, given the above the analogy with unpreventable doxastic events breaks down. Given PIC, which is conceded here, an agent's intentions have no part to play in her doxastic events. Accordingly, belief-formations appear more analogous to unintentional reflex reactions than intentional action with "locked" intentions. The radical special analogy argument has not set up the analogy doxastic voluntarism need.

We are then left to consider the moderate version of the special analogy argument, which aims to demonstrate that blameworthiness for beliefs is indeed analogous to our blameworthiness for certain *doings* not brought about directly as a result of our intentions. Various doings have been offered as candidates for filling in this picture. Again, Sharon Ryan has offered perhaps the most explicit version of the argument:

I think doxastic attitudes are typically held intentionally ... (...), but I think it is often not true that one explicitly intends or decides to believe. I think consciously deciding to do x is sufficient, but not necessary for doing x intentionally. Doing something purposefully is necessary for an action to be done intentionally. And one can do something unconsciously, automatically, and purposefully. The typing of each individual letter of most of the sentences of this paper was done intentionally, but automatically and unconsciously.³⁴⁹

However, there seems to be no need for such "purposeful," yet supposedly *non-intended*, intentional doings to be done unconsciously. Michael Stocker has offered a host of further examples:

[T]here are many physical acts, which we do, and pretty much have to, without moral or practical certainty, but in regard to which we are active and responsible. We could think here once again of the explorer finding the mountain range, of an archer hitting the bull's eye at the very limit of the bow's range, finding the cure for a disease, breaking the world's pole vault record, hurting somebody's feelings, starting a war, concluding a peace treaty, creating an aesthetically satisfying sculpture and so on ... (...) ... Here we should note that physical acts which are large-scale, difficult to achieve, and uncertain of success, like those noted above, are better analogues for belief and believing than are extremely simple acts like holding one's breath.³⁵⁰

At least for some of the actions listed above, it seems clear that one could, under appropriate circumstances, be (morally) blameworthy for performing them. For actions such as starting a war or hurting someone's feelings, this is particularly evident. It even seems imaginable (cf. Ryan) that, under bizarre circumstances, one could be morally blameworthy for unconsciously typing an individual letter of a sentence, if by doing this one inflicted some harm, which one was not appropriately excused for inflicting.

Still, I will argue, such cases of blameworthy action cannot serve as analogues for blameworthy belief.³⁵¹

The trick employed by Stocker and Ryan is arguably that of offering cases of actions, which, *under a different description*, are perfectly *intended*, but which under the description given are not so. For example the action of starting a war may, given

the particular circumstances, be re-described as the action of firing a gun or signing a letter in a certain situation. Hurting my dog's feeling may, given the relevant circumstances, be re-described as the action of uttering certain commands. "Unconsciously" typing a letter of a sentence may be re-described as part of the *conscious* action of writing my paper, etc.

As observed in Section 7.3, arguably any intentional action admits of a description under which it is not intended: Even the supposedly "simple" action of holding one's breath may, given the particular circumstances, be re-described as confining some exact number of air molecules in one's lungs, something hardly ever intended by any agent. Now, following Donald Davidson,³⁵² unless an event admits of a re-description under which it is in fact intended, it does not even qualify as an action. As seen, arguably the "doings" offered by Ryan and Stocker as analogues for (formations of?) beliefs, all uncontroversially qualify as actions on this criterion.

Thus, the proponents of the moderate argument must pursue one of the following two strategies in order to set up a satisfactory analogy between physical doings and belief formation: Either they must present a physical doing for which the relevant agent is blameworthy in a sense analogous to that in which she may be blameworthy for certain actions, but which does not admit of a re-description under which it was produced by intentions. Or they must argue that at least some doxastic events (belief-formation, -preventions, etc.) may be given an appropriate re-description under which they are intentional.

Arguably they can do neither. First, the examples of "unconscious" or "large-scale" actions presented by Stocker and Ryan above, are obviously not good candidates for supporting the first strategy. At least to me it is unimaginable what doings could qualify for this job, short of begging the question by declaring, e.g. the supposed doing of belief-formation an appropriate object of blame.

If the second strategy worked, PIC would definitely be embarrassed for having simply ruled out the possibility of doxastic actions only by focusing on descriptions of such actions irrelevant to their status as appropriate objects of blame. However, it is hard to see how this strategy could be successful: As pointed out by William Alston, the problem with forming beliefs at will is not only that I cannot form a belief by intending to form it. The wider problem is that I do not in the least know how to proceed in intentionally forming a particular belief: If asked to produce a belief with a specific content at will, "I simply wouldn't know what button to push."³⁵³

As seen above (Section 7.2) though, this claim of Alston's does not quite hit the mark. Richard Feldman has pointed to an easily accessible "button" we may sometimes push in order to bring about a certain belief: Sometimes we may directly bring about a state of affairs reliably registered by our cognitive system, and thus form a particular belief regarding that state of affairs: In order to form the belief that the light of my office is on, I simply need to flick the switch, thus turning it on. As remarked above, this kind of control is not deontologically adequate, since it fails to produce epistemically undesirable beliefs.

The voluntarist, however, might insist that even the formation of epistemically undesirable beliefs may sometimes qualify as an action: Suppose that, as an effect

of intentionally reading in a book a sentence expressing the ludicrous proposition *p*, I form the epistemically undesirable belief that *p*. Now, cannot my reading this sentence be *appropriately re-described* as my forming an epistemically undesirable belief that *p*? If so, insofar as I read the sentence intentionally (I read it because I willed to read it), must not my formation of the belief *p* then count as an *action* of forming the epistemically undesirable belief that *p*? This line of thinking may attempt to take refuge in what Donald Davidson, following Joel Feinberg, has termed “the accordion effect”:

A man moves his finger, let us say intentionally, thus flicking the switch the switch, causing a light to come on, the room to be illuminated and a prowler to be alerted ... (...) ... Some of these things he did intentionally, some not; beyond the finger movement, intention is irrelevant to the inferences, and even there it is required only in the sense that the movement must be intentional under some description. In brief, once he has done one thing (move a finger), each consequence presents us with a deed; an agent causes what his actions cause.³⁵⁴

In terms of the above, it seems clear that, e.g. Stocker’s example of the action of starting a war, is “uncertain of success,” exactly because the occurrence of this action may be described in terms of an “accordion” stretched out to a high degree and featuring quite complex events: A woman (intentionally) moves her hand, thus moving a pen, thus writing certain letters, thus signing a letter, thus declaring her hostile intentions, thus starting a war; to take one possible accordion-type description of what went on. In Ryan’s example the accordion might yield, e.g.: A woman (Sharon Ryan) intentionally moves her fingers, thus writing a sentence, thus (unconsciously) writing each letter of the sentence. Each “fold” in the accordion yields an appropriate alternative description of the agent’s action: Stocker’s agent moved her fingers, wrote certain letters, and started a war. Ryan’s agent moved her fingers, wrote certain sentences, and (unconsciously) typed certain individual letters. This because, the event of their intentionally moving their body in a certain way did, under an appropriate description, also count as an event of writing certain sentences, starting a war, etc.

Employing the accordion to the example I proposed on behalf of the voluntarist, it certainly seems warranted to say that the agent intentionally reading the sentence in the book thereby *caused herself* to form the belief that *p* and that her causing herself to form the belief that *p* therefore qualifies as a deed; i.e. as something she was doing, just as agent from above causing a war to start by intentionally signing a certain letter, thereby performed the deed of starting a war.

Now, the crucial question is whether a description as a doxastic event of forming the belief that *p* of the event of my reading the sentence expressing *p* in the example above is an appropriate alternative description of that event, or is only an *elliptic* description of this event in terms of its immediate consequence; i.e. whether the event of forming the belief that *p* is only a consequence of the event of my reading the sentence or is truly *identical* to the event of my intentionally reading the sentence, thus qualifying as an intentional action.

Here, I will argue, the former is the case. In the example, of the warring woman, surely the following events are identical: The woman’s signing the letter *and* her causing the war to start (the woman starting the war). As submitted by Davidson,

it seems absurd to say that by signing the letter, the woman caused herself to start the war;³⁵⁵ clearly by signing the letter she did not cause herself to perform further actions or doings, she performed one action: signing the letter (thus starting the war). However, none of the above-mentioned events is identical to the event that the war started. This event is a genuine consequence of the woman's action and not the action itself. To say that the event that the war started was her action, is either an elliptic description of her action or a simple confusion.

Now, in the doxastic case, the above confusion is particularly prone to arise, for here the event parallel to the event of the war starting is the event of the agent forming a belief, an event, which under this natural description, looks conspicuously like an action. The action, however, involved in "accordions" like a woman intentionally reading a sentence, thus forming a belief, is *not* the event of her forming the belief; it is rather the event of the agent *causing herself* to form a belief, being identical to her intentionally reading the sentence. "She started the war" simply means "she caused the war to start," but not "she caused herself to start the war." However, "She formed the belief that p" means neither "she caused the belief that p to form in her" nor "she caused herself to form the belief that p" (the two last phrases being synonymous). Rather, it is identical in meaning to "a belief that p formed in her," a phrase clearly not indicating an action of any kind.

The fact that in the example discussed it seems natural and appropriate to describe the agent's reading of the sentence as her *causing herself* to form the relevant belief, further marks the event of her forming the belief as a non-action. Had it been an action, this description would have been as absurd as describing the woman's action of starting the war as an action of causing herself to perform the further action of starting the war. It should be remarked that forming a belief, remains a non-action (albeit a consequence of other actions), even when it is an *intended* consequence, as in Feldman's example. Even here, all we can do is cause ourselves to form a certain belief (by manipulating our evidence). We cannot perform an action of forming that belief. Our voluntary control over our doxastic events remains indirect, even under the most favourable circumstances.

In conclusion, Ryan and Stocker has provided a fine stock of examples of actions that can only be performed by performing more "basic" actions, i.e. actions, which, under the given description, may appear late in an accordion-like sequence of appropriate action-descriptions. However, this strategy does not help doxastic voluntarism. Doxastic events remain consequences of intentional action, rather than actions proper. This does not mean that actions of causing oneself to form (prevent, suspend, sustain) a belief that p for some proposition p (i.e. a belief with p as its propositional content) are irrelevant to the present wider context. On the contrary, in Chapter 10 I shall examine how such modes of *indirect* doxastic content control may in fact be deontologically adequate within suitably narrow constraints.

8.3.1.2. Hobartian voluntarism. Mark Heller has adopted the label "Hobartian voluntarism" for his position in order to honour the late R.E. Hobart, whom he takes to have held the view that

What we are really judging when we blame S for a bad act is S's character; we are saying that the act reflects badly on S.³⁵⁶

Heller's objective is to develop an epistemic counterpart to this conception of moral blameworthiness. He makes it clear that Hobartian voluntarism is a *compliant* position in the present sense insofar as it expressly concedes Alston's PIC:

I find Alston's simple argument extremely persuasive. I am convinced that the mere willing to believe a proposition would not produce that belief in me.³⁵⁷

However, Heller claims, Alston has not managed to rule out the possibility of *doxastic freedom* in the sense that really matters to the issue of doxastic control:

Even denying effective will, there may still be a sense in which we have influence over our beliefs. We may still have what I call "reflective will" with respect to our beliefs, and reflective will gives us a significant portion of what we want from free will. The reflective will component focuses on the relation between the agent and the action. S has reflective will with respect to just those acts which reflect S's character or nature. The intuition is that free acts are a manifestation of who the agent is.³⁵⁸

And further:

On this Hobartian conception of voluntary belief, we are responsible for most of our beliefs (or at least, if we are not, it is not the voluntarist requirement that stops us from being responsible.)³⁵⁹

Earlier (Section 7.1) we have already encountered one theory of epistemic blameworthiness that explicitly gave up the need for any type of doxastic control, namely that of David Owens. As pointed out, Owens' theory was for that reason implausible: It opened up the possibility of blaming an agent for being of "bad" quality in any imaginable way, be it blaming her for her bad health or her bad appearance. The plausibility of Hobartian voluntarism therefore hinges upon whether Heller can give a convincing account of the kind of direct doxastic control underwriting ascriptions of epistemic blameworthiness on his theory, without thereby opening up the possibility of blaming anyone for any undesirable property they might have. In what follows, I shall argue that he has not given such an account.

Heller offers the following definition of the notion of epistemic nature (or character, he shifts freely between the terms):

So let us adopt this account of epistemic nature:

S's epistemic nature is her desires to form beliefs in accordance with certain dispositions rather than others.³⁶⁰

He then proceeds to make absolutely clear that an agent's nature or character so conceived cannot influence her beliefs by way of intentions or willings:

"Reflective will" is no longer a good title for that conception of freedom, since the reflective process no longer involves willing.³⁶¹

Still, according to Heller, an agent's epistemic nature may influence her belief-formations in a way relevant to evaluations of epistemic blameworthiness:

S's belief is free and voluntary, according to our revised conception of freedom, just in case it reflects her *epistemic* nature, and this occurs just in case S's desires to have certain epistemic dispositions play

the appropriate causal role in producing the belief in question. She is blameworthy when her belief is appropriately caused by her epistemic nature and yet is formed badly. She is blameworthy in this case, precisely because her belief reflects badly on who she is as an epistemic agent. She identifies with bad reasoning processes.³⁶²

Now, this analysis opens up at least three problems. First, it is highly unclear whether normal agents have an epistemic nature in Heller's sense at all. Second, even if I have an "epistemic nature," it is highly unclear how my desires to have certain belief-forming dispositions can in fact causally affect my belief-forming processes, when intentions are expressly ruled out as a causal nexus. Third, it seems that on Heller's account I could equally have, say a "digestive nature," and be blamed for my undesirable gastric processes, if they are in some mysterious way caused by my digestive desires. Thus we seem to be back with the problems inherent in Owens' account of epistemic blameworthiness.

Heller explicitly responds to the third objection:

I do form beliefs the way I do because of my epistemic nature, but, even conceding that I have a digestive nature, I do not form gastric juices the way I do because of my digestive nature.³⁶³

This reply takes us right back to the second objection, which Heller has only succeeded in sharpening: In fact it does not seem less likely that my "digestive nature" may somehow causally influence my digestion than that my "epistemic nature" may somehow influence my belief formation. We are at a loss concerning how to understand the causal connection upon which Heller imposes such heavy deontic burdens. He gives a hint, when he states that an agent may be "unfree" in the Hobartian sense when "her emotions overcome her reason."³⁶⁴ On a natural interpretation, this could mean that the agent is unfree because her emotions causally overrode the force exerted by her awareness of rationalizing reasons. However, *desires* to have certain belief-forming dispositions have no direct role to play in this causal story. After this remark we are only left more mystified concerning Heller's conception of epistemic blameworthiness. Consequently, even if conceding to Heller that normal agents have "epistemic natures" in his sense (thus waiving the first objection), it is incomprehensible how we should even begin to answer the question of the deontological adequacy of a mode of doxastic control by way of epistemic nature.

I therefore conclude that Heller has not presented a convincing compliant strategy on behalf of doxastic voluntarism. That Heller has obvious difficulties in accounting for the way our "epistemic nature" supposedly controls our beliefs is perhaps not surprising, given that arguably Hobartianism has a wrong grip on the relation between the notions of character and blameworthiness from the outset: It simply does not seem to be the case that ascriptions of blameworthiness of any kind imply judgements of character. What we are judging when judging an agent blameworthy is arguably rather her way of *dealing with* her character than her character itself. As submitted by Marcia Baron:

[W]e hold people responsible for dealing with their character flaws rather than treating them as fixed points, as forces that propel them to act this way or that and that simply have a life of their own. The

violent person, the careless person, the lazy person, the impatient person – each may need to set rules for himself so that in a given situation he does not strike another or fail to attend to dangers.³⁶⁵

A similar point arguably applies to the epistemic realm: An agent's cognitive "character" (her cognitive dispositions) gives raise to epistemic blameworthiness if the agent has not done what she should do in order to alter it or does not do what she should do in order to cope with it in her daily doxastic life. Later (Chapter 11) it will be demonstrated how a working notion of epistemic blameworthiness may operate on this basis.

Short of serious compliant alternatives in the literature, I shall proceed to face two defences of doxastic voluntarism that *defiantly* reject PIC.

8.3.2. *The Defiant Strategy*

The defiant strategy for defending doxastic voluntarism rejects PIC: In fact, its proponents argue, belief-formations, belief-suspensions, etc. are just as intentional and free as even the simplest physical action, and epistemic blameworthiness as applying to holdings of beliefs is therefore a notion perfectly parallel to the notion of moral blameworthiness as applying to action consequences.

I shall consider two arguments employing the defiant strategy. The first *ostensive* argument simply aims to point out concrete down-to-earth instances of doxastic action, thus embarrassing PIC's claim that normal agents are unable to perform such actions. The more theoretically ambitious *compatibilist* argument trades on a particular type of theory concerning the freedom of our actions, which renders their freedom *compatible* with the truth of metaphysical determinism. It is then argued that in fact doxastic events like the formation or suspension of a belief satisfy the relevant compatibilist conditions for freedom and control.

I shall offer reasons to reject both versions of the defiant strategy.

8.3.2.1. The ostensive argument. Almost all doxastic voluntarists have offered examples of supposed doxastic actions, but Carl Ginet more than any other author has counted on his examples to constitute a definitive argument against PIC:

Others have thought, though it is conceptually possible, deciding to believe is never psychologically possible and, if it did occur, it would be quite irrational. Against these views I wish to defend the naïve intuition that coming to believe something just by deciding to do so is possible, that it sometimes seems to us that we do this, and that our doing so need not offend against epistemic reason. My hope is to make it plausible that there is a sort of state that counts as a state of believing a proposition, which state is such that it is clear that one could come to be in such a state simply by deciding to do so, and clear, moreover, that such decisions can (though they need not be) be perfectly rational and motivated entirely by one's appraisal of the available evidence and one's general desire that one's beliefs be true.³⁶⁶

Ginet's argument to this conclusion gives up a great deal of territory to PIC, insofar as he does not think that beliefs produced by intentions (decisions)³⁶⁷ is a common phenomenon at all:

It is fairly clear that in the large mass of beliefs held by any normal person at a given time, the overwhelmingly major part will have come about involuntarily and only a small portion will have been adopted voluntarily (by decision).³⁶⁸

Still, Ginet is willing to take a stand on a few infrequent cases and proceeds to offer some supposed examples of beliefs formed by intention. Before going on to consider some of these examples, however, it is crucial to get in the clear about, what Ginet understands by the term “belief.” His basic definition seems to be the following: An agent A believes that p if, and only if, A “count on its being the case” that p.³⁶⁹ Now according to Ginet, “counting on its being the case that” p amounts to the following: “To count on p is to stake something on p with this sort of dismissive or unconcerned or unready attitude towards the possibility of not-p.”³⁷⁰ The import of this phrase is explicated in the following key passage, where a method for forming a belief simply by way of intentions is also suggested:

To not prepare oneself for dealing with the possibility of not-p is to not think about the possibility that not-p or at least not to give any consideration of what to do if not-p. And in fact, since to not prepare for the possibility of not-p is itself to stake something on p (assuming that one believes that not preparing for the possibility of p, is optimal iff p), choosing to not prepare oneself for the possibility of not-p, resisting an impulse to do so, can be the choice such that it is in making it that one decides to believe that p.³⁷¹

This passage raises several questions, not least concerning the nature of the choice “not to prepare oneself for the possibility of not-p” allegedly sufficient for an effective intention to believe that p. It cannot simply be that this is a choice not to *believe* that not-p, for in that case we are right back to the beginning of the question of the possibility of doxastic action, and nothing would have been explained. Aside from that, intending not to believe that not-p surely does not amount to intending to believe that p; the former intention obviously has success criteria different from those of the latter: I do not at present hold the belief that the mathematical proposition known as Goldbach’s Conjecture is false, but neither do I hold the belief that it is true. Even though I am acquainted with this proposition, I do not have a degree of conviction in either its truth or its falsity sufficient to establish a state of belief. The same could be the case, I submit, for a wide number of propositions.

Consequently, “not preparing myself for the possibility of not-p” must mean something different from believing that not-p, if it is to have the power suggested by Ginet. He clarifies the import of this crucial phrase in the following passage:

In the right circumstances, it can take effort to avoid preparing oneself for the possibility that not-p. Such efforts, to suppress considering that possibility and what to do if it is realized, are what make it apt to sometime describe oneself as making oneself believe (or trying to make oneself believe) something.³⁷²

In this passage Ginet’s talk about “considering a possibility” gets somewhat clarified: “Considering a possibility” is something an agent can actively suppress. From this it seems warranted to analyse Ginet’s “considering the possibility that” as meaning simply “giving conscious thought to the possibility that.” As we shall see below, this reading is also strongly supported by his choice of examples.

At this stage then, it seems safe to conclude that Ginet operates with the following basic analysis of the notion of belief: An agent A believes that p if, and only if, A does not give conscious thought to the possibility that not-p and A does not consciously

think about what to do if not-*p* turns out true. Call a mental state that satisfies these conditions a “Ginet-belief.”

In Section 2.1 I have argued that a belief that *p* is best conceived of as a state of conviction that *p* is true, and that this is arguably the sense of “belief” commonly employed in epistemological contexts. It was argued that the strength of this conviction must be stronger than a mere hunch, but need not be so strong as to issue in knowledge claims. I shall now argue that even this fairly vague definition of belief is still strong enough to show the following: Being a Ginet-belief is neither sufficient nor necessary for being a belief.

Sufficiency first: The most obvious line of argument here is to simply observe that there are a vast number of propositions that an agent does not give the least thought about without therefore believing their opposite. For example typical agents do not give any thought at all to the possibility that Goldbach’s Conjecture is false, or what to do if it should once be proven false. Yet these typical agents do not believe Goldbach’s conjecture either. However, faced with this objection, Ginet would probably just supplement the conditions for Ginet-belief with the condition, that (the truth of) *p* must be something that the agent cares about to a certain degree. At least this seems vaguely suggested by his earlier suggestions that an agent believing that *p*, thereby “stakes something on” *p*.³⁷³

However, this move hardly saves the game: Consider now a mathematician, who cares intensely about Goldbach’s Conjecture, in fact he fancies the idea of proving it some day in order to achieve everlasting mathematical fame. However, he also believes that the evidence for and against the truth of the Conjecture is equally balanced and pessimistically also believes that, since most likely no one will ever be able to prove or disprove the Conjecture, his verdict about the evidence is not likely to change. Now, surely this mathematician need not believe that the Conjecture is true; certainly he is not convinced of its truth to any degree. Still, he also does not give any conscious thought to the idea that the Conjecture is false, and he does not prepare himself for the day it will be disproved. It would seem then, that even though he Ginet-believes the Conjecture he does not believe it. Ginet-belief is not a sufficient condition for belief. Consequently, even if it is the case that an agent can intentionally bring herself into a state of Ginet-belief, it has not been shown that she can intentionally bring herself into a state of belief.

Still worse for Ginet’s purposes, Ginet-belief is not even a necessary condition for belief: Consider, e.g. a very conscientious parent, who believes that her daughter is out playing at the playground just outside the house (We might, e.g. suppose that she just visually observed the daughter less than a minute ago). Naturally she cares about the fact that her daughter is out playing. However, this parent also worries intensely about her child and constantly gives conscious thought to the possibility that her daughter is kidnapped. She even has an elaborate plan for her behaviour if this should happen. Now, this parent does not Ginet-believe that her daughter is playing at the playground (she gives too much conscious thought to other options), yet she certainly believes this proposition (what else should she believe, given the

evidence). This shows that forming a Ginet-belief is not even a necessary stage on the way to forming a belief.

Ginet-belief, then, is a very different mental state from belief. It has turned out to be more like what is normally referred to as “make-believe” or “an elaborate (self-)pretence of believing,” to borrow a phrase from William Alston.³⁷⁴ The following example offered by Ginet only consolidates this impression:

S receives a telephone call from the police, saying that his wife has been involved in a car accident and that she wishes him to come to the scene. The police caller says that she believes there are no serious injuries. While S hastens to the scene, S is, as it seems natural to put it, making himself believe that his wife has not been seriously injured. Part of what S does that seems to deserve that description is to suppress all impulse to imagine what his wife injuries might be or to consider how he will handle it if he does find her seriously injured. (Another part is his repeatedly reminding himself of the evidence he has that she had not been seriously injured, namely, that the police caller would very likely have known and told him if there had been serious injuries.)³⁷⁵

Here it seems pretty clear that the man only needs to go through all the mental exercises, because he is far from *convinced* that his wife is not seriously injured, that is: because he does not *believe* that his wife is unhurt. However, it seems equally clear that the man’s exercises may at most result in a state of Ginet-belief, a state in which, with great strain, he forces himself to reason, *as if* he was absolutely certain that she was not injured. If a belief is established by acting out this pretence, it can only be by sheer accident. Ginet’s example has hardly established a convincing counter-example to PIC.

However, Ginet also offer examples, which (contrary to what he seems to think) are of a different nature from the above. As these examples are very similar, I shall take the following as paradigmatic:

Sam is on a jury deliberating whether to find the defendant guilty as charged; if certain statements of a certain witness in the trial are true, then the defendant cannot have done what he is charged with; Sam deliberates whether to believe those statements, to believe the prosecutor’s insinuations that the witness lied, or to withhold belief on the matter altogether. He decides to believe the witness and votes to acquit.³⁷⁶

Now, in contrast to what was the case in the example of the anxious husband, certainly there is a legitimate sense here in which we might say that Sam *decided to believe* a proposition, namely the proposition expressed by the testimony of the defence witness. The question now becomes, whether a decision in the relevant sense essentially involves an intention to believe. I shall argue that in fact it does not.

This becomes clear once the psychological story of Sam’s deliberation, decision and belief-formation is told in greater detail. What happened could be this: At first Sam believed neither the proposition *p* expressed by the defence testimony nor its opposite not-*p*. However he believed that he had strong *instrumental* reasons to soon reach a state where he either believed *p* or believed not-*p*. This was most likely the case, because he believed (1) that he was appointed as a juror in the case, (2) that the truth-value of *p* should guide his verdict, and (3) that he stood under an obligation to reach a verdict. Now, as a consequence of his belief that he had an *instrumental* reason to form a belief concerning the guilt of the defendant, he *intentionally* started deliberating about whether *p*: He attempted to call to mind the witness’ testifying act

in the hope that there were important aspects he might previously not have noticed or given due weight, etc. Now at some point of this deliberation process, Sam reached a point where he consciously believed that he had good *rationalizing* reasons to believe that p. As Sam is a normal agent, and no relevant external belief-causing factors were active, this second-order belief induced in him the belief that p: He came to believe that the defence witness spoke the truth.

Now, the important thing to notice is that Sam's *intentions* only entered into this story at one spot: The time where, as a result of his awareness of his instrumental reasons for reaching a verdict, he formed the intention "decided" to deliberate about whether p. Nowhere did an intention to believe that p enter the stage. That his deliberations resulted in the belief that he had good rationalizing reasons for believing that p was only up to Sam in a weak sense: All he could do was keeping the process going, until some result or other emerged. However, what was *entirely* outside of his control was the empirical psychological fact that his belief that he had good rationalizing reasons to believe that p induced in him the belief that p. If, e.g. he had been under a severe emotional pressure, his emotions might well have upset this causal process.

Yet, even though his intentions out-played their role in the belief-forming process before it even resulted in *any* belief at all, we still say that Sam "decided" to believe that p. What can this mean?

I believe an analogue from the practical realm will help clarify the intricacies involved here: Suppose I form a conscious belief that I have good moral reasons for performing a particular action: I simply come consciously to believe that, among the options available to me, performing that action is the morally best thing to do. In that case, we might legitimately say that I have "decided" for that action: I have "picked out" that action as the best candidate for moral action. Call such a decision a "value-decision."

Now, a value-decision that a particular action is the best option from a moral perspective is far short of actually *deciding* to perform the relevant action in the sense of *intending* to perform it: If I am a lazy person I might well come to recognize a certain action as the morally superior option, yet not perform it, because I also believe that performing it would cause me a great deal of trouble. Arguably, the typical adult Western European, who does not give as much money to charity as she believes is morally optimal, fits this description perfectly: This agent has made a value-decision in favour of supporting charity, yet she has made no *effective decision* (i.e. formed no intention) to actually support charity.

The crucial point to notice here is the following: In the moral case, an effective decision is needed on top of a value-decision in order to actually result in an action.³⁷⁷ On the contrary, in the doxastic case value-decisions are causally *sufficient* for belief, at least in normal agents under typical circumstances. There is simply no causal work left for doxastic intentions to do.

In a suitably qualified sense then, it is perfectly true that Sam *decided* to believe the testimony: His deliberation resulted in the epistemic *value-decision* that the belief that the testimony was true was indeed the belief best supported by the available evidence. However, this value-decision was only up to Sam in a weak sense and the fact that his

value-decision resulted in a belief was not up to him at all: Nowhere in the process leading to the formation of Sam's belief that the testimony was true, did an intention to form that particular belief play any role at all.

I take it, then, that Ginet is simply guilty of confusing value-decisions with effective decisions (intentions). His court room example has not presented a convincing case against PIC: As seen above, PIC has no quarrel with value-decisions, rather it needs their causal powers to make its point. The failure of Ginet's courtroom example is highly important since parallel examples involving "decisions to believe" following upon a process of deliberation are spread out over the entire voluntarist literature. As seen, voluntarists may well have fallen victim to a collective conceptual confusion or a fit of wishful thinking here. In the following paragraph, I shall consider a particularly elaborate version of this confusion.

8.3.2.2. The general analogy argument. Over the last decades Matthias Steup has been perhaps the most persistent proponent of doxastic voluntarism, as well as the staunchest opponent of Alston's PIC. Steup's strategy for vindicating doxastic voluntarism is *defiant* in a much more radical sense than, e.g. Ginet's ostensive argument from above. Steup argues that a completely *general* analogy obtains between our responsibility for our actions and our responsibility for doxastic events. According to Steup then, even the seemingly most involuntary instances of belief-formation, as those involving beliefs about our present perceptual environment, are really appropriate objects of blame and praise to the same extent as are our deliberate actions. As shall emerge later below, in a sense I agree partly to Steup's conclusions: There is no reason in principle why an agent could not be blamed for, e.g. her holding of the perceptual belief that a leafy tree is now in front of her. However, even if endorsing Steup's general aim in this regard, I have severe reservations concerning his voluntarist strategy for defending it.

Steup's general analogy argument starts by delineating three different theories about voluntary control over actions, two of them compatibilist with respect to metaphysical determinism and one libertarianist, i.e. hostile to compatibilism. The first theory has it that voluntary control over actions is really nothing but

Hypothetical Voluntary Control

I have hypothetical voluntary control over doing x if, and only if, I can do x if I decide to do x, and I can refrain from doing x if I decide not to do x.³⁷⁸

This theory Steup attributes to such venerable authors as Hobbes, Locke, and Hume.³⁷⁹ Despite its historical provenance, though, this theory has obvious shortcomings, which Steup is quick to point out: An agent performing an action due to hypnosis or another mechanism that "locks" her practical decisions may still be acting voluntarily on the above hypothetical account. This is because, even if the agent could not decide (intend) to do otherwise due to the hypnosis, it may still be the case that, *had* she decided otherwise, she would have acted otherwise. This result seems immediately implausible. Thus, a more "refined" compatibilist theory of voluntary control may opt for the following more restrictive conception of voluntary action:

Deep Voluntary Control

I have deep voluntary control over doing x if, and only if (i) I can do x if I decide to x , and I can refrain from doing x if I decide not to x , and (ii) with regard to deciding whether or not to x , I enjoy deep internal control.³⁸⁰

Steup does not clarify what is meant by “deep internal control” except from a remark that “... sophisticated compatibilists insist on a deeper account of free action: one that imposes, in addition to hypothetical control internal conditions to ensure that free action is incompatible with brainwashing, hypnosis, manipulation, paranoia, and the like.”³⁸¹

Despite the openness of his account, Steup opts for “deep voluntary control” as the decisive compatibilist account of voluntary action. However, he also recognizes that some authors have felt discontented with compatibilism and argued that a satisfactory account of voluntary action must demand that an alternative action was *actually*, not just hypothetically, available to the agent. Steup presents the following libertarianist account of voluntary action as meeting these specifications:

Categorical Voluntary Control

I have categorical voluntary control over doing x if, and only if, (i) I can decide to do x and decide to refrain from doing x ; and (ii) I can do x if I decide to do x , and I can refrain from doing x if I decide to refrain from doing x .³⁸²

Making the analogy to the doxastic context, Steup can now present two serious candidates for a theory of “voluntary control over doxastic attitudes.” The first of these theories is compatibilist, the second libertarianist in character:

Deep Voluntary Control over Doxastic Attitudes

I have deep voluntary control over my doxastic attitude toward p if, and only if (i) were I to decide to take an alternative attitude toward p I could take that attitude, and (ii) with regard to deciding which doxastic attitude to take towards p , I enjoy deep internal control.³⁸³

And

Categorical Voluntary Control over Doxastic Attitudes

I have categorical voluntary control over my doxastic attitude toward p if, and only if, (i) I can decide to take an alternative doxastic attitude toward p ; (ii) if I decide to take an alternative doxastic attitude toward p , I can take that attitude.³⁸⁴

Steup proceeds to offer an extensive argument to the dual conclusion, that (1) “[C]onstruing the concept of voluntary control as does refined compatibilism leads to the consequence that we enjoy voluntary control over our doxastic attitudes, and do so to a large extent,”³⁸⁵ and (2) “If the concept of categorical voluntary control can be coherently worked out, and if determinism is false, then categorical voluntary control over doxastic attitudes is possible, albeit rather constrained in its extent.”

In the run of this argument, Steup calls upon the resources of both the ostensive argument and the radical special analogy argument, both of which have been debunked above.³⁸⁶ However, I shall not consider the details of Steup’s argument here. Rather I shall press one overriding objection: Steup is guilty of a basic equivocation of homonyms, he simply uses the term “decision” in two radically different, but equally

legitimate, senses in his accounts of voluntary action and voluntary control over doxastic attitudes.

Kant has been one of many philosophers to notice the striking (and deceptive) similarity between talk of “decision” “Urteilen” in doxastic and practical contexts.³⁸⁷ However once the standard meaning of the term “decision” in these contexts is subjected to a closer examination, the differences become more obvious than the similarities. In fact, once these differences are brought out, the seeming similarity between the respective accounts of doxastic and practical freedom offered by Steup completely vanish, and the analogy between his proposed notions of voluntary control over actions and voluntary control over doxastic attitudes breaks down entirely.

As we saw in the preceding paragraph (Section 8.3.2.1) the phrase “deciding for an action,” may mean two things: (1) forming the belief that an action is the (morally) best candidate for action under the given circumstances. (2) Actually intending to perform that action. I called the first kind of practical decision a “value-decision” and the second an “effective decision.” It was argued, that practical value-decisions are in themselves causally insufficient to bring about the performance of an action and need to be supplemented by relevant effective decisions in order to achieve practical results.

In the epistemological case, however, things are very different: Here the “value-decision” of holding the belief that forming a particular belief is, epistemically speaking, the most desirable option under the given circumstances (in the sense that, under the given circumstances, the agent consciously believes herself to have better rationalizing reasons for believing the propositional content of that particular belief than for believing its negation), is under typical circumstances *causally sufficient* to bring about the formation of that particular belief. Further, as argued at length above (see Section 8.2.2.2), in normal agents doxastic intentions are *causally inert*. If her epistemic value-decisions do not in a particular case bring about a coveted belief, there is nothing a normal agent can do by way of effective decisions to change that fact.

Now, if Steup’s analogies are to carry any weight he must of course use the term “decision” in the same sense in the practical and doxastic definitions. Given the distinctions re-enforced above, he must then either refer to value-decisions or effective decisions (intentions) throughout. I shall proceed to examine both possibilities and argue that none of them will work.

Let us begin by substituting “intention” (i.e. effective decision) for “decision” throughout Steup’s definitions. The definitions of *Deep Voluntary Doxastic Control over Doxastic Attitudes* and *Categorical Voluntary Control over Doxastic Attitudes* then come out thus:

Deep Voluntary Control over Doxastic Attitudes

I have deep voluntary control over my doxastic attitude toward p if, and only if (i) were I to intend to take an alternative attitude toward p I could take that attitude, and (ii) with regard to intending which doxastic attitude to take towards p , I enjoy deep internal control.

Categorical Voluntary Control over Doxastic Attitudes

I have categorical voluntary control over my doxastic attitude toward p if, and only if, (i) I can intend to take an alternative doxastic attitude toward p ; (ii) if I intend to take an alternative doxastic attitude toward p , I can take that attitude.

On this interpretation it is clear that normal agents do neither enjoy deep voluntary nor categorical voluntary control over their beliefs, since, as argued at length above, in normal agents intentions to form, prevent, suspend or sustain a belief are causally inert: even under the most favourable circumstances, it is simply not the case that a normal agent may take up an alternative doxastic attitude simply by intending to do so, which is required on both definitions. Thus, on this reading, Steup fails to demonstrate the possibility of any form of doxastic control at all.

We must then resort to the second option and read “value-decision” for “decision” throughout the definitions. This also seems to be most consonant with remarks by Steup like:

Once we have judged that our evidence supports believing that p , we believe that p automatically and instantaneously. No effort is required, and we certainly are not aware of any act such as “executing” our decision. However, these differences [relative to the practical realm] do not undermine my main point: that one has executed a decision to φ if one has *because* one has concluded that one’s reasons or one’s evidence supports φ -ing.³⁸⁸

However, when we substitute “value-decision” for “decision” throughout, we run into problems with the *practical* definitions, which now come to read:

Deep Voluntary Control

I have deep voluntary control over doing x if, and only if (i) I can do x if I form the belief that x is the most desirable thing to do, and I can refrain from doing x if I form the belief that not doing x is the most desirable thing to do, and (ii) with regard to forming a belief that x is the most desirable thing to do or forming a belief that refraining from x is the most desirable thing to do, I enjoy deep internal control.

Categorical Voluntary Control

I have categorical voluntary control over doing x if, and only if, (i) I can form the belief that x is the most desirable thing to do and form the belief that refraining from doing x is the most desirable thing to do; and (ii) I can do x if I form the belief that x is the most desirable thing to do, and I can refrain from doing x if I form the belief that refraining from doing x is the most desirable thing to do.

It is unlikely that these definitions are going to appeal to any compatibilists or libertarians concerning the voluntariness of actions. First, as noted above, forming the belief that the action x is the most desirable thing to do, is causally insufficient to bring about the performance of x : There are many cases, where I hold it most desirable to perform a certain action, yet is too lazy or irrational actually to perform the action, and thus form no intention to actually perform it.

Thus, the condition (i) of the compatibilist analysis of voluntary control is hardly satisfied, unless the compatibilist is going to appeal to intentions *motivated* by value-decisions. This appeal, however, would immediately ruin the analogy with the doxastic definitions that cannot appeal to causally efficacious intentions. For the libertarian things look even worse. Short of begging the question against PIC, there is no way in which I could be sure to be able to form both of the beliefs now figuring in the condition (i) of the libertarianist analysis of voluntary control: Only effective doxastic intentions could secure that. Thus, Steup has hardly presented any compelling

analogy on this reading either: The accounts of action freedom needed to establish his analogy are simply too absurd to merit serious attention.

It seems, then, that any intuitive force Steup's general analogy argument might command, simply rests on not noticing the shift between value-decisions (in the doxastic definitions) and effective decisions (in the practical definition). The only way out of this predicament seems to be the following: Steup must argue that practical intentions are in fact somehow equivalent to doxastic *value-decisions*. No hint of such an argument can be found anywhere in Steup's writings. I shall proceed to argue that in fact no convincing argument can be set up to yield the desired conclusion.

The argument would need to establish that effective intentions to act are in fact nothing but beliefs that performing an action is the most desirable option from a certain perspective. Furthermore, in order to establish a satisfactory analogy, this perspective has convincingly to translate into the doxastic idiom as an epistemic perspective. Notice that it is not enough to argue on Steup's behalf that an intention is simply a belief that a certain action is overall the most desirable thing to do in a certain situation. This conception would translate poorly into the doxastic context where value-decisions of the causally efficacious kind that Steup need, narrowly concern a belief's desirability from an epistemic, i.e. truth-goal-oriented, perspective. I think it is highly unclear what dimension of practical desirability mirrors epistemic desirability in a degree sufficient to yield the practical-epistemic analogy that Steup is after. For the sake of argument, I shall therefore go with a very broad conception of desirability, and say that the relevant practical perspective is carved out simply by the agent's practical reasons, leaving unspecified what kind of reasons are relevant.

On behalf of Steup I therefore advance the following thesis:

S-intention: Agent A intends to perform action α if, and only if A believes that, of the options she is aware of, performing α is the option best supported over-all by her practical reasons.

As this thesis construes intentions as a kind of desirability judgments it belongs squarely within the class of theories about the nature of intentions generally known as "belief-desire" theories. Such theories have in common the claim that intentions essentially involve "pro-attitudes" such as desires or desirability judgments.³⁸⁹

There are several general problems with belief-desire theories of intention. Arguably, e.g. they miss out on the "planning" aspect of intentions, the aspect that "our intentions concerning our future actions are typically elements in larger plans."³⁹⁰ This is a bill which a humble desire can hardly fill. I shall not pursue this influential line of criticism here as, even granted the general adequacy of the belief-desire view of intention, the suggested thesis does not establish the analogy doxastic voluntarism needs.

Belief-desire theories revolve on the claim that a suitable complex of an agent's beliefs and pro-attitudes (as e.g. desires) may somehow *explain* why that agent performed a certain action;³⁹¹ just the explanatory work, which an appeal to the agent's intentions is generally thought to do. Now, perhaps a reasonability judgment as involved in S-intention may play the "desire role" in such an explanation of action. S-intention, however, has left the *belief*-role vacant. This is not a matter to be taken

lightly as the following example will show: Suppose some agent, Tom, is walking down the street. We now learn that Tom believes that over-all he has better reasons to buy a birthday present for his mother, than for performing any alternative to that action he can think of. Clearly nothing is explained yet. If however we further learn that Tom believes that, *in order to* reach a place where he can buy a birthday present for his mother, he must walk down the street; his action of walking down the street is at once explained.

Now, the problem for S-intention is that it cannot afford to supplement its account of intention with a belief-clause, as this would immediately destroy the analogy with the doxastic realm where reasonability judgments are in themselves causally effective, hence need not be supplemented by further doxastic states in order to explain belief formations.

A proponent of Steup's analogy argument might now want to point out that there are instances of action explanation by way of beliefs and desires where the belief-component comes almost *for free*: Suppose for instance that some man, Adam, raises his arm. We now learn that, just before raising his arm, Adam believed himself to have better reasons for raising his arm than performing any relevant alternative. Now in order fully to explain Adam's raising his arm on a belief-desire model, we need only appeal to his belief in the almost tautological proposition that in order successfully to raise his arm he must just raise his arm. Since it is unimaginable that a normal agent does not believe this proposition, it seems we may safely assume that Adam believes it. However, this assumption, even if trivial, is still indispensable: If Adam somehow believed that in order to raise his arm he should rather try to influence it by some telekinetic mental action (he, e.g. crazily believes that his arm is separated from his body and cannot be moved in the normal way), his arm-raising is not yet explained.

On this line of argument then, belief-formations may still be voluntary in the same sense as certain primitive physical movements that an agent can *just do* without doing them in order to do something else: In these cases, the voluntarist may claim, S-intention yields exactly the analogy doxastic voluntarism needs. She may then simply argue that a causal explanation of a belief-formation by appeal to an epistemic desirability judgment also *tacitly* presupposes that the agent holds a trivial belief with a relevant content of the type: In order to form the belief that p I must just form the belief that p.

I find this line of response highly strained. Surely there seems to be no need to appeal to such "trivial" beliefs in offering a causal explanation of a belief formation: Suppose that some crazy agent somehow or other manages not to believe the relevant tautological proposition. Assuming that such an agent is even normal (perhaps he could not be), would this matter to our explanations of his belief-formations? Surely not! The causal efficacy of his reasonability judgment that he has good rationalizing reasons to believe a certain proposition is hardly affected by his holding or not holding of further beliefs about doxastic action (as on PIC he cannot even perform such action).

This shows that S-intention, even if not ruled out on account of its immediate implausibility, cannot establish a convincing analogy between actions and doxastic events. As a consequence, there is no way Steup can avoid the charge of equivocating

crucially different senses of the term “decision” in stating his general analogy argument. The analogy simply falls apart, once this is recognized. Equally, the ostensive argument (see Section 8.3.2.1) may now be rejected on a more general level: It simply loses out on any possibility of getting intentions to play a further role in a process of belief-formation involving deliberation: The epistemic desirability judgements (or “epistemic value-decisions”) that typically crystallize as a result of deliberation, do *not* require the presence of doxastic intentions in order to be causally efficacious. Thus, once ostensive arguments appeal to such epistemic value-decisions, the case for efficacious doxastic intentions is irredeemably lost.

This concludes my dismissal of the case for doxastic voluntarism. I hope to have shown that as yet no author has succeeded in countering Alston’s psychological impossibility claim in an even remotely convincing way. Thus, I uphold the conclusion that doxastic voluntarism fails to legitimately account for any ascriptions of epistemic blameworthiness, because no normal agent enjoys the direct doxastic content control that doxastic voluntarists need in order to make sense of evaluations of epistemic blameworthiness by reference to an agent’s performances or omissions of doxastic actions.

DIRECT PROPERTY-DIRECTED DOXASTIC CONTROL
OR PROPERTY VOLUNTARISM

It is not that one has a choice in the beliefs that one forms, but that one has a say in the procedures one undertakes that lead to their formation. The notion of “epistemic responsibility” attaches to the undertaking of appropriate procedures.

John Heil³⁹²

Abstract. Following the demise of doxastic voluntarism in Chapter 8, I proceed to discuss whether modes of direct doxastic control are in fact relevant at all to evaluations of epistemic blameworthiness. I argue that they may in fact be so within a limited domain of cases, only the relevant modes of direct control need to be directed against the normatively significant properties of beliefs, rather than their particular contents. In Sections 9.2 and 9.3 I present two authors, John Heil and James Montmarquet, who have made modes of direct *property-directed* doxastic control the backbone of their conceptions of deontic epistemic normativity. In Section 9.4 I then argue that, although these conceptions are incomparably better off than doxastic voluntarism in accounting for actual cases of epistemic blameworthiness, they still fall short of qualifying as globally adequate accounts.

9.1. THE IMPORT OF DOXASTIC PROPERTY VOLUNTARISM

Doxastic voluntarism failed. There simply is no such thing as a deontologically adequate mode of direct content-directed doxastic control. I shall now proceed to examine whether a mode of direct *property-directed* doxastic control (or “direct doxastic property control”) may aspire to deontological adequacy.

As argued above (Section 7.3), the deontological adequacy of a mode of *abortive* doxastic *property* control would most likely demand the deontological adequacy of a corresponding mode of abortive doxastic *content* control: There simply is no imaginable way in which I may suspend or sustain all properties of an undesirable belief relevant to its epistemic undesirability without suspending or sustaining the belief altogether. Now, as seen above, there is no such thing as a deontologically adequate mode of direct abortive content control; this possibility was excluded by the demise of doxastic voluntarism. Thus, I can safely restrict my attention to the dual modes of positive and negative direct *genetic* property-directed doxastic control, i.e. the *direct* control an agent may enjoy over the instantiation or non-instantiation of properties of her epistemically undesirable beliefs relevant to those beliefs’ level of epistemic undesirability.

Consequently there are thus two ways in which a proponent of the deontological adequacy of a mode of direct doxastic property control may hope to account for an agent’s epistemic blameworthiness for holding an epistemically undesirable belief:

1. The agent is epistemically blameworthy, if she holds an epistemically undesirable belief, because she directly brought it about that an epistemically undesirable

belief (rather than, say, an epistemically desirable belief) was formed and she has no appropriate excuse for doing this.

2. The agent is epistemically blameworthy, if she holds an epistemically undesirable belief, because she omitted to prevent that an epistemically undesirable belief was formed, and has no appropriate excuse for that omission.

Since no term is current for a position that aims to account for evaluations of epistemic blameworthiness in any of the two above ways, I shall baptize such a position “doxastic property voluntarism” or simply “property voluntarism,” to contrast it with the standard content-oriented voluntarist position, which properly ought then to be re-labelled “doxastic content voluntarism.” As, however, the term “doxastic voluntarism” has been the label traditionally assigned to the latter type of position, as above I shall remain in line with the terminological tradition.

Property voluntarism does not entail doxastic voluntarism³⁹³ and is not immediately affected by its demise: It does not follow from the fact that an agent could directly bring it about that an epistemically undesirable belief was formed at a particular instance, that she could also at that instance perform the doxastic action of forming the *particular* undesirable belief that she actually formed. Neither does it follow from the fact that an agent could directly prevent that an epistemically undesirable belief was formed at a particular instance that she could, at that particular instance, have performed the doxastic action of *preventing* a belief with the *particular* content of the undesirable belief that she actually formed.

I shall argue below that property voluntarism is in fact the first legitimate strategy for accounting for cases of epistemically blameworthy beliefs encountered in this study: There *are* cases of epistemic blameworthiness that may appropriately be given a property voluntarist analysis. However, as I shall also aim to show, compelling cases of epistemic blameworthiness remain, for which property voluntarism cannot hope to account.

To my knowledge only two authors, John Heil and James A. Montmarquet, have seriously defended anything resembling property voluntarism in the sense presented here, and one of these authors, John Heil, only in a very sketchy vein.

9.2. JOHN HEIL AS A PIONEERING PROPERTY VOLUNTARIST

Heil openly rejects doxastic voluntarism:

On the other hand, believers appear to be passive; beliefs are not chosen or rejected, but simply held or not. In this regard, believers seem to be largely *at the mercy* of their belief-forming equipment.³⁹⁴

However, he does not take the downfall of doxastic voluntarism to be the end of doxastic control:

It is not that one has a choice in the beliefs that one forms, but that one has a say in the procedures one undertakes that lead to their formation. The notion of “epistemic responsibility” attaches to the undertaking of appropriate procedures.³⁹⁵

Heil mentions two kinds of such “appropriate procedures”: “suitable information-gathering” and “belief-refinement.” What chiefly warrants the labelling of Heil as a pioneering property voluntarist, is what he has to say about “belief-refinement,” the second type of “appropriate procedure” to which he “attaches” the notion of epistemic responsibility. First, according to Heil, perception is a “belief-saturated”³⁹⁶ phenomenon:

First I wish to suggest that perceiving is non-contingently a matter of belief-acquisition.³⁹⁷

Now, he continues, perceiving is an activity that can be undertaken in significantly different ways:

Perceivers move around their environment, exploring and investigating and manipulating what they find there. These activities are guided by beliefs already held and by wants and fears. They are guided as well by perceptual “feed-back”: if an object appears in some way odd or suspicious, if the lighting is unusual, one may be obliged to look more closely or carefully than at other times. We keep our eyes peeled, prick up our ears when the occasion demands. If much is at stake we may wish to examine things more cautiously still. These perceptual activities have the effect of refining perceptual beliefs.³⁹⁸

It is not altogether explicit in Heil’s short article, which connection he wishes to draw between the notion of epistemic responsibility and the notion of epistemic blameworthiness, which he only touches upon briefly.³⁹⁹ However, when the passages quoted above are taken together, a type (2) property voluntarist account of some instances of blameworthy belief is at least vaguely suggested: In particular it is suggested that an agent may be epistemically blameworthy for holding an undesirable *perceptual* belief because, in the perceptual process that constituted the formation of this belief, the agent did not guide her perceptual activity in a way that (sufficiently) *refined* that perceptual belief. Remark that this account does not make the doxastic voluntarist presupposition that the belief-forming process was in any way sensitive to the agent’s intentions regarding the *content* of the particular perceptual belief resulting from the “unrefined” perceptual activity. Bluntly put: It is supposed that the agent may choose the epistemic quality of the resulting perceptual belief, but not the resulting belief itself.

The type (2) property voluntarist strategy applied here may be re-stated thus: An agent is epistemically blameworthy for holding a belief, which is epistemically undesirable in virtue of being formed by the unreliable mechanism of “unrefined” perception, if, and only if her belief was formed by the unreliable mechanism of “unrefined” perception, because she did not guide her perceptual belief-forming process in a suitably “refining” manner, and has no appropriate excuse for this omission.

A particular application of this strategy could be the following: Consider an agent, who by looking at a red car forms, in an unreliable way, the belief that she is looking at a brown car. This agent could now be held epistemically blameworthy for holding the undesirable belief that she is looking at a brown car, in so far as her belief was formed in an unreliable way, because she did not “keep her eyes peeled” and has no appropriate excuse for this omission: She was quite simply perceptually lazy at the relevant time and performed an action of forming an epistemically undesirable belief.

I believe there is much to be said for such accounts of epistemic blameworthiness in particular cases like the above, where the property voluntarist account given seems overwhelmingly plausible (given, as we shall see in Chapter 12, that relevant intellectual obligations obtain). However, as we shall see in Section 9.4 below, many cases of epistemic blameworthiness cannot be accounted for by a property voluntarist strategy, and it is questionable whether property voluntarism can even in general account for cases of blameworthy perceptual beliefs. Before proceeding to further discuss the merits of property voluntarism, I shall present a more out-spoken version of the position due to James Montmarquet.

9.3. JAMES MONTMARQUET'S PROPERTY VOLUNTARISM

Montmarquet submits:

My thesis, simply put, is that doxastic responsibility is typically direct and *incomplete*. Doxastic responsibility, I want to say, is grounded in the control we have with respect to one aspect, one modality of the belief-holding process. With respect to other aspects, then, there is no claim to direct or even indirect control. More importantly, there is no claim here, that the existence (or occurrence) of a given belief is itself subject to one's control. In a given case I may be unable to control whether I hold a given belief, even if it is within my power whether I believe this item virtuously or not.⁴⁰⁰

Here we have the property voluntarist combination of a rejection of doxastic voluntarism together with a claim that a notion of epistemic responsibility may instead be grounded in the direct control agents enjoy over the instantiation of normatively significant properties of their beliefs. As indicated in the above passage, for Montmarquet the decisive normative property of a belief is its *being held virtuously*. He elaborates on the nature of this property in the following passage, employing the image of a carver at work:

Let us now broaden our notion of care to embrace the full range of epistemic virtues. How can we do this and why are we entitled to do it? My answer is this: "Care" as the generic term for what our carver is exerting, really embraces many more specific concerns. Thus the carver may be taking care specifically that his knife does not slip, or that he doesn't press too hard, and so forth. But whichever of these more specific concerns he has, it will remain true that "care" represents a modality of his carving and not some prior action. Likewise, then, insofar as the epistemic virtues can be readily seen as ways of exerting forms of care (e.g. care that one remain open-minded, etc.) Any of them may, under appropriate circumstances, be understood as modalities of the belief-forming process.⁴⁰¹

Following this explication, we may well want a fuller account of the virtues or "cares" relevant to belief-formation. I shall not here venture into Montmarquet's full catalogue of epistemic virtues. Rather, what is important here, is whether by "virtuous" Montmarquet means anything close to "epistemically desirable" in the present sense (see Chapter 5), specifically whether he links his conception of epistemic virtue tightly to the truth-goal.

Montmarquet generally holds the truth-goal too narrow to contain the "larger goals of epistemic life": "For ideally, one seeks not only truth but *science*. That is, one seeks a deep explanatory understanding of the world and ourselves as part of the world."⁴⁰² However, Montmarquet does not take his above claim to imply a broadened

conception of epistemic (intellectual) virtue. Here broadened *desires* in the epistemic agent will do:

Certainly, if our goals have broadened, we must suppose that the intellectually virtuous person's desires must be broadened correspondingly. She will seek more than truth and the avoidance of error; ideally she will seek these other [broader] goals as well. Beyond this obvious need to expand one's goals, however it is not clear that any different set of "regulative" traits are now required – beyond impartiality, intellectual courage and the like. For these traits are obviously broad enough to characterize epistemically virtuous activities on a small "everyday" scale, but also those in science as well.⁴⁰³

In other words: A catalogue of epistemic virtues based on the truth-goal need not be supplemented in order to encompass the broader epistemic goals of science and understanding. Epistemic virtues, in Montmarquet's sense, arguably are then simply truth-conducive belief-forming strategies (to which agents are habitually attached⁴⁰⁴). According to Montmarquet the central such virtue is "epistemic conscientiousness": the desire to attain truth and avoid error,⁴⁰⁵ in other words the desire to attain the truth-goal. Now, since Montmarquet explicitly discusses the issue of epistemic responsibility in order to account for ascriptions of epistemic blameworthiness (culpability) for holdings of beliefs,⁴⁰⁶ it seems far from far-fetched to ascribe to him a *global* property voluntarist stand on the issue of epistemic blameworthiness; i.e. the view that *all* instances of epistemically blameworthy holdings of undesirable beliefs may be accounted for by one of the two property voluntarist strategies⁴⁰⁷: As we saw above, he expressly rejects doxastic voluntarism, yet by way of his carpenter image also expressly denies that the relevant kind of doxastic "care" may be expressed in actions *distinct* from the belief-forming processes. The modes of doxastic control held relevant to deontic evaluations thus seem to be *direct* in the terminology preferred here, and thus must be modes of direct *property* control.

Further, even though Montmarquet's conception of epistemic undesirability is virtue-driven, since he ties his conception of epistemic virtue narrowly to the truth-goal, the notion of epistemic undesirability adopted in Chapter 5 cannot be alien to his approach to epistemic deontology. This means that his position translates without too much difficulty into the property voluntarist framework presented in Section 9.1. I shall engage with Montmarquet's position under this heading in what follows.

9.4. THE LIMITATIONS OF PROPERTY VOLUNTARISM

As pointed out above, there are definitely cases of epistemic blameworthiness that may be satisfactorily accounted for by one of the two strategies available to property voluntarism. The most obvious candidates are those where the undesirability of a perceptual belief is due to the agent's crude perceptual laziness; i.e. an inexcusably sub-optimal perceptual performance constituting the formation of an undesirable perceptual belief. We encountered a specific instance of this type in the red car-brown car example at the end of Section 9.2.

However, I shall submit, there are also compelling cases of epistemic blameworthiness, which the property voluntarist is unable to account for. Here again the case of "the educated racist" (first presented in Section 1.2) comes in handy. Suppose, as

before, that the educated racist has formed her racist belief that blacks are generally inferior to whites on the testimony of her favourite demagogue; a highly unreliable belief-forming mechanism. Now suppose also that the critical testimonial act of the demagogue had conspicuous properties, which the agent would perceive even in her dullest, most “vicious” mood. Suppose further, that these properties exactly fitted the educated racist’s working-indicator-properties (see Section 10.6) for credibility judgments concerning contents of the type expressed: The educated racist’s cognitive make-up at the relevant time was simply such that a testimonial act as that of the demagogue *inevitably* induced in her the belief that she had good reasons to believe that blacks are generally inferior to whites. Further, inevitably under the circumstances, that belief again induced in her the belief that blacks are generally inferior to whites.

Now, it seems plausible that this credulous cognitive make-up does not in itself excuse the educated racist from epistemic blame, and that she is therefore epistemically blameworthy for holding the undesirable racist belief in the lack of more appropriate excuses. After all, her credulity in the relevant situation might have been the educated racist’s own fault. Just as we do not excuse a driver for doing harm, just because her driving abilities were impaired due to self-imposed intoxication at the time of harm-doing, we should hardly excuse an agent from epistemic blame, just because the occurrence of the undesirable belief was inevitable given her self-imposed cognitive deficiencies.

However, property voluntarism cannot account for the above case of epistemic blameworthiness: It hardly makes sense to say that, in the decisive moment when the educated racist faced the testimony of the demagogue, she had any active say in her belief-forming process: Due to her credulous cognitive make-up, the undesirable belief formed inevitably and automatically. That is just the kind of doxastic agent she was at that moment, just what her “epistemic character” had in store. This completely rules out applying a property voluntarist strategy to this example, since it demands that the agent’s perceptual refinement or employment of epistemic care in the process of belief-formation versus the lack thereof be the cause of the bad epistemic quality of the outcome of that process.

Montmarquet realizes that his position seems to have a problem with cases of epistemic blameworthiness like the above, but also seems to think that his position holds the resources to deal with this worry:

Suppose that S believes that p unvirtuously; suppose too that S could have been virtuous under these circumstances, but that S would still have believed p. Less abstractly, suppose that Hitler, say, would still have held many of his vicious beliefs even had he been as epistemically virtuous as he was capable of being. In that case, whereas we can still fault him for a lack of virtuous efforts, it seems that we cannot truly blame him for these beliefs themselves . . . (. . .) . . . Let me address these concerns by turning at once to our main case in point. Assuming Hitler to be at least a somewhat sane individual, it is not very likely to be true that he could have virtuously held to these beliefs. For instance, could Hitler have conscientiously believed in the authenticity of the Protocols of Zion and the so-called International Jewish Conspiracy? This does not seem very likely. To that extent, then, I am in a position to claim here that the worrisome supposition does not remain, on reflection, all that worrisome. Either Hitler is only insane – a definite possibility but not one I am exploring in this study – or, as I have been assuming, he is not. If he is not

insane, then it remains highly plausible to suppose that he is morally accountable for his actions by way of being epistemically (and morally) culpable for his belief.⁴⁰⁸

In Montmarquet's initial formulation, his Hitler example is similar to the version of the "educated racist" example given immediately above, where presumably no amount of virtue would have directly prevented the formation of the racist's undesirable belief. I shall hope to show that his way of dealing with this example is inadequate: The "worrisome proposition" *is* in fact worrying to Montmarquet's global property voluntarism.

First, it might well be that, as a matter of historical fact Hitler was in fact too cognitively well-endowed to be able virtuously to form a "vicious" belief like the belief that the Protocols of Zion are authentic. In that sense, the supposition that he could have virtuously held this "vicious" belief is indeed highly implausible, as Montmarquet observes in the above passage. However, this answer provides no substantive shelter for the worries Montmarquet is facing. The *real* problem arises, if Hitler was sufficiently sane to be an appropriate object of epistemic deontic evaluation but was in fact *not* as cognitively well-endowed as supposed above. That is: If it were the case (perhaps contrary to historical fact) that Hitler⁴⁰⁹ would still have formed "vicious" beliefs, no matter how much care he had employed in his belief-forming processes. I shall assume here with Montmarquet that Hitler was in fact mentally sane in the relevant sense (at least for most of his despotic career).

Now, on these suppositions, it suddenly seems less plausible that property voluntarism commands the resources to appropriately blame Hitler for holding his "vicious" beliefs, in fact it now seems quite impossible that this could be the case. Employing Montmarquet's preferred carver-analogy, Hitler would in that case be like a carver with very little talent: No matter what care this carver shows in handling his tools, his carvings still come out horribly ugly. Still, on the supposition that his blameworthiness for the ugliness of his work hinges only on the care he employs when carving, he cannot be blamed for the ugly result of his careful work. Thus, insofar as we take Hitler's epistemic blameworthiness for, e.g. his anti-semitic beliefs to account for the moral blameworthiness of his horrific actions prompted by those belief (and this is clearly Montmarquet's position⁴¹⁰), we have no other choice than to acquit him of moral blame for these actions, if we do not have access to a stronger strategy of underwriting ascriptions of epistemic blameworthiness than that provided by property voluntarism.

Some may now be inclined to respond on Montmarquet's behalf that something has gone wrong here: Surely the miserable carver from above cannot be virtuous since he produces such ugly carvings. Analogously Hitler cannot be a virtuous believer while holding such "vicious" beliefs. Arguably, however, this answer confuses different senses in which an agent can be virtuous. Virtue, even epistemic virtue, is a multi-faceted notion.⁴¹¹ Surely a very broad notion of virtue may take anything appropriately admirable in an agent to be a virtue, and anything appropriately detestable in an agent to be a vice. On such an account our cognitively deficient version of Hitler and our miserable carver are clearly more vicious than virtuous. However, Montmarquet

explicitly operates with a much more constrained notion of epistemic virtue, where such virtue is to express itself only in the “care” with which an agent forms her beliefs. Here it is clear, as Montmarquet also implicitly allows in stating his “worry,” that the cognitively deficient version of Hitler can show all the “care” he can muster, and still not aspire to epistemically desirable beliefs. Property voluntarism simply does not have the resources satisfactorily to evaluate such cases as that of the cognitively deficient version of Hitler.

Some general lessons can be drawn from the above considerations: As argued earlier (Section 8.2.1.5 and Section 8.2.2.1), the human cognitive system seems normally to operate so that an agent’s judgment (belief) that she has good rationalizing reasons to believe a proposition *p* is causally sufficient to induce in her the belief that *p*. It seems, then, that property voluntarism is best fit to account for cases where an agent’s application of epistemic care (under which falls arguably Heil’s “belief-refinement”) can *directly* influence the formation of such second-order attitudes (reasonability judgments). Strategy (1) property voluntarism must then focus on cases of epistemic blameworthiness in which an agent’s “epistemic carelessness” may actively affect her reasonability judgments. In contrast, strategy (2) property voluntarism must focus on cases in which a passive lack of “epistemic care” influenced the agent’s reasonability judgments, so as to bring about the formation of an undesirable belief.

I believe there are successful applications of both strategies. Above (Section 9.2) we encountered a successful strategy (2) application (the red car-brown car example). A successful strategy (1) application is harder to come by, but the following might suffice: Suppose an agent stands under some kind of obligation to visually identify as accurately as possible some distant objects, say, cars coming over a far-away hilltop (perhaps she is part of a team of investigators trying to solve a difficult criminal case). However, bored with her task, she plays a game with herself: In order to make the visual identifications more challenging, she half-closes her eyes at the moment she focuses on a car, so that she can barely see it. In fact this method of identification is highly unreliable. However, as the agent believes herself to be an expert auto connoisseur, she remains confident that she can reliably identify all cars even under these constraints. The following now happens: A Ford Taunus climbs the hill and the agent with her eyes half-shut identifies it as an Opel Manta. As a consequence she confidently forms the belief that an Opel Manta just climbed the hill. Now insofar as this belief is epistemically blameworthy, the first strategy available to the property voluntarist may apply: The agent is plausibly blameworthy for holding her epistemically undesirable Opel Manta belief, because, by her game, she brought it about *in the belief forming process* that she formed an undesirable belief and she seemingly has no appropriate excuse for doing it (boredom hardly suffices).

The requirements for deontological adequacy, which I set up earlier above (Section 7.2), are clearly met in both types of example: It is not assumed that the agents have any extraordinary cognitive powers or mental abilities. This vindicates the naturalistic requirement. Further, it seems clear that taking care in one’s belief-forming process (e.g. in the sense of keeping one’s senses to the best of their abilities), is a norm according to which normal agents can guide their conduct. This takes care

of the activity-guiding requirement. Finally, it is obvious that genetic property control is epistemically relevant: as seen above, an agent's exercise of this type of doxastic control can matter crucially to the epistemic desirability level of her resulting beliefs.

My choice of examples might leave the superficial impression that property voluntarism is in fact well equipped to deal with *any* case of blameworthy perceptual belief. However, there are cases of blameworthy empirical belief that it cannot hope to account for: Suppose, e.g. that, due to the operation of some unreliable belief-forming mechanism, I believe that all dogs with long hair are terriers. Suppose further that I am able correctly to identify a longhaired dog in front of me even in my most perceptually dull and careless mood. I now see a longhaired dog under these conditions and form the perceptual belief that a terrier is in front of me. Really the dog is an English sheepdog. My belief that a terrier is in front of me is clearly epistemically undesirable; my mechanism for forming beliefs about the race of dogs hardly qualifies as reliable. It could now be that I am here blameworthy for holding the undesirable belief that a terrier is in front of me, given among other things that I have no adequate excuse for holding the first undesirable belief that all longhaired dogs are terriers. However, in my process of forming the blameworthy belief that the dog in front of me is a terrier, all the care in the world would not have changed the result.

We may express in more general terms why property voluntarism lacks the resources to deal with examples of epistemic blameworthiness like the above and those of the educated racist and the cognitively deficient version of Hitler: In these cases, no investment of epistemic care in the belief-forming process could in the least have influenced the relevant agents' reasonability judgments. The blameworthy agents and their environments in these examples were simply so constituted in the relevant slice of space-time, so as to make the agents very vulnerable to an unreliable belief-forming mechanism. Put in different terms, their belief-forming processes were simply unable to be decisively influenced by all the momentary carefulness in the world.

It seems clear, then, that the source of epistemic blameworthiness (if such in fact obtain) must in such cases be sought in the *cognitive dispositions* (including, of course, the doxastic background) of the agent at the time of the belief-forming process, rather than in the lack of care employed in her belief-forming processes. To escape an unhappy divorce between doxastic control and epistemic blameworthiness, it must then be argued that for a normal agent such cognitive dispositions are not entirely beyond her control, and that through her influence on these dispositions and her activity *before* a critical belief-forming process, she may *indirectly* influence the particular beliefs that her cognitive system ultimately generates. In the Chapters 10 and 11, I shall consequently be concerned with the deontological adequacy of modes of indirect doxastic control.

INDIRECT CONTENT-DIRECTED DOXASTIC CONTROL
OR DOXASTIC PASCALIANISM

Abstract. In this chapter I discuss the merits of modes of indirect content-directed doxastic control in terms of deontological adequacy, i.e. I discuss whether an agent's indirect influence on the content of her beliefs may play a legitimate role in evaluations of epistemic blameworthiness. I conclude that it may only do so within a very limited domain of cases. In Section 10.1 I present William Alston's particular theory of epistemic blameworthiness based on a mode of indirect influence (control) over belief-contents. I argue that his analysis must be refined in order to merit serious attention, and present a suitably refined version of it. In Sections 10.2 and 10.3 I then argue that, even after a thoroughgoing refinement, Alston's analysis is still inadequate in two respects. First (Section 10.2) it fails to pay attention to the substantive role that the notion of *foresight* must play in an adequate conception of blameworthiness beget by exercises of indirect control. Second (Section 10.3), it is vulnerable to a certain "Frankfurt-style" objection, highlighting that, on Alston's conception, an agent's epistemic blameworthiness implausibly hinges on accidental matters entirely unrelated to the blameworthy agent's past. I proceed in Section 10.4 to establish that modes of indirect control over belief-contents may in fact matter to some evaluations of epistemic blameworthiness. However, the relevant cases are highly exotic, and the appropriate way to bring out the relevance of such modes significantly differs from Alston's efforts in this direction. In Section 10.5 I sum up the merits of appeals to indirect control as a basis for ascribing epistemic blameworthiness. In Section 10.6 finally, I add the ironic twist to the discussion that appeals to indirect control over the content of *other* agents' beliefs may be better fit to ground ascriptions of a species of epistemic blameworthiness, albeit not the species primarily topicalized in the present work.

10.1. WILLIAM ALSTON'S CONCEPTION OF EPISTEMIC
BLAMEWORTHINESS

In Alston (1989), William Alston presents what can only be taken as an attempt to render a mode of indirect content-directed doxastic control globally deontologically adequate. After rejecting doxastic voluntarism, which he takes to offer the "more natural" theory of epistemic blameworthiness, he proceeds to lay out an alternative account, which he considers "the only viable"⁴¹² alternative to doxastic voluntarism:

This suggests that even if propositional attitudes [like beliefs] are not under our effective voluntary control, we might still be held responsible for them, provided we could and should have prevented them; provided there is something we could and should have done such that if we had done it we would not have had the attitude in question. If this is the case, it could provide a basis for the application of deontological concepts to propositional attitudes, and perhaps for a deontological concept of epistemic justification.⁴¹³

Below I will confront Alston's specific suggestion for an account of epistemic blameworthiness (deontological unjustifiedness) along these lines, and argue that his account fails to capture more than a class of highly exotic cases of epistemic blameworthiness, even if remedied of its most obvious shortcomings.

Alston's suggestion for an analysis of the notion of epistemic "intellectual" blameworthiness along the above lines is the following:

S is (intellectually) to blame for believing that p if, and only if, if S had fulfilled all her intellectual obligations, then S's belief-forming habits would have changed, or S's access to relevant reverse considerations would have changed, in such a way that S would not have believed that p.⁴¹⁴

He goes on to identify "deontologically justified belief" with belief that is *blameless* (i.e., not blameworthy) on the above account and concludes that, although he will not deny that this concept of justification is "important and interesting,"⁴¹⁵ there is nothing to be said for it as a "fundamental concept for epistemology."⁴¹⁶ I shall not challenge this conclusion here. Rather, I shall be concerned with whether Alston's above analysis can in fact adequately account for any cases of epistemic blameworthiness at all.

We have already seen that, contrary to Alston's suppositions, certainly his conception cannot be the "only viable" account of the notion of epistemic blameworthiness: Property voluntarism was successful within a limited domain of cases. The interesting problem then becomes, whether Alston's account can take care of any cases not accounted for by property voluntarism. As we shall shortly see, Alston's above analysis suffers from some obvious shortcomings. However, due to its prominence in the literature on deontic epistemic normativity, it seems warranted in the present context to bring these flaws out, in order to see if a suitably remedied version may command any plausibility.

Above all, Alston does not say anything about the epistemic "intellectual" undesirability of the beliefs supposedly blameworthy on his account. This leaves open the rather bizarre possibility that an agent may be blamed for holding even the most every-day and reasonable belief, if contingently it happens to be the case that she would not have held this particular belief, if only she had some time in the past fulfilled some particular "intellectual obligation."

A poignant example might be this: Consider a woman, Anna, who once violated an intellectual obligation by skipping a college class that would have improved her belief-forming habits within a certain domain. However, as a result of skipping this class she strolled through town and met her future husband Benny, whom she would otherwise never have met. Consequently, she came to hold many beliefs about Benny, many of which were not epistemically undesirable at all. For example Anna came to hold the perfectly reasonable belief that Benny has dark hair. Now, certainly, if Anna had fulfilled all of her "intellectual obligations," she would never have formed the belief that Benny has dark hair: In fact she would never have acquired the "habit"⁴¹⁷ of forming beliefs about Benny at all. Thus, on Alston's above account she appears to be blameworthy for believing that Benny has dark hair. This seems utterly implausible.

The above example points to two major shortcomings in Alston's account that must be remedied before it can aspire to any degree of plausibility: First, it cannot be that an agent is epistemically blameworthy for holding a belief that is not epistemically undesirable in any way. Earlier (Section 4.2) I offered reasons to reject the possibility of "epistemic accuses."⁴¹⁸ Second, the violations of intellectual obligations supposedly relevant to an evaluation of epistemic blameworthiness must stand in a tighter

relationship to the holding of the particular undesirable belief under deontic evaluation: Supposedly, Anna also holds a number of epistemically *undesirable* beliefs about Benny. It would certainly be odd to blame her for holding these beliefs with reference to the fact that she would never have held them, had she not skipped class that long past day.

Rather, if Anna is to be blamed for holding such a belief with reference to violated intellectual obligations in her past, surely we must demand that her past violations must at least be more than accidentally related to the specific *type* of sub-optimal cognitive performance involved in the formation of the undesirable belief. If the violations of intellectual obligations pertaining to “belief-forming habits” in Alston’s picture are to carry weight in deontic evaluations, we must, very roughly speaking, demand those habits to concern cognitive dispositions of an ingrained nature⁴¹⁹ and we must further demand that the epistemic undesirability of the relevant belief be made out as a result of an application of a belief-forming habit formed by a violation of an intellectual obligation. For example if Anna unreasonably believes that Benny is a psychopath, and the class, which she once skipped, would have improved her ability to reliably diagnose psychopathy, her skipping of that class suddenly seems relevant to the deontic evaluation of her belief.

So far I have left the “access to relevant adverse considerations” clause in Alston’s analysis out of the picture. According to Alston, this clause becomes relevant in “cases in which we acquire or retain the belief only because we are sheltered from adverse considerations in a way we wouldn’t have been had we done what we should have done.”⁴²⁰ It is not entirely clear which types of cases Alston has in mind here. My best guess is that he is here referring to an earlier passage, where he submits that: “With respect to a particular issue, I have voluntary control over whether, and how long, I consider the matter, look for relevant evidence or reasons, reflect on a particular argument, seek input from other people, search my memory for analogues cases, and so on.”⁴²¹

Alston does not observe that this kind of indirect influence on beliefs is severely restricted by the agent’s cognitive dispositions: Once a belief is formed, the agent is now convinced to a high degree of the truth of its content. Thus she can no longer freely “consider the matter”: If she really believes the proposition, it seems unimaginable how she should be motivated to continue looking for relevant evidence with the same zest as before she came to believe it. Still, up until a belief is formed, there might be room for “activities that bring influences to bear, or withhold influences from, a particular situation involving a particular candidate, or a particular field of candidates for belief,”⁴²² to be of relevance to the deontic status of an undesirable belief that is eventually formed. We thus get a further set of intellectual obligations relevant to an agent’s epistemic blameworthiness, now having to do with the evidence-gathering undertaken *before* the formation of the particular undesirable belief (the first kind of obligations related to “habits,” i.e. cognitive dispositions *when* confronted by a certain set of evidence).

Once the difference between these two kinds of “intellectual obligations” is brought out, it seems clear that Alston errs in demanding for blameworthiness only a decisive

shift in the doxastic states of the agent, if the agent fulfils *all* of her intellectual obligations. Surely it might plausibly be the case that an agent is blameworthy, *only* because she violated her evidence-gathering obligations, or *only* because she violated her obligations pertaining to her cognitive dispositions. In contrast, what would be the case if the agent had fulfilled *all* of her obligations is often irrelevant to the blameworthiness of her beliefs: In a specific case, it could, e.g. be that, if the agent had only fulfilled her relevant *cognitive disposition obligations*, she would not have held the blameworthy belief, no matter how many of her *evidence gathering obligations* she had violated at the particular instance. This is most obviously the case if, given the “right” kind of cognitive dispositions, the falsity of the blameworthy belief would have been so obvious to the agent, that she would have refrained from forming it, no matter the quality of the evidence-gathering she could have undertaken.

Even after taking note of these observations, still a further qualification of Alston’s analysis is needed: Accounting for an agent’s blameworthiness with reference to her past violations of intellectual obligations surely cannot work, unless at least it is the case that the agent is not appropriately *excused* for these violations, that is: unless she is not blameless for violating these obligations. It is not altogether clear, whether Alston uses the term “intellectual obligation” in a sense that renders it impossible to be excused for violating such an obligation. Certainly though, such a usage would be far from the standard usage of the term “obligation” with regard to actions: For example in standard terms we all stand under an obligation to act according to the criminal law of our respective legal systems, but may sometimes be excused for not so acting. Since Alston does not make it clear that he is using the term “obligation” in a non-standard way, I shall submit that his account must be supplemented by clauses securing that the blameworthy agent was not appropriately excused for her failure to adhere to her intellectual obligations relevant to a particular case. Lastly, although this is more of a logically nitpicky point, I will bring out on Alston’s behalf, that the blameworthy agent *actually* did not fulfil her intellectual obligations, which interpreted (over)literally, Alston’s analysis does not imply as it stands.

Suitably clarified and remedied of its most obvious implausibilities, Alston’s thesis then comes out as follows (where I observe a maximum of continuity with Alston’s preferred terminology):

S is (intellectually) to blame for believing that p if, and only if her holding the belief that p is epistemically undesirable and it is either the case that

1. (i) If S had fulfilled all of her intellectual obligations relevant to the particular cognitive disposition “belief-forming habit” active in the formation of her belief that p, then this cognitive disposition would have changed, in such a way that S would not have believed that p. (ii) S did not fulfil these intellectual obligations. And (iii) S is not appropriately excused for her failure to fulfil these obligations Or
2. (i) If S had fulfilled her intellectual obligations in her activity of gathering evidence relevant to the field of candidates for belief of which p was a member, S’s access to relevant reverse considerations would have changed, in such a way that S would

- not have believed that p. (ii) S did not fulfil these intellectual obligations. And (iii) S is not appropriately excused for her failure to fulfil these obligations Or.
3. (i) If S had fulfilled her intellectual obligations relevant to the particular cognitive disposition active in the formation of her belief that p and S had fulfilled her intellectual obligations in her activity of gathering evidence relevant to the field of candidates for belief of which p was a member, then S would not have believed that p. (ii) S did not fulfil these intellectual obligations. And (iii) S is not appropriately excused for her failure to fulfil these obligations.

I shall now prepare the ground for my argument that such an analysis, despite its ingenuity, may not hope convincingly to account for any cases of epistemic blameworthiness, insofar as it is couched in counterfactual terms.

10.2. DOXASTIC PASCALIANISM

From the above it should be clear that Alston is invoking modes of indirect content-directed doxastic control in his conception of epistemic blameworthiness. More specifically he is invoking the mode of *negative genetic indirect content-directed doxastic control*: It is an agent's power to indirectly prevent (refrain from) the formation of a *particular* undesirable belief by way of her control over her cognitive dispositions and her evidence-gathering activity that is to account for her epistemic blameworthiness for holding that particular belief.

Given the taxonomy of modes of doxastic control presented earlier (Section 7.3), it should already be clear that, contrary to Alston's opinion, his analysis is even far from being the only possible strategy for underwriting ascriptions of epistemic blameworthiness based on a mode of indirect content-directed doxastic control: As the combinations of positive/negative versus genetic/abortive modes of doxastic control yield no more than four possible modes of indirect content-directed doxastic control, there are also four strategies for accounting for ascriptions of epistemic blameworthiness on an indirect content-directed basis. Alston has at most spelled out the negative genetic option in some (insufficient) detail.

Now, spelling out all four options in the detailed manner in which I spelt out my refined version of Alston's analysis, is arguably too tedious. It should be at least superficially clear, how these accounts would go: The positive genetic option blames the agent by appeal to her having indirectly brought it about the particular undesirable belief, the positive abortive option blames the agent by appeal to her *not* having brought it about that the undesirable belief was suspended by her means of indirectly controlling that the particular undesirable belief was suspended, and the negative abortive option blames the agent by appeal to her having indirectly brought it about that, once formed, the undesirable belief was sustained.

Trudy Govier has aptly termed indirect content-directed doxastic control "Pascalian control" in honour of the pious French philosopher Blaise Pascal, who held that a belief that certain (Roman catholic) religious dogmas are true may be induced by attending mass regularly, living a moral life etc.⁴²³ In honour of Govier's terminology I shall

here use the term “Doxastic Pascalianism” or simply “Pascalianism” to cover any position that aims to account for legitimate ascriptions of epistemic blameworthiness by appeal to a mode of indirect content-directed control, i.e. to cover any position that holds one or more modes of indirect content-directed doxastic control deontologically adequate for some non-empty domain of undesirable beliefs. I will aim to demonstrate in the present chapter that a Pascalian strategy may well work to account for some cases of epistemic blameworthiness, but that such cases are arguably rare in the actual world. I will be able to do this without significant appeals to the details of the refined Alstonian account given above.⁴²⁴ The difficulties arise on a more general level.

To engage in this argument, however, I must at first clear away an argument to the conclusion that indirect doxastic control cannot matter at all to issues of epistemic blameworthiness. This (surprising) claim has recently been proposed by David Owens:

There are self-help manuals which assure us that by rehearsing positive attitudes in our auditory imagination we can banish negative attitudes and unhelpful beliefs about ourselves. Suppose that were true. Then we could act to get rid of these undesirable mental states solely on the grounds that it was bad for us to have them (regardless of whether they were evidentially well founded or not), and such actions might be perfectly reasonable: it would be rational to induce beliefs which were themselves irrational. The exercise of this sort of belief control is governed by practical and not by epistemic norms, and therefore the fact that we have control of this sort over belief can do nothing to explain why our beliefs are responsible to epistemic norms. But the whole point of trying to construct a notion of control which will cover belief is precisely to explain how beliefs can be subject to epistemic norms in a way consonant with the juridical theory of responsibility.⁴²⁵

This argument is hardly more than a generalization of the fallacious claim encountered earlier above that exercises of direct content-directed doxastic control (doxastic actions), if such exist, cannot possibly be guided by “truth-considerations,” e.g. reflections on one’s rationalizing reasons for belief. I presented reasons to reject this claim in Section 8.2.1.5: There is absolutely no reason to rule out the possibility that an agent recognizes that her believing a certain proposition would be *epistemically desirable*, and decides to form this particular belief for that reason, although, as demonstrated in Chapter 8, such intentions are hardly ever effective in normal agents.

Exactly the same point applies to the indirect control case mentioned by Owens: Choosing a truly “Pascalian” example, we might, e.g. imagine an agent, who after having grasped certain theist arguments, believes that she has good reasons to believe that God exists. However, due to certain emotional barriers (e.g. a resentful attitude towards her overly religious parents) she still cannot believe that God exists. Now, in order to come to believe that God exists, she adopts the lifestyle of a believer, attends mass regularly, seeks the company of devoted believers etc. Imaginably, she might even succeed in this strategy: Her emotional barriers gradually break down, and some day she sincerely believes that God exists.

What this example shows is that, contrary to Owens, an agent’s exercise of her Pascalian control may very well be “governed” by epistemic norms in a sense relevant to ascriptions of epistemic blameworthiness, although typically this is perhaps not the case. Perhaps the background for Owens’ claim that “the fact that we have control

of this sort over belief can do nothing to explain why our beliefs are responsible to epistemic norms” is nothing more than a failure properly to distinguish between deontic and non-deontic epistemic norms: Surely the fact that I may exercise some measure of control over my beliefs, does not go any way towards explaining why we should value reasonable beliefs over unreasonable beliefs from an epistemic perspective, i.e. it goes no way towards explaining why our beliefs stand under evaluative norms in the first place. However, as established in Section 7.1, the fact that we have some (limited) control over our beliefs, “explains” why they may be appropriate objects of *deontic* epistemic evaluation, insofar as, without such control, no such evaluation would arguably be appropriate.

The demise of Owens’ worries means that it is at least legitimate to discuss whether some mode of Pascalian control is deontologically adequate in some cases. Still this is a long and difficult way from actually establishing which cases that could be.

10.3. A FIRST GLIMPSE OF THE FORESIGHT PROBLEM

Even the refined Alstonian version of Pascalianism faces a problem, which, as I shall later argue, must be dealt with by any account of epistemic blameworthiness, but is particularly conspicuous here. The problem concerns *foresight*: It seems plausible to suppose that, even if an action is in itself blameworthy (an inexcusable violation of some obligation), it can only give rise to blameworthy consequences if the agent performing the action either foresaw or could be *required* to foresee those consequences at the time of action, at least under some appropriate description of those consequences.

An example from the practical realm might help clarify the above point: Suppose a manufacturer of canned soup commits a blameworthy action by adding to his soup some cheapening ingredient than he does not mention in the declaration on the soup cans. At the time he commits this misdemeanour, the extra ingredient, although ill-tasting, is generally considered entirely harmless even by the best scientific authority, and the manufacturer consequently believes that his action is quite benign. However, in fact the ingredient is highly cancer provoking in larger quantities. As a direct consequence of the manufacturer’s blameworthy action, many regular soup eaters contract cancer and die, surely an undesirable consequence. Now, even if these unfortunate deaths result from the manufacturer’s blameworthy action, it is not clear that the manufacturer is to blame for the cancer deaths in anything but a causal sense. There simply was no way that he could have been required to foresee the unfortunate outcome of his action, not even under the very general description “some people will get ill or die” and hence no way in which he could have been demanded to take this possibility into account when he acted to add the extra ingredient. He was just very unlucky, to grossly understate matters.

To sum up, in the above example the following seems to hold: (1) The manufacturer performed a blameworthy action in the past; he inexcusably violated some obligation. (2) The cancer deaths occurred as a result of his past blameworthy action. (3) Yet the manufacturer is not to blame for the occurrence of these deaths. He simply neither

foresaw, nor could be required to foresee⁴²⁶ at the time he added the ingredient that his action would bring about anything like the cancer deaths.

This example highlights why holding an agent epistemically blameworthy as a simple consequence of fulfilling one of the conditions of the refined Alstonian account, seems a very odd thing to do: Consider how the endemic example of the blameworthy “educated racist” would be accounted for by the refined Alstonian account. Construe the example as before, where the educated racist simply has cognitive dispositions of such a quality, that she is compelled to form the epistemically undesirable belief that blacks are generally inferior to whites after hearing the testimony of her favourite demagogue. Now assume that no kind of evidence-gathering activity leading up to the formation of the racist belief would have prevented her from forming it when hearing the testimony of the demagogue. The educated racist’s dispositions were simply such as to ignore any evidence that would have raised her epistemic shields in the relevant situation.

If the educated racist is epistemically blameworthy in this example, it must be accounted for by the *first* of the three disjunctive clauses of the refined Alstonian account which must then argue that (1) there are violations of intellectual obligations in the educated racist’s past relevant to her unfortunate credulous dispositions concerning the testimony of the demagogue, (2) that she has no appropriate excuse for these violations, and (3) that had she not violated her intellectual obligations at this or these instances, her credulous cognitive disposition would have changed such that she had not formed the racist belief on the testimony of the demagogue.

Let us grant that all of this is true: For example we may suppose that the educated racist skipped her race issue classes in college, and that she thereby inexcusably violated one or more intellectual obligations. Let us further grant that, had she not skipped those classes, her cognitive dispositions would have changed so as to make her unable to trust the testimony of the demagogue on race issues. Now, is this enough to underwrite the relevant ascription of blameworthiness?

One might doubt if this is really so, given the further supposition that she neither had, nor could be required to have, *any* relevant kind of foresight in the situation where she broke her intellectual obligations. Suppose, e.g. that this agent believed and had good reasons to believe that another lecturer in another class would repeat everything said in the class she skipped, and that she planned to attend this other class. She therefore skipped the first class (may be she had important homework to do). Now, by an unforeseeable accident she breaks her leg between the two classes and never makes the second one. In fact she is hospitalised for a longer period and never makes any further classes on race issues. As a result of this, her skipping the first class begets the long-range consequence that her cognitive dispositions regarding testimony on race issues remain highly credulous. However, given the circumstances she could hardly have been required in any sense to foresee that her skipping the first class would have such long-range consequences. She had any good reason to believe that her skipping that class would not have any long-range consequences at all relevant to epistemic quality of her future beliefs.

Given the above circumstances, it seems plausible that at least we still do not know enough to declare the agent blameworthy for holding her racist belief, even if she broke an intellectual obligation by skipping her original race issues class. If this is correct, even in its refined version Alston's analysis simply has not offered a satisfactory account of how past blameworthy actions or omissions (inexcusable violations of intellectual obligations) matter to the deontic evaluation of present undesirable beliefs. At least his account must be supplemented by clauses concerning the agent's foresight at the time she violated her intellectual obligations. Supposedly this supplementation could be carried out. Unfortunately, though, worse troubles for the refined Alstonian account of epistemic blameworthiness are still in store.

10.4. A FRANKFURT-STYLE OBJECTION

I shall now present an argument to the conclusion that no version of Doxastic Pascalianism trading in counterfactual terms (as does Alston) may hope to provide a convincing account of ascriptions of epistemic blameworthiness.

My argument to this conclusion will turn on a epistemic version of the kind of counter-example to certain theories of responsibility for actions known as a "Frankfurt-example" (after Harry G. Frankfurt's classical 1969 paper "Alternate Possibilities and Moral Responsibility"⁴²⁷). In the typical Frankfurt-example, some mechanism ensures that an agent cannot fail to perform a certain action, no matter what she does to prevent it. For example a woman may be unable not to commit a certain crime, because, if she should falter in carrying out the crime, a demonic scientist triggers a mechanism built into her brain that induces her to commit the crime anyway. Now, Frankfurt famously argued, if the scientist does not intervene, the woman is responsible for committing the crime, even though she could not have done otherwise.

As suggested by John Martin Fischer and Mark Ravizza, Frankfurt-style examples are also highly relevant for considering the issue of moral responsibility for action consequences. Fischer & Ravizza's preferred example is the so-called "missile" case, which I will here present in its least complicated version:

In "Missile 1" an evil woman Elizabeth, has obtained a missile and missile launcher and she has decided (for her own rather perverse reasons) to launch the missile toward Washington D.C. Suppose that ... she has not been manipulated, brainwashed, and so forth. Further imagine that she has had ... [a] sort of device implanted in her brain ... and that there is a counterfactual intervener associated with her who would ensure that Elizabeth would launch the missile, if she were to show any sign of wavering. We also suppose that, once the missile is launched toward the city, Elizabeth cannot prevent it from hitting Washington. Now when Elizabeth launches the missile toward Washington [without intervention], she does so freely, and we believe that she is morally responsible for the occurrence of the consequence-universal that *Washington D.C. is bombed*.⁴²⁸

There are several debatable issues surrounding the relevance of examples like the above to theories of moral responsibility and moral blameworthiness. For example it might be contested, whether in fact there is in any interesting moral difference between a case where Elizabeth bombs Washington without the intervention of the

counterfactual intervener, and cases where she bombs Washington after the intervener has intervened, if in both cases that the bombing was a direct result of Elizabeth's intentions (whether these intentions were prompted by the intervener or not). I shall not address these worries here. Rather I shall proceed to offer an epistemic counterpart of the above example. As intentions are not directly involved in the production of beliefs (see Chapter 8), such an example need not fear worries of the above sort. Neither need it appeal to exotic sci-fi inventory like remote-controlled brain implants.

Consider again our well-worn educated racist. Now suppose that, beside her familiar racist belief that blacks are generally inferior to whites, she also forms other beliefs on the testimony of her favourite demagogue on the occasion, on which she forms her racist belief. Among those beliefs is a belief about a crime statistic. The content of this belief happens to be true. However, the crime statistic belief is formed by exactly the same unreliable belief-forming mechanism as the racist belief, namely by the operation of her credulous cognitive disposition regarding the testimony of the demagogue, and must therefore be blameworthy (or blameless) for exactly the same reasons as her racist belief. Suppose, plausibly, that both beliefs are epistemically blameworthy.

Now suppose that on the same meeting a distinguished criminologist was the speaker before the racist demagogue. In the run of his speech, the distinguished criminologist utters the very same statistical statement that the demagogue later utters. At the meeting the educated racist, infused with racial hatred, pays no attention at all to the criminologist. Now suppose that the following is the case: If the educated racist had not once skipped a college class on racial issues and thus violated her intellectual obligations, her cognitive dispositions would have changed, such that on the meeting she would *still* have formed a crime statistics belief with the very same content, only in this case she would have trusted the reliable criminologist rather than the unreliable demagogue on the issue. Thus, *even* if she had fulfilled her intellectual obligations, she would *still* have believed the same propositional content, only this time around her belief would not have been epistemically undesirable.

Now, if Alston's account (even in its refined version) is applied to this "Frankfurt-style" case, it yields that the educated racist is *blameless* for holding the racist belief on the demagogue's testimony, and this *only* because the racist had the incredible *stroke of luck* that accidentally the condition was *not* fulfilled that, had she fulfilled her relevant intellectual obligations, she would not have held the crime statistics belief. But, intuitively, such luck cannot matter to an agent's epistemic blameworthiness. An agent's epistemic blameworthiness must hinge on the agent's exercise of her doxastic control in past and present: It cannot be that factors entirely outside her control such as counterfactual doxastic interveners decide whether or not she is to blame epistemically for holding a particular undesirable belief.

Here some might perhaps object that many instances of blameworthiness hinge on sheer (bad) luck in a certain sense. For example in another scenario the educated racist would not have formed her undesirable racist belief, because the demagogue had cancelled his speech at the meeting. However, this kind of luck is decidedly different from the above: It concerns the luck an agent may have in not doing something

blameworthy viz. not forming a blameworthy belief. In this sense, an intoxicated driver is also taking part in a “blame lottery,” insofar as she might be lucky enough not to hit anyone on her way home from a party. However, *once* she brings about an accident, whether she is blameworthy for it depends on her conduct, not on additional contingencies. In contrast, Alston ultimately asks us to make sense of the idea that two agents can be identical in their history and conduct, yet one is blameworthy and the other blameless for holding epistemically undesirable beliefs with identical contents under similar actual circumstances. This I take it to be unacceptable.

Alston’s account then immediately loses any plausibility in cases where the possibility of a “counterfactual doxastic intervener” like the bland criminologist in the example above cannot be ruled out in principle.⁴²⁹ Otherwise any ascription of blameworthiness accounted for by Alston may immediately be ruined by the observation that the agent only came out blameworthy on his account due to sheer bad luck.

Now, since arguably any normal agent can be made to believe a great variety of propositions given an exposure to suitable testimony, brain washing, etc., it seems clear that Alston can at most account for cases of epistemically blameworthy belief, where it would have been *psychologically impossible* that the agent could have been brought to believe the content of the relevant undesirable belief in the relevant scenario, if only she had not violated her relevant intellectual obligations. Only these circumstances rule out a counterfactual doxastic intervener. This observation immediately narrows the scope of Alston’s account severely: The violations of intellectual obligations involved must be so severe that they made it psychologically possible for the agent to believe a proposition, which, under the relevant circumstances, she could otherwise not be brought to believe by any means, save by having in her past the very same violations. I am not sure, which cases fit this description at all.

It should be clear that these considerations apply equally to any version of Pascalianism that appeal to general counterfactual conditions, as these conditions can all be “Frankfurtized” (i.e. subjected to an example like the above, where the agent would still have believed the same proposition, even if she had in fact not violated her relevant intellectual obligations). Thus a version of Pascalianism arguing for the deontological adequacy of a mode of positive abortive indirect control (suspension, rather than prevention of belief) is in equal trouble. Such strategies simply fail as serious candidates for underwriting standard ascriptions of epistemic blameworthiness.

10.5. THE LIMITATIONS OF DOXASTIC PASCALIANISM

The argument of Section 10.4 leaves the Pascalianist with the option of pursuing a strategy not involving appeal to counterfactual conditions. In particular Alston’s conception of epistemic/intellectual blameworthiness, even in its refined version, needs to be redressed without the appeal to counterfactuals, e.g. as I have preferred it, in terms of causal relations between exercises of doxastic control and formations of epistemically undesirable beliefs. I shall not spell out such a changed version of Alston’s thesis here. Rather, employing the taxonomy of modes of indirect doxastic content control introduced in Section 7.3, I shall examine the general applicability of a

pascalianist strategy for accounting for particular cases of epistemic blameworthiness. Since my preferred taxonomy allows for four possible modes of indirect content control (positive/genetic, genetic/abortive), four pascalianist strategies are feasible. One possible account (filling in the relevant clause of $\text{Blame}_{\text{epist, control}}$) would go:

An agent is blameworthy for her holding the belief that p *if* her belief that p is epistemically undesirable and

The agent has exercised her positive genetic indirect content control so as to form the belief that p , and she is not appropriately excused for doing this.

This account must immediately face severe restrictions on its domain of applicability. As hinted at in Section 10.1, arguably an agent's primary means of exercising Pascalian – i.e. indirect content-directed – control over her beliefs are the performance of actions that influence either her cognitive dispositions or her evidence-gathering activity. However, these are normally instruments far too blunt to single out a unique propositional content as a candidate for belief or even delimit an enumerable set of such contents.

The reason for this is not hard to see: It simply is beyond human abilities to generally foresee which *particular* beliefs will arise from our belief-influencing actions, let alone which particular beliefs will be influenced by them.⁴³⁰ Planning our actions in order to bring about particular future belief-formations is normally entirely futile.

In fact it is not even easy to generally foresee which *kinds* of beliefs will ultimately be influenced by one's belief-influencing actions. To recall once again our case of the educated racist, this agent might well have been in a position to predict that her persistent failure to stay attentive in her racial issues classes would probably influence her future beliefs on race issues. However, this predictive power is far short of the power she would need to actually single out a specific set of race issue contents, and guide her conduct towards believing one of these contents.

In fact, the only means of exercising a distinctive positive genetic Pascalian control available to a normal agent seem to be more or less benign processes of brain-washing (including hypnosis), either carried out by the agent herself, or by some other agent(s) under the control of the agent. Here the original "Pascalian" example comes in handy: It might be the case that I can single out a group of religious belief (say a list of dogmas) for future belief by regularly attending mass, keeping up close social relations with devoted believers, etc. It might also be the case that, assisted by a hypnotist, I can bring myself to believe, e.g. that cigarettes are highly repulsive. However, the domain of beliefs over which I can actually hope to exercise such control is highly limited. Says Alston:

It is very dubious that we have reliable long-range control over any of our beliefs, even in the most favourable cases, such as beliefs about religious and philosophical matters and about personal relationships. *Sometimes* people succeed in getting themselves to believe (disbelieve) something. But I doubt that the success rate is substantial.⁴³¹

Thus Pascalianism is now left to account for the rare cases, where an agent actually succeeded in indirectly bringing herself to believe a particular in an epistemically undesirable way and is not appropriately excused for doing this. Such a case might

imaginably concern a woman, Betty, who, for the dubious prize of a better conscience, wants to believe that whipped cream is part of a healthy diet. She goes to a hypnotist and asks him to induce in her that belief, which he successfully does (imaginably he also removes other beliefs that might cause her to lapse back in her old belief-state). Now, after the hypnosis, perhaps Betty may be judged epistemically blameworthy for holding the belief that whipped cream is a part of healthy diet on a doxastic Pascalianist account, given that she was not somehow excused for valuing the easy pleasure of a clean conscience above her service to the truth-goal at this instance.

Cases as Betty's, however, are obviously rare in the actual world, and cases in which, e.g. a *negative abortive* Pascalian strategy could apply are arguably even rarer: In such a case an agent would have to be blameworthy for holding an undesirable belief that *p* in virtue of having made it the case that, after she formed a belief that *p*, she sustained (did not again suspend) that belief. Perhaps some advanced form of hypnosis or social scheming could do the trick. However, I take it that the relevance of a negative abortive Pascalian strategy to actual cases of epistemic blameworthiness is negligible. The same sees to go for the positive abortive strategy. Here the agent's blameworthiness would have to hinge on the omission to exert an indirect doxastic influence of the above type.

Perhaps the negative genetic strategy fares slightly better. Consider, e.g. Betty's sister Catty, who, without the use of hypnotism, holds the unreasonable belief coveted by Betty, viz. the belief that whipped cream is part of a healthy diet. Now, in some farfetched scenario, perhaps we may in fact blame Catty for holding this belief with reference to the fact that she holds it because she omitted to pursue a course of action that would have prevented her from forming a belief with exactly the content that whipped cream is part of a healthy diet. It seems clear, however, that Pascalianism cannot hope to account for "main-stream" cases of epistemic blameworthiness where no distinct exercise of Pascalian control accompanied by a relevant foresight or blameworthy inadvertence to risk (see Chapter 13) is plausibly present.

10.6. THIRD-PERSON INDIRECT DOXASTIC CONTENT CONTROL

Above I have argued that modes of Pascalian control are strictly limited in their domain of deontological adequacy, because normal agents rarely stand in a position to control the content of their beliefs even indirectly. However, as I shall proceed to argue, a normal agent may well be in a position to exercise a certain amount of indirect control over the content of *other* agents' beliefs: Even though my intentions to induce in myself a certain belief are impotent, even I restrict myself to indirect means of belief-induction, my intentions to induce in another agent a certain belief may very well be successful.

In order to show this, I shall begin by laying down some terminology: The correct understanding of the term "testimony" in an epistemological context has been the subject of a great deal of recent controversy.⁴³² I shall here make a very relaxed and broad use of that term, meaning by a piece of "testimony" simply any utterance by an

agent intended to express a proposition, which has also been taken as the expression of a proposition by some audience.⁴³³

I shall say that an agent A judges another agent B *credible* with respect to a piece of testimony t uttered by B (in either writing, speech, sign language, etc.), understood by A to express the proposition p if, and only if A consciously believes that the fact that B uttered t gives her a good rationalizing reason to believe that p is true.

Now, I submit, a credibility-judgment is in the following sense an *inferential* judgment: When a normal agent judges another agent B credible with respect to a piece of testimony, she does this, *because* she (consciously or unconsciously) believes B's testimonial act to have certain properties *indicative* of B's credibility. Miranda Fricker has aptly termed such properties taken by an individual or a society to indicate a testifier's credibility, *working-indicator-properties*.⁴³⁴

The view that credibility-judgments are inferential in the above sense, has been defended at length by Paul Faulkner:

Our standing disposition to accept a given testimony, could then be explained as follows. We have a long history of dealing with testimony and on this basis we have learnt to distinguish testimonies into types some of which we take to be credible, some non-credible and others as credible as not. With this background of belief we are in a position to form the judgment, of any given testimony, that it belongs to a certain type, say type x. We believe that type x testimony is credible; we would assent to the statistical generalisation "Testimony of type x is probably true." We then directly infer that this testimony is credible.⁴³⁵

Note that Faulkner is far from stating that the "working-indicator-properties" (or the judgment that a certain piece of testimony has one or more of these properties) must be conscious to the relevant agent:

This is not to claim that audiences explicitly articulate the statistical syllogism when accepting, rejecting or suspending judgement in testimony. It is to claim that an audience's recognition of a testimony as a certain type in conjunction with the belief that such a type is credible provides an explanation of their doxastic disposition.⁴³⁶

I am not sure that the explanatory power of "the statistical syllogism" is at all enhanced by the claim that the agent would (under some unspecified circumstances) *assent* to it (see first quoted passage). Imaginably, an expression of this "syllogism" need not even be accessible to the relevant agent at all. Apart from this reservation, I entirely endorse Faulkner's position.

Now, given the above, it is fairly clear what an agent must do in order to induce in another agent A the belief that p: She must perform a testifying act that fits A's working-indicator-properties and which is taken by A to express the proposition p. This will then instantly induce in A the belief that she has a good reason to believe p, and the belief that p will instantly be formed in A under typical circumstances.

Working-indicator-properties are arguably sensitive to several features of the testifying act: First of all, in normal agents working-indicator-properties are highly sensitive to the content of the proposition taken to be expressed by the testimony: Agents typically have far more demanding standards for trusting statements about traffic rules than, e.g. the location of the nearest bath room. Further, the working-indicator-properties are typically sensitive to, among other things, the context of the testifying act and the properties of the testifier.

All this means that the difficulty of performing an action of inducing in another agent a propositional belief varies greatly, depending, among other things, upon who the testifying agent is, who her “victim” is, what her relation to the “victim” is, what acting skills she has, and what is the content of the proposition in question. However, whereas probably no one is clever enough to induce in a professional astronomer the belief that the Earth is flat by a single testifying act, it arguably requires minimal acting skills for an agent to induce in another agent a great deal of mundane propositional beliefs, e.g. concerning trivial past events, which her victim is unlikely to have any cognitive access to apart from the testimony offered.

I take this observation to underwrite two points: The first is this: Intuitively an agent could be epistemically blameworthy for holding a belief formed by testimony in the way sketched out above, if this belief-forming mechanism was in fact unreliable, and the agent has no appropriate excuse for falling victim to this mechanism. For example the educated racist from our oft-quoted example holds her blameworthy belief, because she took its propositional content to be expressed by an utterance of her favourite demagogue. Intuitively now, due to her background, the educated racist is not appropriately excusable for trusting the demagogue on such issues. That the demagogue actually intended to induce in the members of his audience the racist belief, and was successful in this regard, does not seem to detract from the epistemic blameworthiness of the educated racist. Thus, as a further embarrassment to doxastic voluntarism, it seems that a satisfactory account of epistemic blameworthiness must be able to account for cases of epistemically blameworthy belief, where the only relevant direct doxastic content control was exercised by other agents.

The second point is this: Arguably most of our beliefs somehow derive from testimony⁴³⁷: Now, presumably, among the vast number of testimonial beliefs upon which one of my present beliefs depends, some were unreliably produced, and in some cases the relevant testifier may even have been blameworthy in some sense for the unreliability of their process of formation. This is most clearly the case if she knowingly and intentionally tricked me into holding an unreliably formed belief. This past instance of blameworthiness may easily be relevant for the epistemic undesirability of one of my present beliefs, as the following example will show:

Suppose I am a newcomer to some academic discipline. Now a more experienced malicious co-student of that discipline testifies to me that a certain textbook is a highly reliable source of information on some central issue, well knowing that it contains almost nothing but dated nonsense. As a result I start reading this textbook and come to believe a number of absurdly false propositions. Now, who is to blame for my holding of these undesirable beliefs? It would seem that at least *some* blame attaches to my malicious co-student.

The same point applies to the case of the Morgan racists encountered in Section 1.3, where certainly Mr. Morgan cannot fail to be blameworthy in some sense for inducing the highly undesirable belief in his younger colleagues that his wife would consent to have violent intercourse with them.

I take this to show the following: The notion of epistemic blameworthiness evolved here is not that of *substantive* blameworthiness in a suitably modified Scanlonian⁴³⁸

sense (see Section 3.2): It is not the case that, because an agent is epistemically blameworthy for holding an undesirable belief, *all* legitimate blame for her holding of this undesirable belief befalls *her*. This highlights my point from Section 1.6 that the notion of epistemic blameworthiness evolved here is not the only viable sense of epistemic blameworthiness, although arguably the central one: The sense in which agents may be blamed in some sense related to epistemic goals for beliefs not held by themselves is not captured by my account of epistemic blameworthiness, which concerns only an agent's epistemic blameworthiness for her *own* undesirable beliefs.

However, the gist of doxastic Pascalianism may well be put to better use in exploring another notion of epistemic blameworthiness: The sense in which an agent may be blameworthy for *other* agents' epistemically undesirable beliefs. Here at last, after its long walk in the desert, indirect doxastic content control may finally enjoy its fair share of deontic glory, although for the present I shall not pursue this promising path further.

INDIRECT PROPERTY-DIRECTED DOXASTIC CONTROL OR PROPERTY PASCALIANISM

Abstract. In this chapter I finally present the modes of doxastic control, which I take to be most plausibly involved in standard ascriptions of epistemic blameworthiness: Negative and positive genetic indirect property-directed doxastic control. I lay out in Section 11.1 how these modes of doxastic control may in fact be invoked to underwrite ascriptions of epistemic blameworthiness. In Section 11.2 I acknowledge that I am not the first author to make suggestions of the above kind, insofar as at least John Heil and Hilary Kornblith have uttered remarks in a similar vein, Kornblith being the more out-spoken author. In Section 11.3 I then discuss how agents may in fact exercise the relevant modes of doxastic control. In the process I demonstrate how germane cases of epistemic blameworthiness, such as that of the educated racist from Section 1.2 above, may, under various circumstances, plausibly be interpreted in terms of these modes.

11.1. PROPERTY PASCALIANISM

In this section I shall finally present what I take to be the most versatile strategy of accounting for cases of epistemic blameworthiness: It shall become clear that standard cases like that of the educated racist may most plausibly involve the agent's negative genetic *indirect property-directed* doxastic control; i.e.: the control the agent can exercise so as to indirectly *prevent* the instantiation of the properties of her epistemically undesirable belief making for their epistemic undesirability. However, before going on to scrutinize such examples in greater detail, some initial points must be made.

In the case of property voluntarism (Chapter 9), I denied the deontological adequacy of the *abortive* modes of direct *property-directed* doxastic control on an early stage, due to the observation made in Section 7.3 that such modes of doxastic control cannot be exercised without exercising the corresponding abortive mode of *content-directed* doxastic control. With the demise of doxastic voluntarism, however, it was clear that these corresponding modes of content-directed control were not available to normal human agents.

In the case of indirect doxastic control, the situation is somewhat different: As we saw in Section 10.5, neither positive nor negative abortive Pascalian control is wholly inaccessible to a normal human agent. Pascalianism, however, had severe trouble in establishing the deontological adequacy of either of these modes of doxastic control for anything but highly bizarre or rare cases. Still, it holds (as argued at the end of Section 7.3) that a normal agent cannot exercise any mode of abortive *property-directed* doxastic control without exercising the corresponding mode of abortive *content-directed* control: Plausibly, an agent cannot suspend or sustain the instantiation of the properties making for the epistemic undesirability of a belief, without simply suspending or sustaining that belief altogether.

This means that, in cases where a mode of abortive *property-directed* doxastic control is deontologically adequate, the corresponding mode of abortive *content-directed* doxastic control must also be deontologically adequate. This again means that the exotic cases in which a mode of abortive indirect property-directed doxastic control is deontologically adequate have already been dealt with in Section 10.5. I shall therefore not examine such cases further.

We are thus left to consider the domains of deontological adequacy for modes of indirect *genetic* property-directed doxastic control. To stay in line with the terminology introduced earlier, I shall term a position that argues for the deontological adequacy of such modes of doxastic control “Property Pascalianism” to contrast it with doxastic Pascalianism of the type treated in Chapter 10. Property Pascalianism thus is committed to the following claim for a non-empty domain of undesirable beliefs (filling in the relevant clauses of $\text{Blame}_{\text{epist, control}}$ (Section 7.1):

An agent is epistemically blameworthy for holding the belief that p if her holding of the belief that p is epistemically undesirable and

1. By her means of indirectly influencing the instantiation of the properties making for the epistemic undesirability of her beliefs she has brought it about that her belief that p is epistemically undesirable.
2. She has no appropriate excuse for doing this.

11.2. JOHN HEIL AND HILARY KORNBLITH AS PIONEERING PROPERTY PASCALIANISTS

Only two authors, John Heil and Hilary Kornblith, have seriously discussed anything like a property Pascalianist strategy.

As we saw earlier (Section 9.2), John Heil has advanced remarks more than suggestive of property *voluntarism*. In some passages of Heil (1983), however, he seems to provide the outline of a property *Pascalianist* position, without however drawing the connection to the blameworthiness of singular beliefs:

In general we expect one another to undertake the task of information-gathering in an appropriate way. We condemn S for not looking more closely before reporting his sighting of a blue-crested finch; we chastise T for not listening more carefully to the gurgling sound produced by our De Soto [a type of car]. We find persons liable for things not done, for investigations not undertaken or undertaken with insufficient care. We enjoin one another (and of course ourselves) to “act responsibly” in this regard, I think, because acting in this way tends to make our beliefs more dependable, more reliable, more likely to be true. It is our status as intelligent belief-acquiring instruments that constitutes our status as epistemic agents. We do not choose what beliefs we acquire, but we have a hand in determining how we shall go about the business of gathering those beliefs. And this is a business that may be conducted responsibly or carelessly.⁴³⁹

Here *in nuce* we have the rejection of doxastic voluntarism, together with the insistence of the normative quality (reliability, dependability) rather than the content of our beliefs as the focus of epistemic deontic evaluation (the deontic nature of the evaluation is strongly suggested by Heil’s use of the terms “condemn” and “chastise”). However, Heil does not make the crucial suggestion that our epistemically undesirable beliefs may be epistemically blameworthy insofar as their undesirability

result from “careless” or “condemnable” (blameworthy) information-gathering. Also, as we shall later see, genetic property-directed doxastic control may in fact be exercised in ways not naturally described as information-gathering on a par with a careful search for a blue-crested finch.

Hilary Kornblith has presented, in far greater detail, a position that may plausibly be presented as a pioneering version of property Pascalianism. Kornblith flatly rejects doxastic voluntarism: “it is clearly true that beliefs are not freely chosen.”⁴⁴⁰ However he recognizes that this observation is far from bringing to its knee a (deontological) conception of epistemic justification:

When we ask whether an agent’s beliefs are justified we are asking whether he has done all that he should to bring it about that he have true beliefs. The notion of justification is thus essentially tied to that of action, and equally to the notion of responsibility.⁴⁴¹

Kornblith submits that, when we evaluate an agent’s intellectual conduct along the dimension of “responsibility,”⁴⁴² we examine “whether the extent to which the subject departed from the ideal was *his own fault*”⁴⁴³ (my italics), and further submits that “sometimes when we ask whether an agent’s belief is justified what we mean to ask is whether the belief is the product of epistemically responsible action.”⁴⁴⁴ It seems fairly clear, then, that the notion of justification Kornblith is after is at least very close to that of epistemic blamelessness (non-blameworthiness) in the sense of this study. In the following I shall assume that Kornblith is after such a notion, and that his considerations are for that reason immediately relevant to the present concerns.

Now, the actions Kornblith has in mind above seem to constitute exercises of property Pascalian (i.e. indirect property-directed doxastic) control over beliefs, as particularly emerges from the following passage:

Even in the case of simple perceptual beliefs, which are, admittedly, for the most part arrived at automatically, the agent’s actions play a role in the fine tuning of the belief acquisition process. An epistemically responsible agent must be on the uptake for defects in the process, and act to correct for them. While outright hallucinations are rare, there is a wide range of familiar perceptual illusions which the epistemically responsible agent comes to take account of. A certain amount of accommodation to misleading experience is itself automatic and not the product of free action. Nevertheless, there can be no doubt that one can self-consciously install in oneself a certain circumspection in circumstances when a mistake is likely to occur, such as in emotionally charged situations. Once this circumspection is acquired, more reliable belief-acquisition will occur in one automatically, and without any particular action on the agent’s part. Nevertheless, the presence of such automatic processes may often be traced to free action designed precisely to result in such processes, and when this is the case, it is to the agent’s credit; failure to take such action may be epistemically irresponsible.

In determining the justificatory status of a belief, we must thus look beyond the process which gave rise to it and look at the means by which the process itself was arrived at. This will involve, for example in perceptual cases, events long prior to the agent’s interaction with the object perceived.⁴⁴⁵

Insofar as a belief’s lack of justification on Kornblith’s account implies its epistemic blameworthiness, it seems highly plausible that Kornblith champions a property Pascalianist strategy, even if he does not make it absolutely clear that only epistemically undesirable (in this case, specifically: unreliably formed) beliefs may lack justification on his account⁴⁴⁶. It seems clear that an unreliably formed belief on his conception may be blameworthy (unjustified) because its unreliable formation

resulted from a failure to perform certain belief-influencing actions that the agent should have performed in the past, and that these omissions may be prior to the belief-forming process, narrowly construed. This is enough to render Kornblith a property Pascalianist, although he does not discuss the possibility that an agent may be appropriately excused for her failure to perform the “irresponsible” actions appealed to in an evaluation of her epistemic blameworthiness for holding some belief.

11.3. HOW TO EXERCISE INDIRECT PROPERTY-DIRECTED DOXASTIC CONTROL

Aided by Kornblith and Heil, I shall now proceed to discuss how a normal agent may in fact exercise some measure of indirect property-directed control over her beliefs, thus bringing into a clearer focus how property Pascalianism could in fact hope to underwrite actual ascriptions of epistemic blameworthiness. I take it here that the ways in which a normal agent might exercise property Pascalian control over her beliefs fall in four broad, not entirely discrete groups:

1. Influence on cognitive dispositions.
2. Influence on reasoning processes.
3. Influence on information-gathering activities (*cf.* Heil).
4. Influence on cognitive situatedness.

I shall now aim to make out in some detail which means of property Pascalian control is available to a normal agent within each of these domains, and how this may matter to the epistemic blameworthiness of her undesirable beliefs.

(1) Influence on cognitive dispositions. With the account of epistemic undesirability adopted here, the important kind of influence, which an agent may exercise over her cognitive dispositions, is the influence bearing on the epistemic undesirability of her future beliefs. Earlier I have adopted the empirical psychological thesis that, if an agent judges (has the belief) that she has good rationalizing reasons to hold the belief that *p*, this is typically causally sufficient to induce in her the belief that *p*. However, I believe that all doxastic events of a normal agent are governed by such reasonability judgments: If an agent judges about a certain proposition that she does not have good reasons to believe it, this is normally causally sufficient to prevent her from believing it; If an agent judges about a belief that she holds that she does not have good reasons to hold it, this is normally causally sufficient to bring her to *suspend* that belief; Finally, if an agent judges about a belief that she holds that she has good reasons to hold it, this is normally causally sufficient to bring her to *sustain* (not suspend) that belief.

Given these claims it seems clear that the desiderata of *adequate basing* and *reasonableness* (see Section 5.3) must have a central position in the present context: If an agent could indirectly influence her belief-forming cognitive dispositions such that her reasonability judgments actually track good rationalizing reasons, that is: if she judges that she has good reasons to hold a belief if, and only if, she actually has good

reasons to hold it, she has effectively *prevented* herself from forming beliefs that violate the adequate basing and reasonableness desiderata; the last insofar as she monitors her beliefs persistently. Further, she has now done everything she could possibly have done to secure that, if her beliefs are based on reasons, they are also truth-conducively formed. If they should still happen not to be truth-conducively formed, surely she cannot plausibly lack an appropriately excuse for this: Her management of her resources of indirect doxastic control simply appears to be impeccable.

The central question, however, remains the following: What can an agent do in order to make it the case that her cognitive dispositions promote the goal that her future beliefs are not epistemically undesirable. I believe the crucial notion here to be *education*,⁴⁴⁷ either by self and others. Quoting a phrase of Kornblith's from the passage above, education is what is normally "designed precisely to result in such [reliable] processes [of belief-formation]."⁴⁴⁸ This is plausibly why cases of *educated* agents, like "the educated racist," holding blatantly epistemically undesirable beliefs make particularly compelling cases of epistemic blameworthiness. A basic way, in which an agent could indirectly prevent the epistemic undesirability of her future beliefs, is by actively aiming to benefit from the educational opportunities available to her.

Thus, one way that Property Pascalianism may hope to account for the epistemic blameworthiness of the educated racist from Section 1.2, is by pointing to her missed educational opportunities: The fact that despite the educational opportunities available to her in college, she still was not really educated enough to resist the racist demagogue. The educated racist, so property Pascalianism may claim, has inexcusably failed to exercise her negative genetic property Pascalian control by not actively benefiting from the educational opportunities made available to her, since, had she so benefited, her cognitive dispositions would have changed so that she would not have formed an epistemically undesirable belief when faced with the testimony of the demagogue. Insofar as the educated racist is really epistemically blameworthy, I believe this negative property Pascalianist strategy to constitute perhaps the most appealing account of her epistemic blameworthiness.

(2) *Influence on reasoning processes.* Earlier (Section 8.3.2.1) we considered Carl Ginet's example of the juror Sam who had to reach a verdict about the guilt of a defendant.⁴⁴⁹ We saw that even if Sam reached a verdict, he had hardly "decided" for this verdict in anything like the sense in which he could effectively decide, e.g. to raise his arm. Sam had absolutely no active say concerning the propositional content of the belief that he eventually came up with.

Still, Sam's choice of engaging in a focused process of deliberation certainly affected the epistemic quality of his resulting verdict in the case at hand. By engaging in this process he helped prevent, e.g. that his beliefs would fall prey to his emotional attitudes towards the defendant. One might imagine that in his process of deliberation Sam attended closely to the beliefs he already held, trying to perceive if certain other beliefs were inferentially warranted by these beliefs. It is not hard to see how this type of indirect doxastic influence can be brought to bear on the epistemic blameworthiness of a belief. In terms of the educated racist example: If the content

of the educated racist's belief in fact contradicted certain of her other beliefs, and she would have seen this, if only she had given the matter a more focused attention, her lack of consideration would be a plausible candidate for the source of her epistemic blameworthiness for holding the racist belief.

(3) *Information-gathering.* This is my chosen label for the third bulky class of means, by which an agent may indirectly influence the epistemic quality of her future beliefs: I am here referring to roughly the same class of activities, which William Alston has termed "activities that bring influences to bear, or withhold influences from, a particular situation involving a particular candidate, or a particular field of candidates, for belief."⁴⁵⁰ Consider again the case of the educated racist as construed above. Certainly this agent, by moving around in her environment and making observations could have collected a lot of information relevant to her evaluation of the testifying situation that made her form the racist belief; features relevant to the reasons she might take herself to have for judging the speaker credible in the relevant situation. These information-gathering activities could well go back several years, where the racist would, e.g. have had the opportunity to gather information about the demagogue and his history as well as information about racial issues.

Suppose, now, that the cognitive disposition of the educated racist at the time she faced the testimony of the demagogue was in fact *not* so block-headedly credulous as stipulated earlier above. In particular, suppose that her working-indicator-properties (see Section 8.4) for trusting testimony of the type testified by the demagogue were so that she would not have formed an undesirable belief in the given situation, if only she had been aware of certain background information, say that the demagogue had an impressive past record of public lies. Now, in this scenario, if the educated racist has no appropriate excuse for not having gathered this information, this failure could arguably render her epistemically blameworthy for holding the undesirable racist belief on a negative genetic property Pascalianist strategy.

(4) The control an agent has over her *cognitive situatedness*, is easily overlooked and, to my knowledge, has not been addressed in the literature on deontic epistemic normativity. An agent's exercise of this control, however, might be highly relevant to the present context. For instance, the educated racist, if she had not been present at the demagogue's speech and stayed home instead, would never have formed her undesirable racist belief. Thus, if it could be argued that, given her credulous cognitive dispositions, she was not appropriately excused for being present at the demagogue's speech, this in itself could account for the epistemic blameworthiness of her racist belief.

The way that this type of indirect influence may act as a source of epistemic blameworthiness can perhaps be brought into a clearer light by considering a practical analogy. Suppose that Carla is a woman severely sensitive to intoxication: Normally she is gentle-mannered but as soon as she enjoys even the smallest sip of wine, she gets physically violent. Now suppose that Carla is invited to an official party where the drinking of wine is mandatory. Arriving at the party, she is unable to resist drinking

a glass of wine, and horrible mayhem results. If Carla is blameworthy for the damage she inflicts, it seems plausible that the primary locus of blame here is neither her failure to improve her unfortunate behavioural reactions to intoxication, nor the way in which she drinks the wine or the quantity she ingests. Rather, it seems natural to blame her with reference to the fact that she showed up at the party in the first place, given that she lacks a compelling excuse for not staying at home with a cup of tea. This situation closely mirrors that of the blameworthy racist, where the source of the racist's epistemic blameworthiness could perhaps be tracked to the very fact that she showed up at the demagogue's speech, given that she had no compelling excuse for not staying away.

However, before an agent's exercise of property Pascalian control in any of four ways listed above (or any mode of doxastic control for that matter) can underwrite an ascription of epistemic blameworthiness, I shall now proceed to argue, at least two further requirements must be met. These requirements, however, are both fraught with difficulties: First, the action or omission constituting the relevant exercise of doxastic control must itself be blameworthy: Unless it be the case that the agent *should/should not* have performed the actions/omissions constituting the relevant exercise of doxastic control, and that she had no appropriate excuse for not performing/performing these actions/omissions, no appeal to these belief-influencing actions or omissions can render the agent epistemically blameworthy. Second, the performance of these actions/omissions must stand in an appropriate relation to the later holding of the epistemically undesirable belief under deontic evaluation. In particular, I shall argue, some relevant kind of *foresight* or *blameworthy inadvertence to risk* must have been present. Discussing these requirements will provide the final elements I need in order to present a substantial analysis of the notion of epistemic blameworthiness.

INTELLECTUAL OBLIGATIONS

Abstract. In this chapter I defend the relevance of the notion of an intellectual obligation to the issue of epistemic blameworthiness. I conceive of an intellectual obligation as an obligation to perform a belief-influencing action or omission constituting an exercise of doxastic control. In Section 12.1 I argue that in order to render an agent epistemically blameworthy for her holding of some belief, it must needs be argued that in bringing about the undesirability of that belief she violated an intellectual obligation, lest she be appropriately excused for the undesirability of that belief. In Section 12.2 I make a fundamental distinction between two types of intellectual obligations based on their different sources. In Section 12.3 I proceed to argue that the source of so-called *institutional* intellectual obligations is not that hard to account for, compared to the source of so-called non-institutional intellectual obligations. I argue that a number of attempts to ground such obligations universally in terms of, e.g. truth-oriented desires do not resist closer scrutiny. As a result, even if the notion of *intellectual vice* may offer some assistance, in general the source of intellectual obligations is not easy to specify.

12.1. THE IMPORTANCE OF INTELLECTUAL OBLIGATIONS

If an action or omission is to bring about the epistemic blameworthiness of a belief, whose epistemic undesirability is a consequence of that action or omission, that action or omission must in itself constitute a violation of an obligation under the given circumstances. This can be seen thus:

On $\text{Blame}_{\text{epist, control}}$ (see Section 7.1) an agent cannot be epistemically blameworthy for holding an epistemically undesirable belief, whose undesirability results from an exercise of her doxastic control, if she has *an appropriate excuse* for that exercise. Now, if in a particular case, the relevant doxastic control was not something, which the agent was obliged *not* to exercise, the agent may simply pass the appropriate excuse for the undesirability of her belief that, even though it resulted from an exercise of doxastic control, that control was not something, which she was obliged *not* to exercise. If, however, the agent is indeed blameworthy for the undesirability of her belief due to her non-performance of a belief-influencing action or omission⁴⁵¹ as a consequence of which the belief came out epistemically undesirable, it seems natural to say that the agent stood under an “intellectual obligation”⁴⁵² to perform that action or omission.

The point might gain weight from considering an example from the practical realm: Suppose I have brought about the end of my department’s automatic coffee maker by pushing its on-button: The machine was malfunctioning, such that simply turning it on immediately short-circuited it. However, even though by pushing the button I brought about the unhappy end of the coffee maker, I can hardly be blamed for it: I can simply pass the appropriate excuse that I did nothing I was in any sense obliged

not to do, insofar as the machine was not equipped with any warning signs, I was perfectly entitled to use it etc.

Analogously, it now seems clear that on $\text{Blame}_{\text{epist, control}}$ any ascription of epistemic blameworthiness to an agent need to appeal to her violations of intellectual obligations in the sense introduced above. The crucial question then becomes: Do normal agents stand under intellectual obligations to perform or omit certain belief-influencing actions at all? In the present chapter, I shall provide an affirmative answer to that question and attempt in some detail to track the source of such obligations under various circumstances.

In the literature, intellectual obligations have not loomed large, despite the recent surge of interest in epistemic deontology. Rather, when authors have talked about obligations within epistemological contexts, they have focused narrowly on supposed “epistemic obligations” to hold only beliefs that are in some sense justified or they have simply identified beliefs justified in a deontic sense with beliefs whose holding does not violate any “epistemic obligations.”⁴⁵³ Such obligations are clearly not of the kind sought after here, since they apply to belief-holdings, rather than belief-influencing actions or omissions.

A very clear example of the tendency to focus narrowly on the notion of justification and thereby miss out on the importance of intellectual obligations can be found in Haack (2001),⁴⁵⁴ where W.K. Clifford, the most out-spoken champion of intellectual obligations one could ask for, is interpreted as being concerned only with the justification of beliefs rather than with belief-influencing actions. This in spite of the fact that the headline of a central part of Clifford’s celebrated 1877 essay “The Ethics of Belief” reads “The Duty of Inquiry”! In the run of the present chapter I shall make it clear that Haack’s reading of Clifford simply misses the point.

12.2. INSTITUTIONAL AND NON-INSTITUTIONAL INTELLECTUAL OBLIGATIONS

The predominant focus in the literature on so-called epistemic obligations pertaining to belief-holdings is not that surprising. As shall emerge, it is far from easy in general to point out why an agent stands under an intellectual obligation to perform some belief-influencing action or omission, let alone to establish the relevance of particular intellectual obligations to a particular evaluation of epistemic blameworthiness.

Perhaps the most obvious place to look for intellectual obligations is within *institutional settings* where either an agent has taken such obligations upon herself, or more or less explicitly others have imposed them on her. I shall term intellectual obligations imposed in this way “institutional intellectual obligations.” Not surprisingly, recent authors, when more or less obliquely touching upon the issue of intellectual obligations, have shown a conspicuous taste for courtrooms, classrooms and medical clinics in their examples. For example Susan Haack has acknowledged that indeed there are cases where an agent is held “culpable” for “not having cultivated better judgment” and

There are circumstances where this is appropriate – for example in some cases where it is this person’s (this doctor’s, this lawyer’s, this juror’s, this academic’s) particular responsibility to know about the matter at hand ...⁴⁵⁵

Holly Smith has shown a similar emphasis: She employs the notion of a “benighting act” as “an...act, in which the agent fails to improve (or positively impairs) his cognitive position.”⁴⁵⁶ As she later tracks the blameworthiness of “unwitting acts” based on a “culpable ignorance” to their roots in blameworthy benighting acts, it seems pretty clear that at least some benighting acts are simply belief-influencing actions (or omissions⁴⁵⁷) standing under intellectual obligations in my preferred terminology.⁴⁵⁸ Smith’s most elaborate example of a benighting act is located within institutional settings: It involves the realistic case of a doctor who exposes a baby to an unnecessarily high concentration of oxygen, which, however, he believes to be harmless at the time. Nevertheless the baby loses its sight as a result of the treatment. Smith submits:

Perhaps the doctor should have known that high oxygen enrichment would induce blindness: the latest issue of his medical journal described a study establishing this effect and recommending the use of lower concentrations as equally effective for respiratory problems. The doctor should have read this journal, and if he had done so, would have realized he ought to use less oxygen.⁴⁵⁹

Here of course, the benighting act is the doctor’s omission of reading the relevant journal. However, Smith also offers explicit examples of benighting acts, which do not seem good candidates for being blameworthy on the basis of institutional standards. She, e.g. discusses the case of a slightly near-sighted woman, who performs the benighting act of driving in her car on a foggy morning without putting on her glasses, which, as Smith makes clear, she is not legally required to wear. As a consequence the woman accidentally kills a young pedestrian.⁴⁶⁰ Another of Smith’s examples involves a man, who forgetfully leaves home in the family car in spite of his promise to let his wife use it. Here, according to Smith, the benighting act consists in the man’s failure to “ask himself” to bring to mind his wife’s requests.⁴⁶¹

Further below I shall point to other cases of intellectual obligations, which are not easily construed as institutional. At this stage, however, it is time to get at least partly in the clear regarding the sources of intellectual obligations of either kind.

12.3. THE SOURCE OF INTELLECTUAL OBLIGATIONS

To approach this problem, I shall first argue that intellectual obligations to perform or omit belief-influencing actions may in fact obtain within proper institutional settings. Here the most obvious type of intellectual obligation obtains when an agent has chosen or has been obliged to perform a certain professional function, e.g. that of a medical doctor, that bears with it obligations to perform certain belief-influencing actions or omissions.

Being a medical doctor may, e.g. include an obligation to participate in certain mandatory courses specified by relevant governmental institutions. These are clearly obligations pertaining to actions influencing cognitive belief-forming dispositions. It may also include obligations to show a special care and alertness in diagnosing patients

and run through certain specified double-check routines even after forming a belief on an issue.⁴⁶² Such obligations pertain to reasoning processes. Further a doctor may be obliged to stay up-to-date on research within her field of specialization, an obligation pertaining to information-gathering. On the other hand, it is not clear that the function of a doctor brings with it any obligations pertaining to cognitive situatedness, but other professions may well do so: Consider, e.g. a secret agent. Such an agent may well be obliged to stay away from certain briefings of other agents, because, since she is not an initiate in this other context, she would probably misunderstand the codes used and form epistemically undesirable beliefs about the operation of the service as a result.

These observations bring into play the four basic ways of exercising modes of indirect property control over one's beliefs lined up in Section 11.3 above. However, institutional intellectual obligation may plausibly also concern exercises of *direct* property control (see Chapter 9):

Other functions that involve intellectual obligations may be that of a juror or a proscribed soldier. The soldier and the juror may well be obliged to apply care in their belief-forming processes, e.g. they are not allowed to wilfully distort their vision and hearing when scouting for enemy tanks or listening to defence testimony. Similar obligations may apply to our doctor from above.

It is important to notice that institutional intellectual obligations do not apply to the *role* of a medical doctor, juror, or soldier as such, but to an agent *insofar as* she occupies the role of a medical doctor, juror, or soldier. Thus, violations of such obligations may certainly impose blame, not on "the role,"⁴⁶³ but on the agent occupying the relevant role. If, as in Holly Smith's example, a doctor inexcusably violates her professional intellectual obligation to read a certain journal and thereby comes to hold a certain belief based on inadequate evidence, this could provide the back-ground of a sound property Pascalian account of the epistemic blameworthiness of that belief: Had she in fact read the journal that she was intellectually obliged to read, she would not have formed an inadequately based belief. Further, if in acting on her inadequately based belief, the doctor causes harm to a patient, plausibly (see Section 1.3) she is also blameworthy for her harm-doing.

However, as we shall shortly see, all cases of blameworthy harm-doing on the basis of "honest mistakes" cannot be rooted in blameworthy violations of intellectual obligations within institutional settings. Further, perhaps the most conspicuous institutional locus of belief-influencing action, the general system of education, is primarily attended by children, who have not themselves chosen to attend school or been obliged to do so as legal subjects. Rather they have effectively been forced to school by their parents, who bear the legal responsibility for them. Thus schoolchildren's intellectual obligations are not immediately comparable to those of a doctor who has herself chosen her profession with its inherent obligations or a juror who stands under these obligations according to the laws of her society.

Consider again the *Morgan* case of Section 1.3. In this case three men forced a woman, Mrs. Morgan, to non-consensual intercourse acting on the belief that she consented to the intercourse and their use of physical force. This belief was credulously formed on the testimony of the woman's husband, who assured them that her signals

of dissent during the intercourse were to be interpreted as signs of her “kinky” pleasure. I argued that if the men were to be held blameworthy for the harm they inflicted on Mrs. Morgan, as intuitively they should be, this blameworthiness must have its source in the epistemic blameworthiness of their belief that Mrs. Morgan consented.

Now, as established on this stage of my argument, this belief cannot be epistemically blameworthy unless it is epistemically undesirable (it certainly seemed to be highly unreasonable) and its epistemic undesirability stems from one or more inexcusable violations of intellectual obligations. However, it is far from obvious that any relevant *institutional* intellectual obligations were violated by the *Morgan* rapists, even if part of their block-headed cognitive disposition originated in their loafing in secondary school: As pointed out above, it is far from evident that schoolchildren stand under any institutional intellectual obligations in the sense that jurors and doctors do. Also, the rapists’ duties as legal British citizens did not explicitly include any belief-influencing actions: Legal duty forbids inflicting certain kinds of harm; it does not explicitly forbid stupidity or credulity.

It would seem then that the undesirability of the fateful belief of the *Morgan* rapists (the unfounded belief that Mrs. Morgan consented to violent sex) cannot be traced to their violations of *institutional* intellectual obligations. The following dilemma therefore arises: Either we must at this stage acquit agents like the *Morgan* rapists of epistemic blameworthiness or we must account for the epistemic blameworthiness of such agents in terms of violations of *non-institutional* intellectual obligations. I shall now take grasps with the second horn of this dilemma. I therefore need an adequate understanding of the source of non-institutional intellectual obligations. In order to reach such an understanding I will first rule two misguided suggestions for the location of their source:

As we saw in Section 12.2, Holly Smith has offered a number of examples of agents standing under intellectual obligations of a perspicuously non-institutional kind. One of these concerned a near-sighted woman, who, due to her “benighting act” of driving without her glasses, was blameworthy for hitting a pedestrian with her car. It would now have been highly interesting, had Smith passed a verdict on this type of case, pointing out exactly why the near-sighted woman should be held blameworthy for her “benighting act.” However, all Smith offers is a general claim that acts are *morally* blameworthy, if a “reprehensible configuration of desires and aversions” gave rise to their performance.⁴⁶⁴ Now, it is far from clear what “reprehensible desires and aversions” are present in the benighting acts from Smith’s examples. Smith offers no guidance here, only she seems to presume that in her example of the near-sighted woman, her “willingness to leave home without the glasses” was reprehensible enough to render her act morally blameworthy.⁴⁶⁵

However, first it is not easy to see that this willingness is an aversion or desire in any standard sense of the terms at all. Second, even if such willingness is indeed “reprehensible” under the given circumstances, this observation does not seem to provide a good reason for holding the relevant benighting act morally blameworthy. Smith arguably has things the wrong way around here: Actions are not blameworthy because certain “reprehensible” desires gave rise to them. Rather, if any desires are

“reprehensible” it can only be because they (typically) give rise to blameworthy actions. Third, hypostasing the “reprehensibility” of certain desires or motivations is simply uninformative: Since it is far from intuitively striking that, e.g. the slightly near-sighted driver’s willingness to leave home without her glasses should be judged reprehensible at all, insofar as she was not legally required to wear them, Smith owes us an account why we should think it so, an account that she does not give. I will conclude then that Smith fails to shed any real light on the source of intellectual obligations, neither inside nor outside of institutional settings.

Hilary Kornblith has made an approach to the source of non-institutional intellectual obligations very similar in kind to Holly Smith’s. Rather than invoking a general notion of reprehensible desires, Kornblith focuses more narrowly on a lack of *truth-oriented* desires.

In Kornblith’s terminology the notion of an agent’s “epistemically responsible conduct” plays approximately the role that the notion of an agent’s “fulfilment of her intellectual obligations” play in my preferred terminology: As we shall shortly see, an agent is [deontologically] justified (blameless) in Kornblith’s sense in holding a belief, only if that belief was not a result of the agent’s prior “epistemically irresponsible” conduct (see also Section 11.2). Now, Kornblith conceives of “epistemically responsible” conduct in terms of certain truth-oriented *desires*:

*An epistemically responsible agent desires to have true beliefs, and thus desires to have his beliefs produced by processes, which lead to true beliefs; his actions are guided by these desires. Sometimes when we ask whether an agent’s belief is justified what we mean to ask is whether the belief is the product of epistemically responsible action, i.e. the product of action an epistemically responsible agent might have taken.*⁴⁶⁶

In my preferred terminology, Kornblith thus seems to defend roughly the following thesis: A belief-influencing action or omission constitutes a violation of an intellectual obligation, if, and only if, it was *not* guided by truth-oriented desires. At least it is clearly the case that on his view, if a belief is epistemically blameworthy (not justified), its blameworthiness may derive from prior belief-influencing actions (or omissions) not guided by truth-oriented desires. Kornblith offers the following example of an agent, who is supposedly epistemically blameworthy for (not justified in) holding a belief, due to its being the result of belief-influencing actions not guided by the right truth-oriented desires:

Consider the case of Jones. Jones is a headstrong young physician, eager to hear the praise of his colleagues. After Jones reads a paper, a senior colleague presents an objection. Expecting praise and unable to tolerate criticism, Jones pays no attention to the objection; while the criticism is devastating, it fails to make any impact on Jones’ beliefs because Jones has not even heard it. Jones’ conduct is epistemically irresponsible; had Jones’ actions been guided by a desire to have true beliefs, he would have listened to the objection. Since his continuing to believe the doctrines presented in his paper is due, in part, to this epistemically irresponsible act, his continued belief is unjustified.⁴⁶⁷

For the sake of argument, let us grant here that, if Jones had in fact paid careful attention to the objection of his senior colleague, Jones would have suspended his unreasonable belief that the doctrines of his paper were true. Still, I find it completely unconvincing that the truth of this conditional essentially hinges on Jones’ *desire*

to have true beliefs: At most it was contingently the case that, as claims Kornblith, “had Jones’ actions been guided by a desire to have true beliefs, he would have listened to the objection.” If Jones was sufficiently arrogant and guided by a desire to have true beliefs, he would probably just have ignored his colleague’s objection to an even higher extent, perhaps he would even have stuffed his ears for fear that his “pristine” body of belief should be tainted by the testimony of an intellectual inferior.

Now, it is hard to see why this arrogance should make Jones any less epistemically blameworthy (unjustified) for holding his unreasonable belief, even though, if he had been sufficiently arrogant and guided by truth-oriented desires, his belief would then have been the product of “epistemically responsible action” in Kornblith’s sense. However, as argued in Section 12.1, if Jones is indeed epistemically blameworthy, he must have violated some intellectual obligation in the past. Consequently the fact that an undesirable belief results from epistemically responsible action in Kornblith’s sense hardly suffices to render it epistemically blameless: Being guided by truth-oriented desires is not *sufficient* to satisfy one’s intellectual obligations.

Now, even worse for Kornblith’s conception of intellectual obligations, it does not even seem to be the case that the guidance of truth-oriented desires is *necessary* for the fulfilment of one’s intellectual obligations: It is entirely unclear, why it cannot be acknowledged that an agent can fulfil her intellectual obligations, without therefore being very susceptible to the motivational power of a desire to believe the truth. Perhaps in the “Jones” example, what Jones really needed to function better intellectually was not so much a desire to believe the truth as a desire to be on better terms with his colleagues, it being the latter desire rather than the former which would have motivated him to pay sufficient attention to the senior colleague’s well-put objections. However, once this point is recognized, it seems immediately implausible that an appeal to any single type of desire may shed any light on the nature of non-institutional intellectual obligations. This point, of course, should not be taken to imply that the source of such obligations might not be tracked to the value of an agent’s pursuit of the truth-goal (I shall in fact consider this option below). Only a universal grounding of non-institutional intellectual obligations in the notion of a truth-oriented *desire* has been challenged at this point.

Leaving Smith and Kornblith behind I shall proceed to discuss two alternative basic ways of grounding non-intellectual obligations: Either in the value of avoiding harm or in the value of an agent’s pursuit of the truth-goal. I shall not make any final decision between the rival approaches, but will restrict myself to pointing out important *caveats* in relation to both of them. Also, I shall argue, on pains of plausibility both approaches must acknowledge that non-institutional intellectual obligations are sensitive to the cultural embeddings of agents.

Let us consider first, how the notion of a non-institutional obligation may be grounded in the notion of harm-doing. Notice here that if an evaluation of epistemic blameworthiness is to appeal to a violation of a non-institutional intellectual obligation, the source of her intellectual obligations relevant to that evaluation cannot very well be tracked to the *particular* harm-doing which their violation brought about.

This is because it seems overwhelmingly plausible that, if a belief is epistemically blameworthy, it remains so, whether or not it actually prompts harmful actions or omissions. At least this seems to match our reactive attitudes: In the case of the educated racist, e.g. it does not seem to detract from the epistemic blameworthiness of her racist belief, if accidentally she should never come to harm someone on the basis of it.

We might then attempt to ground non-institutional intellectual obligations in the general *likelihood* of harm brought about by their violation. However, this approach does not seem very plausible either:

In fact it is not too hard to imagine a violation of an intellectual obligation, which is not likely to cause any significant harm at all. Consider, e.g. the case of the mathematician Dora. Dora believes that she has proved some highly esoteric theorem on which she has worked for many years. However, this belief is entirely induced by wishful thinking, and she has no good rationalizing reasons to hold it at all. Plausibly, now, Dora's belief may be epistemically blameworthy, if she violated due standards of care in not double-checking her proof as she should have, an omission of an exercise of property Pascalian control, had she not omitted it, she would not have held an undesirable belief with the said content. Nevertheless, far from bringing about any harm or even the likelihood thereof, Dora's violation of this intellectual obligation was actually a great benefit to her family, friends, and colleagues, who suffered greatly under her relentless obsession with the obscure theorem. Overall, then, the benefits resulting from Dora's violation of her intellectual obligation far outweigh the negligible harm it might ever bring about.

Faced with considerations like the above, maybe the best way to tie together non-institutional intellectual obligations and harm-doing is to opt for a *rule-based* approach, saying, e.g. that a belief-influencing action or omission constitutes a violation of a non-institutional intellectual obligation, if it constitutes the violation of a rule for belief-influencing actions or omissions, such that, if all agents always adhered to this rule, a minimum of harm-doing would result.

On this approach the mathematician as well as a harmless educated racist could be said to have violated non-institutional intellectual obligations, insofar as, not having double-checked the supposed proof or having remained credulous, they have not adhered to rules of the kind in question and this even if the particular rule-violations of either agent did not in fact bring about harm-doing or even the likelihood thereof.

A rule-based approach like the above, however, suffers from some general problems, not least concerning how to carve out rules of the relevant kind. Perhaps, though, an even more disturbing problem arises from the observation that different cultures recognize very different rules for intellectual conduct. As pointed out by William Alston, e.g. cultures look very differently upon an agent's actively questioning the traditional beliefs of her community:

[W]hat can reasonably be expected of a subject with respect to, for instance, critical examination of beliefs and their bases will differ across cultures. We require adults in our culture to be critical of "tradition," but this is a relatively new phenomenon, given the time humans have been on earth; it cannot be reasonably expected of anyone in any society.⁴⁶⁸

Now, suppose that, if in an ancient culture, say the Aztec, everyone had been highly critical and inquisitive regarding the truth-value of some traditional myths, eventually less harm-doing would have resulted (say, less violence and blood-shed). Even leaving aside the problem of imposing a cross-cultural standard of harm-doing (To an ancient Aztec, giving up human sacrifices to the gods would arguably count as a great harm.), is this observation enough to impose on any Aztec the intellectual obligation actively to question the cherished myths of her society? At least to this author, this seems overly zealous. Remark that I do not, e.g. claim that the fact that an agent is an ancient Aztec in itself constitutes a reason for holding her liable to different intellectual obligations than pertain to a modern Westerner. This would have been a very strong form of cultural relativism, which I do not wish to endorse. Rather, I simply take it that specific circumstances like a highly conservative cultural embedding may sometimes alleviate an agent of intellectual obligations to, e.g. question the traditional beliefs of her society, which may well apply to agents under different cultural circumstances.

Before returning to the problem of cultural embedding, let us turn to the second of the alternative conceptions of non-institutional obligations lined out above; the conception in terms of the independent value of an agent's pursuit of the truth-goal. W.K. Clifford is a particularly out-spoken proponent of this approach:

We all suffer severely enough from the maintenance and support of false beliefs and the fatally wrong actions which they lead to, and the evil born when one such belief is entertained is great and wide. But a greater and wider evil arises when the credulous character is maintained and supported, when a habit of believing for unworthy reasons is fostered and made permanent ... (...) ... If I let myself believe anything on insufficient evidence, there may be no great harm done by the mere belief; it may be true after all, or I may never have occasion to exhibit it in outwards acts. But I cannot help doing this great wrong towards Man that I make myself credulous. The danger to society is not merely that I should believe wrong things, though that is great enough; but that I should become credulous, and lose the habit of testing things and inquiring into them; for then it must sink back into savagery.⁴⁶⁹

Earlier in his 1877 essay Clifford has made it clear that a "bounden duty to mankind" to "guard the purity of his beliefs with a very fanaticism of jealous care" is not only incumbent upon "the leader of men, statesmen, philosopher, or poet" but also upon "every rustic that delivers in the village alehouse his slow, infrequent sentences[!]"⁴⁷⁰ It thus seems safe to conclude that, according to Clifford, *all* agents do *always* stand under non-institutional intellectual obligations in the present sense, and that these obligations at least pertain to belief-influencing actions of the three first types listed in Section 11.3: In the passage above Clifford alludes to our influence on cognitive dispositions "a habit of believing on unworthy reasons," our influence on reasoning processes "testing things," as well as to our influence on information-gathering "inquiring into them." Further, with his unrelenting stance on these issues he probably would have explicitly included obligations pertaining to cognitive situatedness, had he recognized them as a distinct category. It should be clear, then, that Haack's interpretation of Clifford as concerned narrowly with beliefs and their justification mentioned at the end of Section 12.1 is far too narrow.

A plausible interpretation of the reasons Clifford gives as to why we generally stand under intellectual obligations seems to be the following: We stand under intellectual

obligations in all circumstances because we owe a certain duty to “Man,” “the fate of mankind,” or “the sacred tradition of humanity”⁴⁷¹ (as Clifford likes to put it). This tradition he regards as inherently valuable: “the very sacredness of the precious deposit imposes upon us the duty and the responsibility of testing it [the tradition], of purifying and enlarging it to the utmost of our power.”⁴⁷²

In slightly more earth-bound terms, what Clifford seems to value so highly is the tradition of Enlightenment science, especially “truths which have been established by long experience and wanting toil, and which have stood in the fierce light of free and fearless questioning.”⁴⁷³ At the bottom-line, then, it seems that Clifford is most plausibly interpreted as anchoring the existence of intellectual obligations in the demand society has on each individual member to pursue the *truth-goal*: The goal of holding only true beliefs.⁴⁷⁴ This is a goal, Clifford clearly thinks, that each member of society can only aspire to reach by participating in a collective effort to promulgate the ideals and practices of the Enlightenment scientific tradition in even the lowliest stratum of life.

Clifford’s global grounding of intellectual obligations in our unrelenting duty to pursue the truth-goal is grand-mannered to say the least, perhaps also too grand-mannered to be taken seriously in the present cooler academic climate. However, as I shall point out further below, we might well be forced to consider quite lofty intellectual ideals if we are not willing to give up on the feasibility of ascribing epistemic blameworthiness in many compelling cases.

Clifford’s position may be labelled “deontic epistemological purism”: the hypostasis of epistemic values in no substantial need of vindication in terms of moral values as basic to epistemic deontology. An initial temptation in this vein might now be to say simply that an agent stand under an intellectual obligation to perform or omit a certain belief-influencing action, if this performance or omission makes her more likely to reach the truth-goal. However, this thesis seems overly strong. I wholly endorse James Motmarquet’s point that

For Einstein not to have persisted in his theoretical researches, say, in 1904, *merely* because these were not supported by the greater share of the physicists of his time would have marked a failure of intellectual courage.⁴⁷⁵

If Einstein had not persisted in his research, perhaps he would not have served the truth-goal as well as in fact he did. However, if he had given up his research at an early stage due to, e.g. material hardships, blaming him for not performing extremely virtuous actions like the courageous pursuit of groundbreaking scientific results in the face of poverty would have been a strange thing to do. This is plausibly because his failure to perform these actions could then hardly have been made out as an expression of intellectual *vice* either. A more plausible thesis thus seems to be the following: An agent stands under a non-institutional intellectual obligation to perform or omit a certain belief-influencing action, if her failure to perform or omit that action would constitute a *vicious* deviation from her ideal service to the truth-goal.⁴⁷⁶

Now, arguably a notion of intellectual vice also comes handy to the rule-based conception of non-intellectual intellectual obligations in terms of harm-doing presented

in brief out-line above: One adequate conception of the rules for belief-influencing actions or omissions involved in this conception could well be in terms of intellectual vice. A reason, why our conservative Aztec from above should go free of epistemic blame for her traditional mythological beliefs, could well be that even though her unquestioning attitude towards the mythology of her culture is not exactly a paradigm of intellectual adventurousness, it is not intellectually vicious in any straightforward sense either: Those beliefs are simply too ingrained in her general way of life that we could reasonably ask her to question them.

On the other hand, the credulous college student and the careless mathematician from above both seemed intellectually vicious in their omissions to improve belief-forming habits or double-check a supposed mathematical proof. These observations suggest that a rule-based conception of non-institutional intellectual obligations in terms of harm-doing would be well served with the following thesis resembling the one suggested for adherents of the independent value of the truth-goal above: A belief-influencing action or omission constitutes a violation of a non-institutional intellectual obligation, if it constitutes an *intellectually vicious* violation of a rule for belief-influencing actions or omissions, such that if all agents always adhered to this rule, a minimum of harm-doing would result.

What we then need is a general account of what constitutes a *vicious deviation* from an agent's ideal service to the truth-goal or adherence to a rule for belief-influencing actions or omissions. James Montmarquet's and Linda Zagzebski's thorough treatments of the notion of intellectual virtue⁴⁷⁷ are of help here only in the negative sense that they provide us with a fair guide-line as to where *not* to look for intellectually vicious actions/omissions: The lack of a certain intellectual virtue is not necessarily an intellectual vice: for example a lack of intellectual boldness, though unvirtuous, is not obviously vicious either.

Intellectual sloppiness and intellectual laziness with regard to important topics, however, may be two fair candidates for central intellectual vices. Thus, e.g. the educated racist, by her skipping her college class without an appropriate excuse, may have displayed a marked intellectual laziness with regard to a topic that was clearly important from a general perspective (the kind of topic we want members of our community to have epistemically desirable beliefs about). For that reason, it seems plausible to say that she stood under an intellectual obligation to attend to that class: On a conception guided by the truth-goal, we might say that she not only deviated from an ideal service to the truth-goal; but did so to a highly objectionable degree. On a conception guided by the notion of harm-doing we might plausibly claim that a rule requiring agents to maintain a general degree of intellectual curiosity would prevent much harm-doing, and that the educated racist's failure to adhere to that rule is objectionable enough to be qualify as vicious in the relevant sense.

Also, if we are to maintain the epistemic blameworthiness of the *Morgan* rapists for their unreasonable belief that Mrs. Morgan consented to sexual intercourse, we may very well have to hypostase a more general intellectual vice like lack of sufficient curiosity. Claiming that credulity is in itself an intellectual vice will not do much work here, insofar as the rapists' credulity (cf. the demise of doxastic voluntarism)

relevantly manifested itself as a certain belief-forming disposition rather than as one or more belief-influencing actions or omissions. If the rapists were to have avoided credulously trusting Mr. Morgan on the matter of the consent of his wife, perhaps they should have gathered more extensive evidence about typical female behaviour and marital relationships, such that their working-indicator properties (See section 10.6) regarding testimony on such matters would have changed. In omitting to do this, perhaps they displayed a vicious lack of intellectual curiosity relative to their over-all life-style.

Still, prospects for a general catalogue of intellectual vices are dim indeed. The present discussion could easily be augmented and very plausibly, another author may be able to offer a more penetrating treatment of the issue. However, we should hardly expect too much precision or exactness on a topic like this. It is more than tempting in the present context to invoke Aristotle's famous words that

We must be content, then, to speak of such subjects and with such premises to indicate the truth roughly and in outline, and in speaking about things, which are only for the most part true and with premises of the same kind to reach conclusions that are no better.⁴⁷⁸

We might, however, aspire to a greater amount of precision that has been achieved so far on matters of deontic epistemic normativity by discussing at some length the relevance of the notions of foresight and inadvertence to risk to the present concerns. This I shall proceed to do below.

FORESIGHT AND BLAMEWORTHY INADVERTENCE TO RISK

Abstract. In this chapter I discuss the notions of foresight and blameworthy inadvertence to risk. It is maintained that an agent's violations of her intellectual obligations can only matter to the epistemic blameworthiness of her beliefs, if those violations were performed with a relevant foresight or blameworthy inadvertence to the risk of epistemically undesirable consequences. In Section 13.1 I take upon myself to define the relevant notion of *foresight*, which turns out to concern an agent's acceptance that, as a result of her belief-influencing actions or omissions, an increase in risk obtains that certain occurrences will occur. In Section 13.2 I then discuss the complimentary notion of *inadvertence to risk*. Two types of such inadvertence are distinguished, but it is argued that they are not relevantly different from the present perspective. In Section 13.3 I take up a difficult discussion of the relevance of particular contents of foresights or inadvertences to evaluations of epistemic blameworthiness. I manage to set up only very rough relevance criteria. Section 13.4 discards some comments due to Larry Alexander, which may be taken as objections to the present argument. Still in Section 13.5 I recognize that the present discussion of foresight raises fundamental worries concerning some of my central claims. I present ways of overcoming these problems.

13.1. DEFINING FORESIGHT

As argued earlier (Section 10.3), the fact that an action is morally blameworthy and has harmful consequences does not make the agent morally blameworthy for these consequences, if she could not in any way have foreseen these harmful consequences at the time of action. Similarly, I shall now argue, no matter how many epistemically undesirable beliefs result from a violation of an intellectual obligation, none of these undesirable beliefs are epistemically blameworthy just in virtue of resulting from that violation.

For example “the educated racist,” even if standing under an intellectual obligation to attend her racial issues classes, can hardly be blameworthy for her later undesirable belief that resulted from her violation of her intellectual obligation to attend the class, *unless* she could at least have foreseen that her skipping this class would result in undesirable consequences of a relevant sort. Otherwise, blaming her for these consequences (e.g. the undesirable racist belief) would be like blaming the canned soup manufacturer from the example of Section 10.3 for the deaths he unwittingly and unforeseeably brought about, which seemed a very unjust thing to do.

The urgent questions then become: (1) When violating an intellectual obligation, which kind of foresight should an agent have or be required to have with respect to (the risk of) an undesirable belief resulting from that violation, if that violation is to ground an ascription of epistemic blameworthiness for her holding of an undesirable belief? (2) How is this foresight to manifest itself? The first of these questions is highly intricate, and I cannot attempt even the sketch of an answer without addressing the second question, as I shall now proceed to do. In order to achieve a maximum of clarity

in this entangled context, I will work out an answer in a moral context, before engaging with the questions narrowly relevant to the notion of epistemic blameworthiness.

Recent authors treating the relation between foresight and moral blameworthiness for consequences have generally been less than willing to explicate the notion of foresight involved in their considerations.⁴⁷⁹ Still, Michael Zimmermann, being an exception to this rule, has proposed the following definition:

P foresees at t1 (to some degree) that *e* [an event] will occur at t2 if and only P occurrently believes at t1 that (there is some probability that) *e* will occur at t2. [where t1 is earlier than t2]⁴⁸⁰

I shall need to make quite a few modifications to this definition in order to reach anything directly useful to the present context. First, there are some terminological problems: As holdings of beliefs are not in any normal sense “events” I will change “event” to the more neutral “occurrence.” Also I will allow that “t2” might be a(n) (open-ended) time-interval such as “next year” or “the future.”

Second, there is a problem involved in the “to some degree.” Presumably, this qualification is supposed to correspond to the second parenthesis, such that, according to Zimmermann, an agent foresees to a 60% degree that an event will happen tomorrow if, and only if, she consciously believes that there is a 60% probability that that event will happen tomorrow. However, this seems a slightly unnatural way of expressing the latter proposition. If the agent has this belief, it arguably seems more natural to say that she simply *foresees* that the event will happen than to say that she foresees it to any limited degree: After all, she believed that the event was more likely to happen than not. I shall therefore prefer to leave out the strange “to some degree” qualification.

Another unnaturalness to Zimmermann’s definition is the following: It would seem that, on the most natural understanding of foresight, if I foresee that an occurrence will occur, I simply take it that it is more likely than not that this occurrence will occur, i.e. that it is more than 50% probable that the occurrence will occur within a designated time-interval. In contrast, if I take it to be less likely that an occurrence will occur than not occur (e.g. if, on Zimmermann’s definition, I “foresaw” it to a 40% degree), it seems highly odd to say, if it occurred, that I nevertheless foresaw it: If I believed it only 40% likely that my home town football team would win the championship and it won nonetheless, I am hardly entitled to say that I foresaw this outcome. Thus the notion defined by Zimmermann is not exactly a standard notion of foresight. For reasons that shall soon emerge I shall go along with the basics of it nevertheless.

However, some crucial modifications must still be made. Earlier I went to some length in distinguishing between the doxastic attitudes of *acceptance* and *belief*.⁴⁸¹ Already this distinction has proven highly important, not least in dismissing certain counter-examples to the Psychological Impossibility Claim regarding doxastic voluntarism (see Section 8.3.2.1). In the present context, again the distinction is of decisive relevance: When an agent is to be judged blameworthy for the undesirable consequences of a blameworthy action or omission, it seems inessential that in performing the action she was actually convinced (i.e. *believed*, see. Section 2.1), that there was a positive probability (a risk obtained) that some undesirable consequences would

occur. What matters is rather, if in her practical reasoning leading her to perform the blameworthy action, she posited as a premise the proposition that such a risk obtained, in other words: if her *acceptance* of that proposition led her to act as she did.

Recurring to the canned soup manufacturer example might help clarify this point. Suppose that, after reading various popular medical articles, the canned soup manufacturer is agnostic about the prospective harmfulness of adding the cheapening ingredient to the soup, in the sense that after confronting differing opinions, he neither believes that the ingredient is harmless nor that it is dangerous. However, finding the evidence for the danger slightly stronger, he accepts that the ingredient is cancer provoking, i.e. he goes along with this proposition for his practical reasoning about how to mix the soup: Being an avid gambler, he reasons that his prospects of wealth together with the unlikelihood that he will ever suffer punishment for his actions outweigh his modest concerns for the health of his customers, and he proceeds to add the ingredient. Surely, in such a situation whether or not his acceptance that risk obtained was prompted by a corresponding belief, it is intuitively sufficient to render him blameworthy for the harm resulting from his action.

There is also an issue concerning the “non-zero probability” involved in Zimmermann’s account, that must be resolved if the resulting notion of foresight is going to be relevant to a deontological context. It might seem that there are probabilities so small that we are entitled to ignore them in our practical reasoning, in the sense that we cannot be blamed for any undesirable consequences of an action or omission, even if, in performing them, we believed or accepted that such a small risk obtained. Remember that, on reflection, we must acknowledge that there is a tiny chance that somebody will die or get hurt from almost anything we do.

Imagine, for instance, that I bow down to tie my shoe-laces. Now, imagine that I believe (correctly, even) that as a result of my bowing down, I might by an incredible misfortune move the surrounding air in a way that ultimately brings about a lethal thundercloud on the other side of the globe, killing hundreds of people in Australia.⁴⁸² Suppose now that my bowing down to tie my shoe-laces at that point was somehow blameworthy, e.g. I had an obligation of some sort not to bow down at the time and no excuse for violating this obligation. Still, does this blameworthiness and my “foresight” that I might also by bowing down kill hundreds of people in Australia make me the least blameworthy for these deaths if they actually occur. Hardly!

However, examples like this do not really point to a level of risk that I am entitled to ignore as insignificant to my actions or omissions. In the above case, a tiny risk *also* obtained that some great harm would result from my *not* bowing down to tie my shoe-laces. In fact then, no obvious *increase* in the risk of a thundercloud in Australia resulted from my bowing down. However, if the risks adhering to the performance and non-performance of an action are not balanced in this way, things look differently:

Imagine, e.g. that our canned soup manufacturer accepted that there was a very tiny increase (say 0.00001) in risk that someone would die as a result of ingesting his soup, once he added the cheapening ingredient, and that such an increase in risk actually obtained. He nevertheless went on to add the ingredient. Now, if someone

actually dies as a result of eating the soup, the soup manufacturer certainly seems to be blameworthy for this, insofar as he accepted the *increased* risk of this undesirable outcome. This example points to the fact that it is acceptances of an *increase* in risk, rather than an acceptance of risk proper that matter to the deontic evaluation of action consequences.

Finally, Zimmermann's definition arguably suffers from the defect that it does not count foresight as a *factive* attitude in the sense that one can only foresee what does in fact happen.⁴⁸³ It thus divorces the notion of foresight from the obviously factive notion of *sight* (seeing that), which seems a controversial move, even if, admittedly, there is a non-factive sense in which we may sometimes talk about foreseeing something. For example even if today is dry weather we may truly say that yesterday I foresaw that it would rain today, meaning nothing more than yesterday I somehow made that unfulfilled prophecy. However, our present interest is in the degree to which an agent *foresaw* – or could have *foreseen* what did in fact happen, e.g. the occurrence of a certain undesirable belief. Leaving aside quarrels whether the factive or the non-factive sense of foresight is the more common or standard one, certainly nothing is lost by sticking to a factive notion.

We might then take as a starting-point the following account of foresight, inspired by Zimmermann's definition:

(Foresight): An agent A foresees at *t* that the occurrence *o* will occur in the time-interval T, if, and only if, A accepts⁴⁸⁴ at *t* that there is a risk (non-zero probability) that *o* will occur within T, where no point of time falling within T is earlier than *t*, and *o* does in fact occur within T.

As remarked above, this notion of foresight is arguably rather non-standard. Still, what is really relevant to the present context is a notion of *conditional foresight*; what an agent foresees will happen *as a result of* one of her actions or omissions.⁴⁸⁵ Such a notion can now be analysed, guided by the above discussion and using my modified version of Zimmermann's analysis as a framework:

(Conditional foresight): An agent A conditionally foresees at *t* that, given her performance of a certain action or omission C, the occurrence *o* will occur in the time-interval T, if, and only if, A accepts at *t* that, if she performs C, *as a consequence of* her performance of C the risk increases that *o* will occur within T, and, as a consequence of C *o* does in fact occur within T.

Again, no point of time falling within T is earlier than *t*. The reason why conditional foresight is the most relevant notion to a deontic context, is that it cannot matter much from a deontic perspective that an agent foresees some occurrence, if that foresight is not tied to her actions or omissions at all. The kind of foresight that an agent should take into account in her conduct is clearly a foresight pertaining to the results of her actions or omissions.

In what follows, I shall simply refer to such conditional foresight as "foresight." If an agent at the time she violates an intellectual obligation, in this sense *foresaw* that, if she performed the action or omission that she had an intellectual obligation not to perform, undesirable beliefs would occur, plausibly she has already taken a huge step towards being blameworthy for the undesirable beliefs resulting from that

violation. Remember here that, if her action or omission constituted a violation of an intellectual obligation in the present sense, it must (cf. Section 12.1) constitute an exercise of doxastic control over the undesirable beliefs resulting from it. Thus, of course, if the term “control” is to remain worthy of its name, it must be the case that the violation of the intellectual obligation actually increased the likelihood of such undesirable beliefs. We have then a parallel of the case of the profit-focused soup manufacturer rather than a parallel of the tier of shoe-laces: An agent of the type in question accepts an increase in risk that actually obtains: If the risk actually manifests itself in undesirable consequences, plausibly the agent is blameworthy for them.

However, intuitively, an agent may still do something wrong, and be blameworthy for the undesirable consequences of this wrong-doing even if she did *not* at the time of wrong-doing foresee that the undesirable consequences would occur. It seems to be enough that the agent *should* have foreseen the undesirable outcome under the given circumstances.

Returning again to the canned soup manufacturer example might help explicate the abstract points above. If the canned soup manufacturer, at the time he added the fatal ingredient to the soup, neither did foresee nor *should* in any sense have foreseen that anything bad would happen as a result of his adding the ingredient to the soup, it is not even clear that he did anything he was obliged not to do by adding it (assuming that, e.g. he did not break the law, thus violating his legal obligations): Certainly it would be odd even to demand an excuse for his action of adding the fatal ingredient under these circumstances. However, if the manufacturer foresaw that it was likely that people would die as a result of his adding the ingredient, it seems clear that, by adding it, he was doing something that he was morally obliged not to do and should be blamed for unless he had an appropriate excuse for adding the ingredient despite his foresight. Also it seems clear that he is then blameworthy for the deaths that later occurred among the soup consumers, insofar as these deaths occurred as a result of his inexcusable violation of his moral obligations.

13.2. BLAMEWORTHY INADVERTENCE TO RISK

It does not seem farfetched to claim that, if the canned soup manufacturer from above did not foresee the deaths he caused but *should* have foreseen them at the time when he added the fatal ingredient, this suffices to make him blameworthy for the deaths that he should have foreseen.

This claim, however, demands some explanation, not least concerning what sense can be made of the statement that an agent, who did not foresee any undesirable consequences of his action or omission, nevertheless was *obliged* to foresee these consequences.

If an agent does not have any foresight regarding the undesirable consequences of an action or omission, that is: if she does not accept that an increase in the risk of undesirable consequences results from her performing that action or omission, two options are open: Either (1) she accepts the proposition that her action or omission does *not* give rise to an increase in risk that an undesirable consequence will occur

as a result of her actions or omissions. Or (2) she is *agnostic* regarding risk in the following sense: In her relevant practical reasoning she does not take account of any proposition at all regarding an increase in risk that an undesirable consequence will occur as a result of her actions or omissions. This is perhaps what is most commonly the case when an agent performs a belief-influencing action or omission.⁴⁸⁶

With regard to the option (1) it seems fairly clear that the agent may be blameworthy for the undesirable consequences of an action or omission, if, at the time she performed it, her acceptance that no increase in a relevant risk resulted from it was *itself* blameworthy under the given circumstances.

However, spelling out a general theory of (epistemic) blameworthiness for acceptances is no easy matter. In the present context, perhaps the “easiest” case is the one, where an acceptance of a proposition is prompted by a corresponding belief. Here it seems fair to say that the acceptance is blameworthy if, and only if, the corresponding prompting belief is epistemically blameworthy. Cf. Section 1.3, this is simply a special “honest mistake” case. However, when the agent does not also believe the proposition she accepts, things quickly get more complicated. Complex pragmatic concerns now become relevant. For example in a case of acute emergency, a doctor may need to *accept* a proposition that a certain medicine is the right one to administer to a patient, even if she would perhaps be epistemically blameworthy for *believing* this proposition under the given circumstances. In such cases we would have to judge, whether the circumstances were such that she did what she could be expected to do in the process that led her to accepting the relevant proposition, that is, if she had appropriate non-epistemic reasons to accept the proposition under the given circumstances, and if not, whether she was appropriately excused for accepting it nevertheless. I shall not try to develop a general theory of such excuses in the present context, and will have to leave this part of my analysis open-ended.

With regard to option (2), arguably we already have the tools present to account for the blameworthiness it may give rise to. Consider again the canned soup manufacturer example. Under which circumstances should we hold the manufacturer blameworthy for the cancer deaths in virtue of his omission to hold back the soup with the fatal ingredient from the market if, at the time he could have done this, he was entirely *agnostic* (in the sense introduced above) concerning an increase in risk of undesirable occurrences resulting from that omission?

I will offer the following answer to this question: The manufacturer should be held blameworthy for the undesirable consequences of his omission if, and only if, *had* he accepted the proposition that *no* increase in the risk of undesirable occurrences obtained, this acceptance would have been blameworthy.

I take this answer to be correct, simply because I do not perceive any interesting difference in blameworthiness between an agent who does not take into account the possibility of an increase of risk at all, and an agent who does so only by acting on the premise that no increase in risk obtains, but is *blameworthy* for hypostatizing this premise. In a sense the latter agent had no *right* to this premise, and should therefore hardly be held less epistemically blameworthy than the former agent, who ignored such a premise entirely. If this equivocation seems to rash, consider that for the former

agent, no matter if in her practical reasoning, she had accepted the proposition that an increase in risk obtained (foresight) or the proposition that no such increase in risk obtained, she would have been blameworthy for the consequences of that action or omission. To this author at least, it would seem wrong to acquit her, simply because she happened to accept neither.

About both kinds of agent, then, I shall say in the following that they displayed *a blameworthy inadvertence to risk*. Blameworthy inadvertence to risk is present, when an agent was obliged to foresee certain consequences of an action, but inexcusably did not foresee those consequences. As made out above an agent can be blameworthy inadvertent to risk in two cases: (1) She accepted the proposition that *no* increase in risk obtained, that undesirable consequences would follow from her actions or omissions and she is blameworthy for this acceptance. (2) She did not accept that an increase of risk of the above-mentioned sort obtained, but *neither* did she accept that an increase in risk of the above-mentioned sort did *not* obtain. In other words, she was entirely agnostic concerning such risk. However, if she had accepted that no increase in risk of the above-mentioned sort obtained, she would have been *blameworthy* for that acceptance.

13.3. RELEVANT FORESIGHT AND RELEVANT INADVERTENCE TO RISK

At this point a central question must be addressed: In order that an agent may be held blameworthy for the undesirable consequences of an action or omission due to the fact that this action or omission was performed with either foresight or blameworthy inadvertence to risk, it seems essential *which* risk was involved in the agent's foresight or blameworthy inadvertence. For example if I did something inexcusably wrong by parking my car in a disabled person's parking spot and did so with the foresight that I might thereby harm a disabled person, this does not seem relevant to my possible blameworthiness for the consequence that my illegal parking accidentally disrupted a police car chase later that day.

In the canned soup manufacturer example the central question may be framed thus: What exactly, in order to be blameworthy for the cancer deaths resulting from his adding the fatal ingredient to the soup, should the canned soup manufacturer foresee or be obliged to foresee at the time he added this ingredient? Was it, e.g. enough that he foresaw or was blameworthily inadvertent to the risk that some consumers would *get ill* from eating his soup, or should he foresee or be blameworthily inadvertent to the risk that people would actually *die*?

I shall attempt only a rough outline of an answer below. First note that, if the manufacturer only foresaw that his adding the ingredient to the soup would possibly cause a few headaches, and he was not blameworthily inadvertent to the risk of bringing about more severe forms of harm, it seems overly zealous to blame him for the deaths that actually resulted.

On the other hand, if the manufacturer foresaw that some people would be severely injured, or was blameworthily inadvertent to the risk of injuring somebody severely, it seems plausible to blame him for the deaths of the soup-consumers, even if, when

adding the ingredient, the manufacturer neither foresaw those deaths nor was blameworthy inadvertent to the risk of causing deaths. Plausibly, an agent's foresight or blameworthy inadvertence need only concern *related* undesirable consequences of *roughly the same degree of undesirability or worse*, in order that she may be blameworthy for some particular undesirable consequences of her actions or omissions. In the present case, severe bodily harm plausibly qualifies as being of roughly the same degree of undesirability as death. However, I do not believe that any general guidelines can be offered for comparing undesirable consequences in this manner. We are moving in an inherently vague territory here.

There is also a temporal version of the above problem: What if the deaths occurred after 5 years of soup consumption, but the manufacturer was entirely agnostic of risks within 5 years, and had only, e.g. a blameworthy acceptance that there was no increase in the risk of deaths or injury in the interval from 10 to 15 years after adding the fatal ingredient to the soup? Here, I will submit, the temporal perspective involved in the foresight or blameworthy inadvertence does not matter at all.

For example in the case of the canned soup manufacturer it seems inessential to his blameworthiness for the deaths he brought about, whether his foresight or blameworthy inadvertence concerned only the proposition that undesirable consequences of this magnitude would result in 25 years: It is the *severity* of the consequent harm foreseen that matters, not its exact spatio-temporal location.

At last, we can now attempt to apply the above observations to the epistemic context. To keep matters as concrete as possible, consider again the case of the educated racist. Assume as several times above that the educated racist's belief that blacks are generally inferior to whites was brought about as a consequence of her skipping a racial issues class in college. Imagine further that, by skipping this class, she violated an intellectual obligation. Assume again that her particular epistemically undesirable belief that blacks are inferior to whites occurred much later after having credulously trusted the testimony of a racist demagogue. What kind of foresight should the educated racist have had at the time of her college class in order to render her epistemically blameworthy for holding her later racist belief, over which she exercised a certain degree of property Pascalian control by her choice of skipping class when in college?

Plausibly, analogous to the canned soup manufacturer above, she need not have had any foresight or blameworthy inadvertence concerning *exactly* the risk of coming to hold an undesirable belief of the racist kind, which she ultimately formed. Arguably it is enough that she, e.g. foresaw or was blameworthy inadvertent to the risk that she would sometime in the future form an epistemically undesirable (e.g. unreasonable) belief *with a racist content* as a result of her skipping the racial issues class.

As in the case of the canned soup manufacturer above, however, I believe that an account of the conditions under which the educated racist is epistemically blameworthy in virtue of violating her intellectual obligations, must remain vague. In the present case it even seems harder than in the example of the soup manufacturer above to determine what should count as a related undesirable consequence of roughly the

same degree of undesirability or worse. Should we, e.g. declare the educated racist blameworthy for her undesirable belief, if neither her foresights nor her blameworthy inadvertences in the relevant college situation involved beliefs about racial issues, but only beliefs about, say, Scandinavian film-making (suppose the class featured a showing of a Scandinavian film)? Since arguably our intuitions about epistemic blameworthiness are not even nearly as strong as our intuitions concerning cases of moral blameworthiness, such a question is not that easy to answer.

Ultimately then, all I am willing to conclude concerning the relation between violations of intellectual obligations and ensuing cases of epistemic blameworthiness is the following vague statement:

An agent is blameworthy for holding an undesirable belief that *p* if, and only if, she holds it as a result of inexcusably violating one or more intellectual obligations and she violated these obligations on either (1) a foresight that, *as a result of* her performing the action or omission constituting a violation of an intellectual obligation, an undesirable belief of roughly the same kind⁴⁸⁷ as her particular belief that *p*, or of a kind including her particular belief that *p*, would occur; or (2) a blameworthy inadvertence to the risk that, *as a result of* her performing the action or omission constituting a violation of an intellectual obligation, an undesirable belief of roughly the same kind as her particular belief that *p*, or of a kind including her particular belief that *p*, would occur.

I take it that this claim captures at least my personal intuitions that if none of the foresights or blameworthy inadvertences involved in an agent's violation of an intellectual obligation concerned an undesirable belief of a kind related to the belief that *p*, it is hard to see how the agent can be blamed for holding the belief that *p* by appeal to her violations of intellectual obligations. In the case of the educated racist the relevant kind of foresight could, e.g. be her acceptance that, as a result of her skipping the racial issues class, an increase in risk obtained that some unreasonable belief about racial issues, or about dark-skinned people, or an unreasonable belief formed on demagogic testimony, would occur.

Below (Section 14.1), I shall attempt to wield these observations into a canonical analysis of the notion of epistemic blameworthiness also taking into account the variety of deontologically adequate modes of doxastic control worked out in Chapters 8–11 above.

13.4. LARRY ALEXANDER'S OBJECTIONS

In Sections 1.3 and 1.4 I argued that an agent's epistemic blameworthiness for holding a certain belief might sometimes account for her moral blameworthiness and even her criminal liability for her harm-doing. The roots of this liability I have now traced further back to violations of intellectual obligations under certain conditions of foresight or blameworthy inadvertence to risk.

However, one author, Larry Alexander, has presented objections to drawing such a connection. He presents the "honest mistake" case *State vs. Williams* (see Section 1.4), a case where two uneducated parents omitted to bring their dying son to a doctor,

acting on the belief that he was only suffering from a benign toothache in no need of professional medical attention. Alexander now rejects the suggestion that the supposed “culpability” of the Williamses for the death of their son might be tracked back to their *blameworthy* choices in, e.g. not attending classes on children’s health care, which, had they attended them, would have changed their cognitive dispositions, so that they would no longer have held unreasonable beliefs about their son’s health condition.

The relevance to the present context should be obvious. According to the present argument, if the Williamses, with a suitable foresight or blameworthy inadvertence to risk, violated an intellectual obligation in not attending the health class, this would render their undesirable belief about their son’s health condition epistemically blameworthy, which would again, insofar as their failure to take their son to the doctor was prompted by that belief, make them blameworthy for the death of their son. If Alexander can present compelling objections to drawing these connections, the present argument would be in serious trouble.

The objections occur in the following passage:

First, in tracing the negligent act [here: the Williamses’ letting their son die] to the prior choices that led to the crucial failure to advert to the risk, we may find a prior choice that was culpable – but we may not. Second, if we find a prior choice that was culpable, it may be quite remote in the causal chain. Third, and perhaps related to the second point, the degree of culpability displayed by that prior culpable choice may be quite disproportionate to the punishment specified for the negligent act. An act of very minor culpability, such as cutting language classes for soldiers stationed in a foreign country, might lead to a major harm, such as rape (through misunderstanding “no” and “yes”). On the other hand, many quite reckless choices will produce only minor harms. And, as I said, many prior acts will not be culpable at all.⁴⁸⁸

The first point made by Alexander above is this: Possibly the mistaken belief in an “honest mistake” case is not due to any previous blameworthy “culpable” actions or omissions “choices” at all. This observation, of course, presents no obstacle to the present position. In that case, the mistaken belief, even if epistemically undesirable, simply fails to be epistemically blameworthy and cannot give rise to blameworthiness for harm-doing at all. Since Alexander seems to opine that the Williamses should be acquitted of all criminal charges for the death of their son, he should welcome this possibility.

Alexander’s second point arguably just amounts to the observation that a violation of an intellectual obligation may be much prior in time to the belief or harm-doing resulting from it. Above I have argued that, when rendering an undesirable belief epistemically blameworthy, it does not matter essentially to its relevance how remote in time or causal proximity a violation of an intellectual obligation is. What matters, is rather that the epistemic undesirability of the undesirable belief resulted from that violation and that the agent performed the violation with the relevant kind of foresight or blameworthy inadvertence to risk as outlined above.

Alexander’s third point in the above passage, however, raises a significant concern: Suppose, in the *Morgan* case, that the undesirability of a rapist’s belief that Mrs. Morgan consented, resulted from the fact that he had skipped a class on sexual awareness in secondary school. Here Alexander’s objection may be stated thus: Even

though the *Morgan* rapist might have violated an intellectual obligation of some sort by skipping the class, and might even have foreseen that his failure to attend to this class would render him less likely to reliably form beliefs about women's consent in the future than he would have been otherwise, is it not overly harsh to track his moral blameworthiness for an offence as serious as a *rape* back to this "minor" incident? Further, if we are punishing this rapist severely on this background without punishing in a similar way the many who have committed comparable or worse violations of epistemic obligations, are we not subjecting him to an unjust "punishment lottery"⁴⁸⁹?

I shall address these questions in the order they were presented. First, if we still hold the soldier's cutting language class in Alexander's original example "an act of very minor culpability," *even* if this soldier foresaw or was blameworthily inadvertent to the risk that, as a result of his cutting class, he would unreliably form beliefs about the sexual consent of local women, perhaps our evaluative scale is simply ill adjusted. Given the significant risk of severe harm resulting from the soldier's action (or omission, depending upon how we describe the case), surely his transgression is not to be taken lightly. In general, the considerations of the wider significance of the notion of epistemic blameworthiness offered here, may well be taken as a reason to reconsider our attitudes toward violations of intellectual obligations, insofar as our default position, c.f. Alexander, is to regard such violations as acts of "very minor culpability."

Second, regarding Alexander's "lottery" point: It is quite true that violations of intellectual obligations, even when performed with the foresight or blameworthy inadvertence to risk sufficient to render resulting undesirable beliefs epistemically blameworthy, may nevertheless, due to sheer luck, never result in any serious harm-doing. In this sense violators of intellectual obligations enter a lottery: Their violations may or may not cost them later counts of moral blameworthiness or criminal liability. However, this "lottery" is no different from, e.g. the "punishment lottery" an intoxicated driver enters by rolling out of her carport: If she does not get stopped by the police no punishment arises; if she get stopped and her blood is tested, she may face a fine and loose her license; if she is unlucky enough that a child crosses the street in front of her car, she may face a harsh prison sentence and a life of terrible self-reproach.

I shall not here comment on the general justness of this lottery, as this would involve me in a complicated digression concerning the wider justification of punishment. In the present context, I will simply offer the double rejoinder to Alexander's objection, that (1) The present account of the source of moral blameworthiness in "honest mistake" cases like *Williams*, does not conflict with the principles underlying our standard legal practice insofar as "lotteries" are concerned. (2) In cases like that of the intoxicated driver mentioned above, it does not seem to detract from her blameworthiness for killing a child that she could avoided this outcome by sheer luck. Likewise in the epistemic case, if an agent's violations of her intellectual obligations ultimately result in harm-doing, surely the fact that by sheer luck it could have turned out otherwise, does not make a significant difference.

13.5. FORESIGHT, INADVERTENCE AND HARM-DOING

Having put aside Alexander's objections, I shall proceed to consider what I take to be perhaps the most problematic objection to the account of epistemic blameworthiness offered above, regarding its fitness to underwrite ascriptions of moral blameworthiness in "honest mistake" cases: The foresight or blameworthy inadvertence to risk I have demanded above in order to tie together the epistemic blameworthiness of undesirable beliefs and the violations of intellectual obligations from which their undesirability originated, concerns only the epistemic undesirability of beliefs, *not* the harm-doing that may result from acting on such beliefs. This means in effect that on the present account, an agent may be blameworthy for harm-doing due to her violations of intellectual obligations although she did not foresee, nor was required to foresee, any harm of this type as resulting from these violations.

This runs counter to a common intuition that an agent cannot be morally blameworthy for *unforeseeable* consequences of her actions, even if these actions are themselves blameworthy.⁴⁹⁰ Two options are open here: Either we allow that epistemic blameworthiness is an exception to this rule: An agent can be morally blameworthy for harm-doing resulting ultimately from her violations of epistemic obligations, even when, at the time she violated those obligations, she could not foresee that any such harm-doing would result from them. Or we acknowledge that the epistemic blameworthiness of the prompting belief must be supplemented with demands of a relevant foresight or blameworthy inadvertence to *harm-doing* at the time, when the relevant intellectual obligations were violated, if blameworthiness for harm-doing in "honest mistake" cases is to be tracked back to violations of intellectual obligations.

It is far from easy to pass a general verdict here. However, I believe that the first (perhaps more controversial) option is defensible, at least in some cases: Consider again the educated racist. Now suppose that, just after her violation of her intellectual obligation to attend the racial issues class, she is taken away to serve a life sentence in a very restrictive solitary confinement, where her credulous disposition towards racist testimony can lead to no harm at all. Suppose further that, at the time she skipped the racial issues class, she was aware of her future prospects and had been presented with overwhelming evidence to the conclusion that she was going to spend the rest of her life in this prison. At the time she violated her intellectual obligation, she could not then possibly foresee any other development, than that her credulous disposition would never pose a threat of harm-doing to anyone. However, totally unforeseeable at the time, a radical shift in the political climate of her country occurs, and the racist is pardoned and set free by a new lenient government. Once free, she quickly amasses a host of racist beliefs and someday commits a hate crime against a person with dark-coloured skin acting on the belief that this person deserved to be beaten up due to her "inferior" skin colour.

What are we to say here? Is the racist to go free of moral blame, or is she morally blameworthy due to the origin of her mistaken belief in her past violation of an intellectual obligation, even if, at the time she violated it, she could not have foreseen that morally significant harm-doing would occur as a result of her violation? My personal

intuitions lean slightly to the latter solution. I acknowledge, however, that this point is highly controversial.⁴⁹¹

What must be remarked still is that, even if foresight of harm or blameworthy inadvertence to harm is judged to be what ultimately drives home the connection between blameworthy “benighting”⁴⁹² actions or omissions and harm-doing based on “honest mistakes,” the notion of epistemic blameworthiness is still crucial to this connection, since the source of harm in such cases is the epistemically undesirable belief prompting the harmful action. However, if we judge on the basis of cases like the above that, even if the prompting belief is *epistemically* blameworthy on the notion worked out above, this still does not suffice to render the agent *morally* blameworthy for her harm-doing, it would seem that we will be forced to distinguish between *two* types of epistemic blameworthiness, significantly different in their explanatory powers.

The first kind, which is essentially the kind treated above, we might then call “pure” epistemic blameworthiness. An agent can be “purely” epistemically blameworthy for holding an epistemically undesirable belief, insofar as the undesirability of this belief results from past violations of intellectual obligations, performed with a foresight or blameworthy inadvertence risk narrowly concerning *epistemically undesirable* consequences.

In contrast to this notion we might then introduce the notion of *epistemic-cum-moral* blameworthiness for holding an epistemically undesirable belief. This type of blameworthiness would arise, when the epistemic undesirability of a belief held by an agent results from her violation of intellectual obligations with a foresight or blameworthy inadvertence to risk pertaining to *harm* resulting from these violations.

If we acquit the educated racist of moral blameworthiness in the last-mentioned “prison” scenario, it seems clear that only the notion of epistemic-cum-moral blameworthiness for holding an undesirable belief, can do the job I devised for the notion of epistemic blameworthiness in Sections 1.3 and 1.4, namely the job of accounting for ascriptions of moral blame and criminal liability in “honest mistake” cases. Even on this position though, “pure” epistemic blameworthiness may play a role in justifying our reactive attitudes towards other agents holding undesirable beliefs under certain circumstances (Section 1.2), insofar as “pure” epistemic values are important to us.

Also, it should be held in mind that epistemic-cum-moral blameworthiness is in a sense much more epistemic than moral: Unless the belief under deontic evaluation is *epistemically* undesirable in the first place, moral concerns do not even enter the picture.

EPISTEMIC BLAMEWORTHINESS ANALYSED

Abstract. In this chapter I first (Section 14.1) present my preferred analysis of the notion of epistemic blameworthiness in full detail. In Section 14.2 I then provide some initial comments upon the issue of appropriate excuses from epistemic blameworthiness, especially concerning what it takes for an agent to violate an intellectual obligation incumbent upon her and still be appropriately excused for this violation. I submit that no general answer to this question can be provided. Still, in Section 14.3 I proceed to discuss the merits of two types of excuses from epistemic blameworthiness discussed in the literature: Cultural isolation (Section 14.3.1) and cognitive deficiency (Section 14.3.2). I pull on the resources of my detailed analysis of Section 14.1 in order to demonstrate that these types of excuses are only appropriate within certain constraints. In Section 14.4, finally I confront an objection presented by James Montmarquet that an analysis of the notion of epistemic blameworthiness like the one presented in Section 14.1 is inadequate, since it is victim to a vicious circularity. I argue that Montmarquet's worry is greatly exaggerated: Even though some kind of circularity may indeed arise in exotic scenarios, my analysis is hardly inadequate for that reason.

14.1. A DETAILED ANALYSIS OF THE NOTION OF
EPISTEMIC BLAMEWORTHINESS

Given the footwork undertaken in the paragraphs 5–13, I am finally in a position to present my final analysis of the notion of epistemic blameworthiness:

$\text{Blame}_{\text{epist, final}}$: An agent is **epistemically blameworthy** for holding the belief that *p* if, and only if,

1. (Belief). She **believes that** *p*.
2. (Epistemic Undesirability). Her holding of the belief that *p* is **epistemically undesirable**.
3. (Doxastic Control). She enjoyed some mode of **doxastic control** *M* belonging to the set \mathcal{M}^* over her holding of the belief that *p* and it is either the case that
 - a. The epistemic undesirability of her belief that *p* was caused by an exercise of *M*. Or
 - b. The epistemic undesirability of her belief that *p* was caused by a failure to exercise *M*.
4. (Intellectual Obligation). By performing the action or omission *P* constituting either the exercise of *M* mentioned in 2a or the failure to exercise *M* mentioned in 2b, she violated an **intellectual obligation** incumbent upon her, and did this with either
 - a. A relevant **foresight**.
 - b. A relevant **blameworthy inadvertence to risk**.
5. (Lack of appropriate excuse). She had no **appropriate excuse** for performing *P* under the above circumstances.

Let me briefly recapitulate my central points concerning the notions involved in the above conditions:

Belief: I say that an agent *believes* the proposition p if, and only if, she is in a mental state of conviction to a high degree that p . In Section 2.1 I contrasted the doxastic attitude of belief with the attitude of *acceptance that* p , which I took to be a policy of positing p as a premise in one's practical and theoretical reasoning.

Epistemic undesirability: I say that an agent's holding of a particular belief under particular circumstances is *epistemically undesirable* if that belief is either not formed by a reliable belief-forming mechanism, not based on adequate evidence or not reasonably held, in the senses of these terms laid down in Chapter 5. I do not claim that no further dimensions of epistemic undesirability (epistemic indesiderata) are relevant to the notion of epistemic blameworthiness, but have assumed that the above three epistemic indesiderata are the ones relevant to "standard" cases of epistemic deontic evaluation, such as those examined in this study.

Doxastic control: As worked out at length in Chapters 8–11, out of the sixteen possible modes of doxastic control acknowledged in the taxonomy of Section 7.3, ten are, for different reasons, entirely irrelevant to actual ascriptions of epistemic blameworthiness. These entirely unfit modes are: The four modes of direct content-directed control (the ones dear to doxastic voluntarism), the two modes of abortive direct property-directed control (following the demise of doxastic voluntarism), and the modes of negative genetic and positive abortive indirect content-directed control (victims of the Frankfurt-style argument from Section 10.4), together with the modes of negative and positive abortive indirect property-directed control, which had no independent relevance (see Section 11.1).

This means that the variable "M" in the second clause of $\text{Blame}_{\text{epist, final}}$ needs to range over the following set \mathcal{M}^* of six deontologically adequate modes of doxastic control: {positive genetic direct property-directed control, negative genetic direct property-directed control, positive genetic indirect content-directed control, negative abortive indirect content-directed control, positive genetic indirect property-directed control, negative genetic indirect property-directed control}. As we saw in Section 10.4, the two modes of indirect content control may at most be invoked in bizarre cases involving hypnosis or other more or less exotic means of self-imposed "brain-washing." On the other hand, the direct and indirect modes of doxastic property control appeared to be relevant to many evaluations of epistemic blameworthiness. To sum up my conclusions from Chapter 9, an agent may typically exercise a mode of *direct* doxastic property-control by applying more or less "care" (cf. Montmarquet, see Section 9.2) to her perceptual processes. In loose terms, she may express such care in the attention she gives to keeping her senses to the best of their abilities and focusing her mind on the evidence presented to her. On the other hand, as made out in Chapter 11, an agent may exercise a mode of *indirect* doxastic property control via her ubiquitous influence on her cognitive dispositions, reasoning processes, evidence-gathering and cognitive situatedness.

Intellectual obligations: I conceive of an intellectual obligation as an obligation to perform or omit certain belief-influencing actions or omissions. It was argued in Chapter 12 that such obligations clearly obtain within proper institutional settings, but may plausibly also obtain outside of such settings, rooted in the general value of the agent's pursuit of the truth-goal or the value of avoiding harm. However, I followed William Alston in recognizing that it is not possible to offer a detailed list of intellectual obligations applying to all agents, since the service of the truth-goal expected of an agent varies greatly across cultural contexts.

Relevant foresight: In the present context, a "relevant foresight" is understood thus: In performing the action or omission P, an agent acted with a *foresight relevant to her epistemically undesirable belief that p* if, and only if, in performing P either she accepted the proposition that as a result of her performance of P, the risk increased that she would come to hold an epistemically undesirable belief B, B being of roughly the same kind as the belief that p. Or she accepted the proposition that as a result of her performance of P, the risk increased that she would come to hold an epistemically undesirable a belief of a certain kind T, T being a belief-kind to which her belief that p belongs (see Section 13.3).

Relevant blameworthy inadvertence to risk: In the present context, a "relevant blameworthy inadvertence to risk" is understood thus: In performing the action or omission P the agent acted with a *blameworthy inadvertence to risk relevant to her epistemically undesirable belief that p* if, and only if, in performing P, she accepted the proposition that *no* increase in risk obtained as a result of her performance of P that she would come to hold an epistemically undesirable belief B, B being of roughly the same kind as the belief that p. And she was blameworthy for accepting this proposition. Or In performing P, she accepted the proposition that *no* increase in risk obtained as a result of her performance of P that she would come to hold an epistemically undesirable belief of a certain kind T, T being a belief-kind to which the belief that p belongs. And she was blameworthy for accepting this proposition. Or it is the case that, if, in performing P, she had accepted any of the above-mentioned propositions, she would have been blameworthy for so accepting.

Lack of appropriate excuse: I shall proceed to discuss the condition (5) of $\text{Blame}_{\text{epist, final}}$ in Section 14.2 below.

It should be clear by now that the conditions 2–5 of $\text{Blame}_{\text{epist, final}}$ are really a further analysis of the second clause (lack of appropriate excusability) in the initial formulation of $\text{Blame}_{\text{epist}}$ from Section 4.1, lack of appropriately excusability being simply an adequate label for the set of further conditions an agent must satisfy in addition to holding an undesirably belief in order to be epistemically blameworthy for holding it.

The above analysis of the notion of epistemic blameworthiness is admittedly far from neat and tidy. Also it creates quite a barrier to overcome when justifying actual ascriptions of epistemic blameworthiness: It is far from easy to determine whether all of the complicated and vague conditions have been satisfied in a given case. However,

I have evaluated and rejected, I take it, all serious attempts in the literature to provide a simpler analysis of the notion. Thus, in lack of better options, I take it that we will have to live with $\text{Blame}_{\text{epist, final}}$, until a better analysis arrives.

Also, $\text{Blame}_{\text{epist, final}}$ is not without important theoretical and practical consequences: If it is even partially adequate, it supports the thesis that a notion of epistemic blameworthiness cannot be easily reduced to non-deontic notions, but rather depends on other such notions as worked out above: not least the notion of an intellectual obligation.

Further, its very intricacy is important to our actual social practices, insofar as we want these practices to be fair and just: Given the difficulty of establishing beyond reasonable doubt the epistemic blameworthiness of an agent in realistic cases, we should first of all be very cautious in endorsing reactive attitudes like indignation and resentment towards an agent holding an epistemically undesirable belief, insofar as (cf. Section 1.2) we want to justify such reactive attitudes by appeal to the epistemic blameworthiness of that agent. Arguably, such reactive attitudes can be very crippling to the recipient, not least with regard to her social life.

Also, in cases of harm-doing based on “honest mistakes” we should be very cautious in ascribing moral blameworthiness to the harm-doer for the harm done, not to mention criminal liability based on such ascriptions, since, as worked out in Sections 1.3–1.4, such ascriptions must ultimately rely on ascriptions of epistemic blameworthiness to the harm-doer for holding the mistaken belief that prompted the harm-doing. This is not to say that harm-doers like, e.g. the actual *Morgan* rapists should go free of punishment for their horrible harm-doing. In such cases, the claim that an “honest mistake” occurred, is common, but can rarely command much credibility. However, in cases like *Williams* (see Sections 1.4 and 13.4), where it was clearly an honest mistake that led to the death of an infant, pity and education rather than punishment seems the appropriate response, given that the unhappy parents could hardly have been epistemically blameworthy for their false and unreliably formed belief that their son was not ill enough to merit professional medical attention.

The general recommendation of caution in ascribing criminal liability and moral blameworthiness in such cases stands, even if the criteria of epistemic blameworthiness offered in $\text{Blame}_{\text{epist, final}}$ should ultimately be nothing more than “platitudes which contextually define the terms of a final vocabulary currently in use,” to use the scathing words of Richard Rorty,⁴⁹³ and thus have no relevance outside, say, Western Europe and the U.S. in the twentyfirst century. After all human beings are resented, blamed and convicted to harsh punishments based on their beliefs in what may plausibly be referred to as “our current practice.” No dose of post-modern irony will make the importance of substantial recommendations for that practice disappear, and nothing short of a thorough analysis of our basic deontic concepts seems to be able to provide a fair ground for such recommendations.

However, I still need to confront a line of questions concerning the above analysis, not least I need to answer the worry that it involves me in a vicious regress. To confront this worry in a satisfactory manner, I shall first try to answer in outline the question what can count as an appropriate excuse on the condition (5) of $\text{Blame}_{\text{epist, final}}$, i.e.

what may excuse an agent from epistemic blame, even if she violated an intellectual obligation with a relevant kind of foresight or blameworthy inadvertence to risk.

14.2. WHAT COUNTS AS AN ADEQUATE EXCUSE FOR VIOLATING AN INTELLECTUAL OBLIGATION?

If the condition (5) of my above final analysis of the notion of epistemic blameworthiness is not to be entirely vacuous, I will need to answer at least in outline the question posed in the headline of the present paragraph.

I shall approach a more general answer by considering again our stock example of the educated racist: Suppose that the educated racist violated an intellectual obligation by skipping her racial issues class in college. Suppose further that in doing so, she accepted that, as a result of her not being present at this class, an increase in risk obtained that she would in the future form beliefs about racial issues on inadequate evidence. Now clearly, if she later forms an inadequately based racist belief on the testimony of a demagogue, and the inadequate basing of this particular belief resulted from her once violation of an intellectual obligation by skipping the class (insofar as, had she been present at that class, at least her belief with the said content would have been adequately based had she nevertheless formed it), $\text{Blame}_{\text{epist, final}}$ yields that she is epistemically blameworthy for holding that belief, unless she was appropriately excused for her once violation of her relevant intellectual obligation.

The question now becomes: What, given her foresight or blameworthy inadvertence to risk, could appropriately excuse her from skipping class? An initial answer must be: She had an appropriate excuse, insofar as her service to a *more important* goal than avoiding an increase in the said risk prevented her from fulfilling her intellectual obligation. However, what we should allow to count as a more important goal here is not easily pointed out in any detail: If she were prevented from attending the class because she pursued a morally desirably goal such as saving another person from severe harm or injury, this plausibly constitutes an appropriate excuse. If, on the other hand, she spent her time comfortably in bed when she should have attended her class, she does not seem to be appropriately excused for her absence. In between these extremes lies a very fuzzy territory, where the agent's promises and service to other valuable goals may inform our decision.

In the case of the *Morgan* rapist Smith, an obvious shot on declaring him epistemically blameworthy for his belief that Mrs. Morgan consented to have sex with him, seems to be the claim that he violated his intellectual obligation to exercise negative genetic direct property-directed doxastic control on his belief by not focusing his mind more clearly on Mr. Morgan's malicious testimony. Had he done this, we might suppose, he would not have formed an epistemically undesirable belief about Mrs. Morgan's consent in the situation. Further it does not seem far-fetched to suppose that at the critical time he was blameworthily inadvertent to the risk of forming an unreasonable belief about Mrs. Morgan's consent. The elements are then in place for declaring epistemically blameworthy his belief that Mrs. Morgan consented.

Now, what if Smith objects that he was appropriately excused for violating his intellectual obligation to focus his mind on Mr. Morgan's testimony because he was severely drunk and thus unable to fulfil it? Clearly, intoxication can only excuse here, if Smith's drunkenness really made it impossible for him to honour his intellectual obligations and was not something that he could be blamed for. Again we are quickly moving into the general territory of blameworthiness for actions.

However, appropriate excuses may also narrowly concern epistemic goals. If, for instance, the educated racist chose to skip the racial issues class because she wanted to hear an indispensable guest lecture in her field of specialized study, should we then hold her appropriately excused for ignoring the risks pertaining to her skipping the racial issues class? Again no definite answer to such a question is ready at hand. We would here have to draw on wider concerns in weighing, e.g. the agent's need for self-fulfilment against her service of her community's wider goals (e.g. of diminishing racism amongst its member). No calculus easily applies to such matters.

Rather, everything seems to depend on our concrete value judgments in concrete cases aided only by rather abstract principles like the one invoked above that only the service of a more important goal may appropriately excuse from the violation of an intellectual obligation.

However, even this principle reveals a substantial feature of ascriptions of epistemic blameworthiness: Insofar as the legitimacy of an ascription of epistemic blameworthiness comes down to the condition (5) of $\text{Blame}_{\text{epist, final}}$ such ascriptions may ultimately hinge on *moral* values, insofar as we must judge whether the agent served moral goals more valuable than the truth-goal neglected by her violation of her intellectual obligations. The complicated nexus between moral and epistemic normativity here comes full circle: moral blameworthiness may sometimes hinge on epistemic blameworthiness, and epistemic blameworthiness may sometimes hinge on moral values. However, as should be clear from this study, epistemic and moral normativity do not therefore reduce to one another either way.

I shall now proceed to discuss two types of excuses from epistemic blameworthiness suggested by William Alston. As shall emerge, the viability of these excuses is not without wider social significance.

14.3. PARTICULAR TYPES OF EXCUSES FROM EPISTEMIC BLAMEWORTHINESS

14.3.1. Cultural Isolation

William Alston has pointed to "cultural isolation" as a general type of appropriate excuse from epistemic blameworthiness. He presents the following case:

S has lived all his life in an isolated primitive community where everyone unhesitatingly accepts the tradition of the tribe as authoritative. These have to do with alleged events distant in time and space, about which S and his fellows have no chance to gather independent evidence. S has never encountered anyone who questions the traditions, and these traditions play a key role in the communal life of the tribe. Under these conditions it seems clear to me that S is in no way to blame for forming beliefs on the basis of the traditions. He has not failed to do anything he could reasonably be expected to do. His beliefs about, for

example, the origins of the tribes stem from what, so far as he can see, are the best grounds one could have for such beliefs. And yet, let us suppose, the traditions have not been formed in such a way as to be a reliable indication of their own truth. S is deontologically justified [i.e. epistemically blameless], but he is not believing in a truth-conducive way.⁴⁹⁴

Plausibly, Alston has chosen the “primitive” community as his standard example, because it supposedly offers us a chance to establish from a more advanced cognitive position the failure of the primitive tradition to conduce truth. However, to look for an esoteric tribe with a tradition for unquestioned internal history writing, we would arguably have to search no further than certain branches of modern institutionalised science. As famously pointed out by Thomas Kuhn⁴⁹⁵ and others, especially the community of physicists has sometimes maintained a tradition of transmitting its own history to neophytes in a way that, cf. Alston, “plays a key role in the communal life of the tribe,” but is not traditionally questioned within it.

Thus, a typical physics student or even a practising physicist, might well have found herself in a cognitive situation very similar to that of Alston’s primitive islander: Basing her beliefs about the history of her community entirely on the testimony of elder tribe-members, and never having had the chance to gather independent evidence about it. Alston’s claim can then be generalized: An agent in the above cognitive situation, i.e. “isolated” within an authoritarian culture in which she is raised, can never be epistemically blameworthy for believing the received unquestioned views of her tribe, even if these beliefs are epistemically undesirable (not reliably formed).

Matthias Steup has sharply criticised this claim, referring to Alston’s above passage:

Alston argues as follows. There is nothing the agent could have done which would have exposed him to any counter-evidence. Hence there is nothing he could reasonably be expected to have done in order to have a different belief. But that, it seems to me, is wrong. In order to show why, I shall first consider how we think about our ancestor’s belief in witchcraft and then apply the result to the agent in the insulated community. When we, in retrospect, judge about our ancestor’s belief in witchcraft, we certainly do blame them, epistemically, for having believed such a foolish thing as that. The evidence for this belief was worse than poor. There actually was no reason at all to believe in witchcraft. And that is why our ancestors weren’t justified in believing in witchcraft, and that is why the agent in the insulated community isn’t justified in believing in witchcraft. The fact of the matter, then, is that we are dealing with belief that is held contrary to evidence.

What conditions would epistemically excuse an agent for believing contrary to evidence? Cultural isolation, or a lack of opportunities to get second, critical opinions, just isn’t good enough for an epistemic excuse. No matter how grim the circumstances are, if an agent holds a belief contrary to evidence, it is within his power, given he is a *rational* agent, to *reflect* upon his belief and thereby to find out that he had better withhold it, or even assent to its negation. Being a rational agent, I would say, involves the capacity to find out, with respect to any belief, whether or not it is being held on good grounds. Hence I will conclude that there is just one condition that epistemically excuses believing contrary to evidence, namely the condition of being incapacitated as a rational agent.⁴⁹⁶

In my view, Steup commits several mistakes in the above passage. First, he fails to acknowledge the importance of testimony as a source of evidence: A sixteenth century Western European peasant girl raised in a culture where the potency of witchcraft is taken for granted would, before coming of age, most likely have encountered dozens of reports testifying to the potency of witchcraft on specific occasions. Thus, even if this agent has not witnessed a single instance of potent witchcraft herself, possibly

many of the people around her, whom she has every good reason to consider reliable on such “colloquial” matters, will have testified to her that they have witnessed potent witchcraft. In this sense, her epistemic situation concerning witchcraft matters will be comparable to that of an average modern college student regarding “black holes” or other exotic entities posited by our present astronomy: She has not actually seen them, nor can she present any good reasons to hold any beliefs about them apart from an appeal to the testimony generally regarded as authoritative within her culture.

Here, however, Steup would perhaps reply that, even if the actual cognitive situation of the peasant girl and the college student are very similar, their *potential* cognitive situation is very different: The college student could take up studies in astronomy or read popular journal articles which would eventually bring her into a position where she has excellent reasons to believe that black holes exists, whereas the peasant girl, if inquiring into the matter of witchcraft or *just by reflecting on the matter* (!), insofar as she is rational, would come to realize that she has *no* good reason to believe that witchcraft is potent at all.

However, this line of reasoning overlooks that standards of appreciating evidence are also historically and culturally contingent.⁴⁹⁷ When it comes to such subjects matters as witchcraft, in fact even obviously so: It is well-known, that in cultures where curses cast by “witches” are thought to be potent, an agent’s mere belief that she has been cursed by a supposed witch is often enough to elicit a strong physiological reaction. Within the framework of modern psychology, such a reaction is likely to be taken as a symptom of hysteria rather than as significant evidence in favour of the potency of witchcraft. However, before the advent of modern Western psychology, what could be better evidence for the potency of witchcraft than the observation that a “cursed” person often falls ill, goes insane, or even dies?

There is also a semantic problem involved here: Much of what was classified as “witchcraft” in the sixteenth century, we would probably today classify as, e.g. herbal medicine, a lot of which we generally take to be quite efficient (although perhaps not exactly due to the outlandish ingredients sometimes added to the witch-brews of yore). Further, even with the most advanced knowledge of internal medicine available in her days, it is far from clear that even the most rational sixteenth century rational agent could, “on reflection,” tell the causal difference between cases of healing or illness brought about by the ingestion of herbal potions, and cases of healing or illness brought about by, e.g. curses, so as not to inductively infer from the potency of the former to the potency of the latter. After all, this was also an age, where mere *readings* from the bible supposedly held magical powers over mischievous hob-goblins and other magical creatures.

In conclusion, it is far from clear that our ancestors’ beliefs about witchcraft were held “contrary to evidence” or were even “foolish”: They were plausibly most often backed up by massive testimonial evidence and perhaps in many cases even carefully underpinned by observation. At least a typical agent would not have encountered any evidence suggesting that witchcraft was in fact impotent. Steup’s example simply appears to be very badly chosen for his purposes: Typical sixteenth century agents had no obvious reason to withhold their belief that witchcraft is potent or assent to

its negation. To appreciate his point charitably we will therefore have to alter his example quite a bit.

Let us therefore examine the case of a *modern* Western European believer in the potency of witchcraft, born and raised in a culture permeated by the authorities of modern natural science and secular psychology. In such a society it is not uncommon that an agent comes to believe that witchcraft is actually potent, perhaps on testimony encountered somewhere in the vast stream of popular television shows endlessly promulgating such beliefs. Assumedly, to Steup such a belief would qualify as a prime case of foolishness, and this even if the agent is insulated within an illiterate suburban culture without access to any evidence pointing to the conclusion that witchcraft is impotent.

However, even in this case, to claim with Steup that this illiterate agent may, *by sheer reflection*, determine that she has no good reasons to believe that witchcraft is potent, is simply absurd. As argued at length in this study, an agent's ability to influence her doxastic attitudes is seriously constrained by her cognitive dispositions: Given her credulity concerning televised testimony on magical subjects, by reflecting on her available evidence the illiterate agent would most plausibly only come to cherish her witchcraft-beliefs to an even higher degree.

Thus, if we are to get any force out of Steup's objections, on the bottom-line the crucial feature of the modern witchcraft-believer must be her *potential* cognitive situation, which, unless she is severely cognitively impaired, is certainly very different from that of the typical sixteenth century witchcraft believer: The modern witchcraft-believer supposedly could, *by suitable inquiry*, amass evidence that witchcraft is impotent, but failed to do so. If her "cultural isolation" excuse is to be deprived of power, it must be argued that in fact her belief that witchcraft is potent is not only epistemically undesirable (unreliably formed, based on terribly bad testimonial evidence), but also epistemically blameworthy, insofar as, among other things, she had an intellectual obligation to *transcend* her cultural cognitive confinement in her inquiries concerning witchcraft matters.

The interesting point of debate here is then really the following: Insofar as an agent is "in a poor position to get the truth"⁴⁹⁸ concerning certain subject-matters due to the cognitive isolation of her culture or sub-culture, to what extent does she stand under an *intellectual obligation* to transcend the cognitive confinements of her culture in her inquiries concerning such subject matters?

The correct answer to this question depends crucially on the further circumstances of the concrete case. To take the most extreme case first: One of the salient features of Alston's islander could be that he simply could not possibly have had any evidence *whatsoever* of the existence of other societies with a more advanced cognitive approach to the past of his community. Under these circumstances it is impossible to see, how on $\text{Blame}_{\text{epist, final}}$ the islander could ever be blameworthy for adhering to the received views about his island's past, since it is unimaginable how he could ever come to foresee or be blameworthyly inadvertent to the risk that, by not transgressing the cognitive boundaries of his community, he would come to hold undesirable beliefs about the past of his island:

Unless the islander has unusual imaginative powers, there is no way in which a proposition concerning such a risk could enter his practical reasoning, and further, should he even come to accept that no such risk obtains, it is hard to see how we could hold him blameworthy for this acceptance: He simply could have had no reason of any kind to accept otherwise. To claim the opposite would be like blaming a modern Western astronomer for not taking into account evidence about the birth of the Universe presented in messages transmitted by highly advanced aliens from the Andromeda Galaxy, given that such messages (unbeknownst to us) presently exist in the actual world.

The other extreme of the spectrum would be the modern illiterate suburban witchcraft believer from above. Here, even though her witchcraft belief was induced by a mechanism peculiar to the credulous sedentary lifestyle of her suburban community, it is not therefore clear that there is nothing she could and should have done in order to transgress the narrow cognitive borders of her illiterate background. This is especially so if she has been offered a general education. Although I have pointed out in Section 12.3 that a child's intellectual obligations to attend to an education are not naturally interpreted as *institutional* intellectual obligations, this does not mean that such obligations do not obtain: A general system of education, I will submit, functions as a wider community's primary defence mechanism against the ills of sub-cultural cognitive idiosyncrasies. For that reason it is of great social value that all individuals attend to it (the importance of this value is also recognized in many contemporary legislations, imposing penalties on parents for their failure to send their children to school).

Aristotle famously observed about the legal practices of his time:

Indeed we punish a man for his very ignorance if he is thought responsible for that ignorance.⁴⁹⁹

In a certain sense the same is true of ascriptions of epistemic blameworthiness on $\text{Blame}_{\text{epist, final}}$: If an agent is blameworthy for remaining in the falsity-conducive cognitive community in which she is born, she cannot offer her "cultural isolation" as an appropriate epistemic excuse. Unless cognitively impaired, the suburban witchcraft believer plausibly would not have held an epistemically undesirable belief in the given scenario, if she had taken full advantage of the mind-broadening opportunities offered to her in her obligatory education. Further, it is not farfetched to assume that when she did not take full advantage of these opportunities she often acted with a blameworthy inadvertence to risk: If, in her practical reasoning leading her to be inattentive in school, she included the premise that no increase in risk of impairing her cognitive abilities resulted from her laziness, this would probably have been a blameworthy acceptance, given the massive explicit information to the opposite conclusion she would most likely have encountered from her teachers.

Some might well opine that I am overly paternalistic in declaring this woman epistemically blameworthy for her witchcraft belief in this case. Here I can only respond that of course many excuses for violating her intellectual obligations may still be appropriate: She could be facing heavy resistance to the school system from her parents and peers, she may have had a very hard childhood making her unable

to concentrate on her classes or she might simply be severely cognitively deficient (more about this type of excuse below). However, it is to cases like the sub-urban witchcraft believer we should look for counter-examples to Alston's suggestion that "cultural isolation" in itself excuses from epistemic blame.

In the case of the suburban witchcraft believer, her supposed cultural isolation excuse deteriorated, as she was blameworthy for her cultural isolation. Plausibly such a situation can arise when, as above, the agent's intellectual obligation to transgress her cultural cognitive confinement is imposed on her by a wider community to which she also belongs. However, it is less obvious that this situation can arise, when the relevant educational opportunity is offered by an "intruding" alien culture, or when it is entirely up to the agent herself to transgress the cognitive boundaries of her culture.

An instance of the first type of case could arise if Alston's secluded island community was approached by a foreign culture offering truth-conducive testimony about the past of the island community as well as educational opportunities designed to make the islanders more truth-conducive in their belief-forming practices. In such a case, even if now an islander confronts truth-conducive testimony contradicting the beliefs shared by members of his tribe, does he therefore become epistemically blameworthy for holding them?

To answer this question, observe first that, given his cognitive dispositions, the islander would most naturally hold the elders of his tribe more credible on matters of the island's past than any intruding stranger. Thus, whatever blame may attach to the islander for still holding an epistemically undesirable belief about these matters, it must root in an intellectual obligation to change his cognitive dispositions so as to be open to the testimony of the strangers, a process best undertaken if he is, e.g. submitted to a thorough-going "re-education" by intruding missionaries.

As pointed out in Section 12.3, it is no easy matter to determine when an agent stands under a non-institutional intellectual obligation to perform some belief-influencing action. However, even if it could be argued in the style of W.K. Clifford (see Section 12.3) that the islander is intellectually obliged to pursue *any* inquiry that will improve the truth-conducivity of his belief-influencing dispositions, it must also be acknowledged that allowing himself to be "re-educated" would mean a major sacrifice to him. Remember that in Alston's example the traditionally shared beliefs played "a key role in the communal life of the tribe."⁵⁰⁰

Thus, pursuing an education or a process of inquiry that would lead him to question the tradition of his tribe, would supposedly alienate the islander severely from his fellow tribe-members and leave him in a kind of cognitive exile. Determining whether he is actually intellectually obliged to break with his cognitive past, and if so, whether the sacrifices involved therein are great enough appropriately to excuse him for violating such an obligation, is a question to which I will not try to provide a reasoned answer. Suffice it to remark that, at least to me, in such a case "cultural isolation" still seems a fair excuse from epistemic blameworthiness.

In the second type of scenario out-lined above, I will provide a more confident answer. Here it was assumed that, if the islander were to question the received views, he would have to collect the contrary evidence entirely by himself, unaided by, e.g.

intruding members of an alien cognitive culture. As Alston has submitted, perhaps there are transcultural epistemic standards such as, e.g. consistency⁵⁰¹ Thus if the tribe's traditional views are in fact inconsistent, perhaps a tribe-member can be held blameworthy for still believing them, given, among other things, that he would have recognized this inconsistency if only he had fulfilled certain intellectual obligations. However, if the traditional views are not inconsistent, it is hard to see how he could have violated any obligation in not coming to question them.

To add some force to this conclusion, observe that an unaided tribesman critical of the received views of his insular community would be much like the solitary genius from the romantic myths of popular science, single-handedly questioning and revising our culture's most ingrained views about, e.g. the nature of the physical universe. Perhaps anyone who is not a prodigy in this sense (but has the rare raw potential to be one) fails to perform one or more maximally intellectually virtuous actions. However, as remarked in Section 12.3, her actions are not exactly intellectually vicious either, and it is the presence of intellectual vice rather than the absence of intellectual virtue that provides the basis of epistemic blameworthiness.

14.3.2. Cognitive Deficiency

Apart from "cultural isolation," William Alston has suggested "cognitive deficiency" as a generally appropriate excuse from epistemic blame:

The other case I presented was a "cognitive deficiency" case. It concerns a college student who doesn't have what it takes to follow abstract philosophical exposition or reasoning. Having read parts of Book IV of Locke's *Essay*, he takes it that Locke's view is that everything is a matter of opinion. He is simply incapable of distinguishing between that view and Locke's view that knowledge is restricted to one's own ideas. There is nothing he could do, at least nothing that could reasonably be expected of him, given his other commitments and obligations, that would lead him to appreciate the difference. Hence he cannot be blamed for interpreting Locke as he does; he is doing the best he can. But surely this belief is outrageously ill grounded, based as it is on the student's dim-witted impressions of Locke.⁵⁰²

Alston soon generalizes from this case:

We have such a case whenever one forms a belief, on poor grounds, on something beyond one's intellectual capacity; and this is surely a common occurrence. Just consider a person who forms the belief that socialism is contrary to Christianity, for the reasons that often given for this view by the New Right, and who is intellectually incapable of figuring out how bad these reasons are.⁵⁰³

As in the case of the "cultural isolation" excuse, Mathias Steup radically dissents:

Our student does deserve epistemic credit for being the kind of student he is. That he isn't smart enough to reach a correct understanding of Locke's essay should not be held against him. It remains unclear, though, whether this constitutes an *epistemic excuse* for his interpretation of the *Essay*. The interpretation the student has reached, namely,

(L) "According to Locke, everything is a matter of opinion," is an object of doxastic attitudes. It was within his power to believe (L), to disbelieve (L), or to withhold (L), and he has taken the attitude of *believing* (L). It is an important part of Alston's story that there is nothing such that, had the student done it, would have led him to a better understanding of Locke's philosophy. But we can easily admit this point and nevertheless insist that there is something such that, first, our student could have done it, and second, had he done it, he would have *withheld* (L). I am claiming, then, that there is something our student could have done that would have led him to a doxastic attitude more reasonable than the one he took in fact.⁵⁰⁴

What, according to Steup, the deficient college student could have done to withhold his undesirable belief, is later brought out:

The question that still remains open is, “Is it within the student’s cognitive power to figure out that, for him, withholding (L) is more reasonable than believing (L)?” Nothing in the case, as construed by Alston, suggests a negative answer.⁵⁰⁵

Seemingly then, paralleling his approach to the case of “cultural isolation,” what Steup considers significant is that, *simply by reflection*, the cognitively deficient college student could have realized that he did not have good rationalizing reasons to hold the undesirable belief that according to Locke, everything is a matter of opinion, and that hence the student is epistemically blameworthy for holding this belief, insofar as, *had* he reflected appropriately on his rationalizing reasons, he would have reached the conclusion that he did not have good reasons for holding this belief, and consequently would have “withheld” it.

As in the case of the “cultural isolation” excuse above, Steup’s rejoinder to Alston cannot command much force. First, contrary to what Steup claims at the end of the last-quoted passage, Alston’s example *strongly* suggests a negative answer to Steup’s question: After all, the college student in Alston’s example simply “doesn’t have what it takes” to follow Locke’s reasoning.⁵⁰⁶ Since then, supposedly, his inadequate analysis of Locke strikes him as the correct one (since he believes it), no amount of reflection on his reasons for that analysis would probably lead him to recognize that he lacks goods reasons for it. If his cognitive deficiency is really severe, supposedly no amount of further reading of Locke could make him suspend his ill-based belief either.

Steup, as we have seen in Section 8.3.2.2, is a doxastic voluntarist. However, given the demise of this position, his claim that, in Alston’s example, after reading Locke’s *Essay* it was within the student’s power either to withhold (prevent), to “disbelieve” (suspend) or to believe the proposition L, is simply false. As argued in Section 8.2.2 above, normal agents simply do not have such powers over their beliefs. However, as I shall proceed to argue, Steup may still be right that in some cases an agent’s “cognitive deficiency” does not excuse him from epistemic blame, insofar as “there is something [the agent] could have done that would have led him to a more reasonable attitude than the one he took in fact.”⁵⁰⁷ Thus, as in the case of cultural isolation above, I shall aim to demonstrate that Alston’s invocation of “cognitive deficiency” as an all-out epistemic excuse only succeeds within certain constraints.

To do this, however, I first need to get into a clearer light Alston’s notion of “cognitive deficiency.” It seems that roughly Alston conceives of an agent’s cognitive deficiency as her disability to form an adequately based belief after having been confronted with evidence, upon which a cognitively adequate agent would not have formed an inadequately based belief.⁵⁰⁸

I shall generalize this definition, employing the terminology preferred here, and say that an agent is *Alston-deficient* with respect to a subject matter M and a certain set of evidence E in a situation S if, and only if, *when* confronted by E in S, she inevitably forms an epistemically undesirable belief about M based on E and an agent with satisfactory cognitive powers would not have formed an epistemically undesirable

belief about M based on E in S. Thus, e.g. the college student from Alston's example is supposedly Alston-deficient with regard to the subject matter of philosopher's opinions and the evidence presented by their texts in the situation where he has both read the text and had it lectured to him.

To approach the question under which circumstances an agent's Alston-deficiency may excuse her from epistemic blame, consider first the following practical example:

Suppose that Fred is a motorically deficient agent in the following sense: Unlike the average agent, it is impossible for Fred, even when employing the greatest care, to control his arm movements when standing close to a well-stacked dinner buffet. Being near to such a buffet invariably triggers a mechanism in his nerve system that makes him arm jitter. Fred is now invited to a party, and as one would expect, upon engaging with the buffet his erratic arm movements topple over a pyramid of wine glasses. Under which circumstances is Fred blameworthy for the harm he has caused?

Here it seems clear that, even if at the time he approached the buffet the accident was unavoidable, there were at least two lines of action Fred might have taken that would have prevented the accident: First, he could have previously entered a treatment, taken medication, or perhaps even tied his arm to his body so as to alter his unfortunate motoric dispositions. Second, he could have stayed away from the buffet altogether, letting someone else pick up his food for him. However, simply keeping his arm in place he could not do. Fred crucially lacked a direct control over his erratic arm movements. Making the parallel to an Alston-deficient cognizer, it is clear that such an agent cannot in fact wield any relevant direct doxastic control over her undesirable belief: In the relevant situation, she *must* form an undesirable belief. However, this does not rule out that prior to the relevant situation she could have exercised some mode of *indirect* doxastic control over that belief: In particular she could perhaps have exercised a measure of negative genetic property-directed doxastic control over it by influencing her cognitive dispositions so that she would not have formed an epistemically undesirable belief when confronted by the relevant evidence in the relevant situation.

Corresponding to the case of Fred, Alston's student could have brought his cognitive deficiency upon himself by a constant record of drug abuse, well foreseeing that his use of drugs would lead to appalling cognitive dispositions or he might have skipped certain preparatory courses, had he only followed them, he would not have been such a miserable interpreter of Locke's *Essay*.

Alston's student could also have exercised a measure of indirect property-directed control over his Locke belief through actions influencing his cognitive situatedness (see Section 11.3), corresponding to Fred's control over his position in the dining room. He could, e.g. be blameworthy for holding the ludicrous belief about Locke, insofar as his cognitive deficiency in the given scenario was simply a result of his presence in the philosophy course in the first place and by being so present, the student inexcusably violated an intellectual obligation.

In particular the following might perhaps be the case: Although the student is Alston-deficient with regard to philosophers' opinions and their books within typical course settings, he is also aware of this and correctly foresaw, at the time he chose

his courses, that should he enter a philosophy course, a heavy increase in the risk that he would form ill-based beliefs about the texts of philosophers would result. However, he is also aware that he has a great talent for, say, electronics. Suppose further that no external pressure forced him to pursue an education in philosophy rather than an education in electronics. The only thing that motivated him to enter the philosophy course was the desire to meet attractive girls together with the belief that the chances of meeting such girls were substantially higher within philosophy than within electronics.

In such a case it does not seem farfetched to say that, by enrolling in the philosophy course instead of, e.g. the electronics course, he inexcusably violated an intellectual obligation and did so with a foresight relevant to the epistemic undesirability of his later Locke belief. Further, had he not violated this obligation, he would never have formed an undesirable belief about Locke's doctrines. In that case then, on $\text{Blame}_{\text{epist, final}}$, he would be epistemically blameworthy for holding his Locke belief, despite the fact that, once enrolled in the philosophy course, he was radically Alston-deficient with regard to all philosophical matters and texts, and could have done nothing to remedy this.

The thrust of Steup's objection to Alston can thus be preserved in the following modification of Alston's claim in its generalized version: An agent's Alston-deficiency with regard to the subject-matter of an undesirable belief in a certain educational context can only excuse her from epistemic blameworthiness, insofar as her Alston-deficiency in the case was not brought about by one or more inexcusable violations of intellectual obligations performed with a relevant foresight or blameworthy inadvertence to risk. In contrast, as seen above Alston-deficiency does not excuse, if the agent's Alston-deficiency was a result of her inexcusable violations of intellectual obligations performed with a relevant foresight or blameworthy inadvertence to risk.

14.4. DOES THE PROPOSED ANALYSIS OF THE NOTION OF EPISTEMIC BLAMEWORTHINESS HAVE A VICIOUS REGRESS PROBLEM?

James A. Montmarquet has argued that any defence of epistemic deontology, which, like the present, includes an appeal to indirect modes of doxastic control, is for that reason involved in a vicious regress. He presents his argument to this conclusion in the following passage:

Many philosophers will readily concede that we bear a kind of *indirect* responsibility at times for what we believe – indirect because it depends on a more direct responsibility for actions (and omissions of actions). The main problem, however, with this indirect view is that it fails to push the issue far enough. Suppose that an agent does culpably fail to take some epistemically needed action (e.g. fails to look for more evidence in support of her opinion, when more evidence is clearly required). Notice that in such an instance she will believe that no more evidence is required. Thus, if the argument alluded to earlier [the one conceded by “many philosophers”] is sound ..., if we are to hold her culpable for omitting to look for more evidence, we must hold her culpable for believing that she needn't so look. Moreover, if the latter culpability is to be made out in terms of some further action or omission – she failed to check to see whether her belief

that further checks were unnecessary was really justified – I will argue simply that we are on the road to a regress.⁵⁰⁹

Montmarquet obviously believes this regress to be *vicious*. Later he takes great care in spelling out that his property voluntarism only appeals to “modalities of the belief-forming process,” in particular the maintenance of “an appropriate frame of mind,” and does not involve appeals to “prior” actions or omissions.⁵¹⁰

I shall argue that in fact this regress, even if perhaps interminable in rare instances, is not a severe threat to my preferred conception of epistemic blameworthiness. Moreover, Montmarquet himself hardly escapes a similar regress worry.

To make good the last claim, consider again the details of Montmarquet’s property voluntarism: Montmarquet argues that the epistemic blameworthiness of a belief hinges entirely on the “care” that the believer invested in its process of formation. As he puts it:

Just as the carver’s care is expressed *in* his work, the believer’s care is expressed in the attitude (the care) she actually uses, or fails to use, in the process of evaluating and ultimately accepting or not accepting the propositions in questions.⁵¹¹

By now the limitations of this view should be obvious: As pointed out several times in this study, the typical belief-formation is not preceded by any “evaluation” of its content that may be described as carried out more or less carefully. Typically beliefs are simply forced upon us without preceding controlled deliberation or even the psychological possibility thereof. Further, an agent’s ability to undertake such “care” is seriously constrained by her general cognitive dispositions. As we saw in Section 9.3, this insight forces Montmarquet to acquit agents of epistemic blameworthiness for holding even very “vicious” beliefs, as long as these belief are “carefully” formed, a worry that Montmarquet recognizes but, as we saw in Section 9.3, his position did not command the resources to counter successfully.

However, these limitations are not even redeemed by the cost of avoiding a regress. Montmarquet’s own position has similar regress problems. In Montmarquet (1999), he goes to some length in arguing that applying “care” to a belief-formation is not something distinct from forming a belief in a careful way:

Doing something in a certain way is not to be likened to the familiar case of doing one thing by doing something else – something relatively more “basic” – as in the case of flicking a switch by moving a finger ... (...) ... Nor would it be very intuitive to analyze “whistling carefully” as involving an act of care causing certain physical changes in one’s lips, throat, or other physically relevant places. Return now to the target case of belief and believing carelessly. Here, by parity of reasoning, it should be clear that we can speak of an individual as having direct control over his degree of care and, to that extent direct (albeit incomplete) control over his forming the belief that he is forming. Thus to say that he has only indirect control on the grounds that his belief is a consequence (in part) of his lack of due care is a non sequitur, akin to maintaining that I had only indirect control over my whistling because I would not have whistled that way had I exerted due care.⁵¹²

In Section 9.2 I classified Montmarquet as a property voluntarist, since in some passages he clearly seems to reject doxastic (content) voluntarism. This rejection

he obliquely refers to above, when submitting that an agent has only “incomplete” control over the beliefs she is forming: In particular Montmarquet denies that an agent has complete control over her beliefs in the sense that “there is no claim here that the existence (or occurrence) of a given belief is itself subject to one’s control.”⁵¹³ There is no hint that he should have explicitly changed his position on this central issue between 1993 and 1999.

However, in the passage quoted above apparently doxastic voluntarism comes sneaking through the backdoor: Surely the proposed analogy between belief-formation and whistling is out of place, for whistling is an *action*, no matter the care involved. However, given the demise of doxastic voluntarism, forming a belief is *not* an action at all. Thus, it might very well be true that “careful whistling” is a specific *kind* of whistling rather than a distinct act, and that “applying care to whistling” and “whistling carefully” are only equally legitimate descriptions of the very same action. However, even if Montmarquet’s *non sequitur* is relevant to whistling, it hardly applies to belief-formation: Since doxastic voluntarism fails, there is no action that “forming a belief carefully” could be. “Applying epistemic care,” insofar as it is an action, thus cannot be another description of *that* (non-)action.

Still, this does not mean that an “application of epistemic care” is “doing something to do something else,” as in the case of flicking a switch by moving a finger. In fact, when applying “care” in forming a belief, the application of care is all that is really *done*, and what is done here can only be described as an effort to influence a certain result: the belief-state actually formed. This means, that all “epistemic responsibility” is really *indirect* in Montmarquet’s sense: Even on accounts of epistemic blameworthiness like Montmarquet’s, which appeal exclusively to modes of *direct* doxastic control, epistemic blameworthiness is a species of blameworthiness for the consequences of actions or omissions.

As a consequence, Montmarquet’s property voluntarism has no real advantage over, e.g. property Pascalianism when it comes to meeting his own regress objection: In both cases the agent is deemed blameworthy for holding a belief in virtue of having performed certain actions or omissions influencing the relevant belief, and in neither case can this blameworthiness be established, unless these belief-influencing actions or omissions were inexcusable violations of the agent’s intellectual obligations. In answering Montmarquet’s regress worry, as I shall now proceed to do, I am answering a worry confronting any viable account of epistemic blameworthiness, including Montmarquet’s own.

It should be noticed that in formulating his objection, Montmarquet certainly constrains the “epistemically needed action” referred to in an unnatural way by only mentioning a “culpable” lack of evidence-gathering in a situation where the agent has *already* formed an “opinion.” It seems quite correct, as Montmarquet submits, that in such a case the agent will typically not believe that additional evidence-gathering is needed, and will in effect then be able to offer an “honest mistake” defence for failing to perform the “epistemically needed action” (i.e. for violating an intellectual obligation, in my preferred terminology). As we have seen in Section 1.3, if an

honest mistake is really present, such a defence can only be countered by establishing that the belief that no further evidence-gathering was needed was itself epistemically blameworthy, and a regress of some kind arises.

However, as argued in Section 10.5, most “epistemically needed actions” (or in my preferred terminology: actions, which the agent stands under an intellectual obligation to perform), occur at a time when the agent has as yet no opinion concerning the *particular* propositional contents of the undesirable beliefs that may result from her failure to perform such an action. A ready example here is that of the educated racist, as the example has been presented above: At the time the racist supposedly violated an intellectual obligation by being inattentive in class, she need not have had any opinion concerning the proposition that blacks are generally inferior to whites, the proposition that later became the content of her epistemically blameworthy belief. This is so, even if she were able to foresee that she would be likely to form *some* undesirable racist belief as a result of her inattentiveness. Thus, in such cases Montmarquet’s regress worry as it stands is beside the point: Here the agent has not formed any “opinion” that may in the least affect her motivation to undertake further evidence-gathering.

Still, however, a regress worry lurks in the vicinity. As argued in Section 13.2 a violation of an intellectual obligation committed on the acceptance that no increase in the risk of epistemically undesirable occurrences would result, could be blameworthy, only if this acceptance was itself somehow blameworthy. Further, I claimed that if this acceptance was prompted by a belief with the same propositional content, this acceptance is blameworthy in the relevant sense, only if the prompting belief is epistemically blameworthy. In effect I am here simply pushing a second-order “honest mistake” case: If the supposed intellectual violation was violated in the “honestly mistaken” *belief* that no increase in the risk of undesirable consequences obtained, this belief would have to be again epistemically blameworthy to drive home the epistemic blameworthiness of the undesirable belief resulting from the violation of the intellectual obligation. Regress threatens again.

However, I take this regress to be quite benign. In fact it is not different from the regress worry that befalls any viable analysis of a notion of non-excusability for bringing about some undesirable result, be it undesirable from a moral, epistemic, aesthetic, or other perspective: No matter which conditions not involving excusability are set up such that, supposedly, if an agent satisfies these conditions, she is not appropriately excused for bringing about certain undesirable consequences, it must plausibly be allowed that the agent can get off the hook again, if she is only appropriately *excused* for not satisfying these conditions. That is: in order to be blameworthy for bringing about the undesirable consequences, she must not only satisfy certain conditions, these conditions must also *include* the condition, that she is *not appropriately excused* for not satisfying the other conditions. In this sense, non-excusability is a non-reducible notion: No matter how many levels of conditions for non-excusability are explicated, there is no way to get rid of an additional non-excusability clause.

However, this does not mean that, on all levels, the relevant non-excusability condition is not satisfied. Only this situation would cause a vicious regress problem, since

in that case no agent could be blameworthy for anything after all. To concentrate on the case of $\text{Blame}_{\text{epist, final}}$, there are two regress worries here: one arising from condition (4b) and one arising from condition (5). The worry arises, because in both cases the agent might sometimes satisfy the condition, only if she is epistemically blameworthy for holding a further belief: In (4b), this problem occurs if her inadvertence to risk was prompted by a belief that no increase in risk of the relevant sort obtained. In (5), the problem occurs if, in performing the actions or omissions constituting the violation of the relevant intellectual obligations, the agent believed that she did not violate any intellectual obligations at all; or if she believed that she had an appropriate excuse for violating her intellectual obligations in the relevant situation. Clearly, to defeat an “honest mistake” defence here, cf. Section 1.3 we will have to argue that the agent was epistemically blameworthy for holding such a belief.

It should however be clear on this stage that the regress comes to a benign halt, unless on *all* levels either (1) the agent’s violation of her relevant intellectual obligation was prompted by a belief that no violation occurred or that she had an appropriate excuse *or* (2) she violated this obligation with a relevant inadvertence to risk prompted by a belief, rather than, e.g. with foresight. This does not seem very plausible: Certainly we would think that a normal agent may often violate an intellectual obligation without being “honestly mistaken” in doing so and that she performs at least some of her belief-influencing actions or omissions with a relevant foresight.

Consider again our stock-example, the educated racist: Most plausibly, when she skipped her racial issues class, as a result of which she later came to hold an epistemically undesirable racist belief, she did not do this with a foresight relevant to the deontic evaluation of her later racist belief. More likely, she did not give a thought to the risk of undesirable consequences resulting from her absence. However, she may well have skipped the class with a blameworthy inadvertence to risk: If she had in fact skipped the class on the acceptance that no increase in a relevant risk obtained, this acceptance would plausibly have been blameworthy. Since she did not *believe* the propositional content of the above-mentioned acceptance, no “honest mistake” defence applies here.

Thus, the racist could only get off the hook if she did not satisfy condition (5) of $\text{Blame}_{\text{epist, final}}$, that is: if she was appropriately excused for skipping the class. This could, e.g. be the case, if her roommate administered a sleeping pill to her without her knowledge etc. However, it is easily imaginable that no excuse of either the “honest mistake” or a more profane type applies. Thus the educated racist can be held epistemically blameworthy for her racist belief without regress worries.

It seems safe to conclude, then, that $\text{Blame}_{\text{epist, final}}$ is not severely threatened by the regress worry obliquely presented by Montmarquet. The envisaged regress is not likely to be interminable in standard cases. Further, if it should prove interminable in some exotic cases, probably the only viable response is to admit that to such cases of undesirable belief the notion of epistemic blameworthiness evolved here does not meaningfully apply.

EPISTEMIC AUTONOMY

Abstract. In this chapter I suggest one viable analysis of the notoriously elusive notion of epistemic autonomy. I maintain that a fruitful way of understanding the notion of an autonomously held belief may be as a belief for which the believer is responsible, yet is also epistemically blameless in holding. In Section 15.1 I present in outline the Kantian notion of autonomous action. I then argue in Section 15.2 that the analysis of epistemic autonomy presented here maintains some measure of continuity with the Kantian notion, despite the demise of doxastic voluntarism. In Section 15.3 I proceed to present an influential alternative notion of epistemic autonomy, applying to certain roles in a community of cognizers rather than to individual agents. This is a notion, with which the present notion should not be confused. Finally, in Section 15.4 I briefly argue that my proposed understanding of epistemic autonomy at least satisfies some central intuitions concerning epistemic autonomy and avoids some imminent excesses in the bargain.

15.1. THE KANTIAN CONCEPT OF PRACTICAL AUTONOMY

In his *Grundlegung zur Metaphysik der Sitten* Immanuel Kant introduces the concept of practical⁵¹⁴ autonomy (Autonomie) as *self-legislation*. He here famously argued that a truly free, or autonomous, will is *in itself* the only fundamental good or value, not in virtue of complying to external standards, but simply in virtue of *being* autonomous in its motivations of action.

In the present context I am sadly unable to do anything like full justice to the subtleties of Kant's *Grundlegung*. However a brief discussion of the complex and entangled structure of the Kantian concept of practical autonomy will prove highly relevant to the further proceedings.

According to Kant, an autonomous action is an action governed by the agent's *self-legislation*, i.e. motivated by respect⁵¹⁵ for the maxims that the agent herself sets for her practical conduct. Being so motivated is a matter of the action's being entirely determined by the *will* to conform to these maxims, unconstrained by will-external causal factors:

Der *Wille* ist eine Art von Kausalität lebender Wesen, sofern sie vernünftig sind, und *Freiheit* würde diejenige Eigenschaft dieser Kausalität sein, da sie unabhängig von fremden sie *bestimmenden* Ursachen wirkend sein kann.⁵¹⁶

A necessary condition for autonomous action, then, is the agent's *unconstrained voluntary control* over that action. However, the matter quickly gets complicated: Not just any self-imposed maxim will do for autonomous action. If the agent bases her maxims even partially on properties of the external world, her actions inevitably become non-autonomous, i.e. *heteronomous*: Her will has then allowed itself to be determined by factors outside of itself, and thus, in Kant's opinion, cannot operate without constraints:

Wenn der Wille irgend *worin anders*, als in der Tauglichkeit seiner Maximen zu seiner eigenen allgemeinen Gesetzgebung, mithin, wenn er, indem er über sich selbst hinausgeht, in der Beschaffenheit irgend eines Objektes das Gesetz sucht, das ihn bestimmen sollen so kommt jederzeit *Heteronomie* heraus. Der Wille gibt alsdann sich nicht selbst, sondern das Objekt durch sein Verhältnis zum Willen gibt diesem das Gesetz.⁵¹⁷

This line of reasoning implicates that in order for the will to be free and thus capable of autonomous action it must apply only maxims springing from *itself*, i.e. maxims not dependent upon any state of the world or “properties of objects” (Beschaffenheit irgend eines Objektes). On the face of it, it seems hard to imagine, which kind of maxim could satisfy this criterion and still have implications for actual action, but Kant believes that exactly one maxim does the job: The famous *categorical imperative* or *Moral Law*:

Der *Wille* ist *schlechterdings gut*, der nicht böse sein, mithin dessen *Maxime*, wenn sie zu einem allgemeinen Gesetze gemacht wird, sich selbst niemals widerstreiten kann. Dieses Prinzip ist also auch sein oberstes Gesetz: handle jederzeit nach derjenigen *Maxime*, deren *Allgemeinheit* als Gesetzes du zugleich wollen kannst; dieses ist die einzige Bedingung, unter der ein *Wille* niemals mit sich selbst im Widerstreite sein kann, und ein solches Imperativ ist kategorisch.⁵¹⁸

That is: As long as an action is only motivated by the maxim to act only according to the maxims that the agent can will to establish as general laws of conduct, the agent’s will is unconstrained by external factors and thus capable of autonomous action. But this has a very important implication: In order to be autonomous, an action must comply to the Moral Law and thus be *simply good* (schlechterdings gut): A necessary condition for autonomous action is thus that the action is good, or, to put it shortly: According to Kant an action is only good if it is free, and only free if it is good! Further, since that the ideal of autonomy is clearly prescriptive⁵¹⁹; an action is only autonomous if it is as it *should* be or *ought* to be. Kantian goodness is arguably a deontic concept.

However, as complicated as matters already are, they can get still further entwined: According to a reading of *Grundlegung* proposed by Christine Korsgaard, Kant’s ethics revolves around the issue of *self-constitution*: Being good in the Kantian sense of self-legislation allows us to constitute ourselves as unified agents.⁵²⁰ But, equally important:

Kant thinks that what makes an action attributable to the person is that it springs from the person’s autonomy, or self-government. The exercise of the person’s autonomy is what makes the action *his*, and so what makes it an action.⁵²¹

This immediately poses the following problem, as Korsgaard is quick to notice⁵²²: We must assume that in order for normative predicates like “good” or “bad” to apply to an agent or person with respect to an action, the action must be *that* agent’s action. However, if the attributability of an action to the agent demands that action to be autonomous, hence good, it is hard to see how any agent can ever act badly. This presents a problem, since we believe many actual actions to be bad ones.

Korsgaard tries to solve the problem by arguing that bad actions may still be attributed to the agent *qua* defective agent. Hence a bad agent is really not her full

integrated self in her bad-doing, but still enough of herself for attributability. I will not here comment upon this solution. My purpose is rather to bring out the implications of a decisive third condition for Kantian practical autonomy: An action is only autonomous, if it is attributable to the agent. cf. Korsgaard attributability in the relevant sense may depend upon actually realizing the ideal of autonomy to a high degree, i.e. acting freely and as one should, which according to Kant comes to one and the same. To rephrase somewhat dangerously: For an action to be autonomous, the agent must be responsible⁵²³ for that action, and responsibility hinges on the agent's freedom or voluntary control of that action.

It is time to sum up the three intertwined components of the subtle Kantian concept of practical autonomy: An action is (practically) autonomous in the Kantian sense if, and only if,⁵²⁴ it satisfies the following conditions:

1. $Aut_{\text{practical, volitional}}$: The action is under the agent's direct and unconstrained voluntary control.
2. $Aut_{\text{practical, normative}}$: The action is morally *good* according to a prescriptive criterion, i.e. not morally blameworthy.
3. $Aut_{\text{practical, metaphysical}}$: The action is the agent's *responsibility* in the sense that it is *attributable* to the agent in the sense required for application of deontological predicates (like "blameworthy") to the agent with respect to that action.

As should be clear from the above discussion, the above three conditions are not logically distinct. David Owens is basically well guided, when he represents "Enlightenment rationalism" as committed to the view that "normativity implies responsibility and responsibility requires control."⁵²⁵ However, if Kant is taken as an Enlightenment rationalist, and I think this is unavoidable if this appellation is to carry any weight, the relationship between control, responsibility and normativity is still more complex: for example control requires normativity as much as normativity requires control.

15.2. APPROACHING A NOTION OF EPISTEMIC AUTONOMY

This is not the place to evaluate, e.g. how the concept of practical autonomy out-lined above stands with respect to our uses of the concept of autonomy in standard ethical evaluations.⁵²⁶ Rather this section will take its outset in the following question:

Does the above highly unified concept of practical autonomy have an epistemological counterpart? In other words, is it possible to construct a fruitful concept of *epistemic autonomy* on the exact model of Kantian practical autonomy?

My answer to this question will be a resounding "no!." Epistemic autonomy fails to resemble the unified Kantian concept of practical autonomy for the simple reason that practical volitional autonomy (the first condition above) has no relevant epistemological counterpart. As argued in Chapter 8, belief-contents may not be brought under systematic direct voluntary control by the believing agent. In fact, as emerged in Section 10.6, if our belief-contents are ever controlled by a will in any sense, that will is most likely not our own. In this sense our prospects of Kantian-style doxastic self-government are dim indeed.

Now, given the intricate interweaving of the three aspects of Kantian practical autonomy, it might be tempting to conclude at this point that a concept of epistemic autonomy can have nothing to do with the Kantian concept of practical autonomy at all: As we saw above, the very content of the Kantian notions of responsibility and goodness depended on the possibility of voluntary control over actions. However, as has emerged from my analysis of the notion of epistemic blameworthiness, this notion is closely linked to the notions of responsibility (authorship of the epistemic quality of beliefs) and voluntary control (over belief-influencing actions and omissions). This being said I shall proceed in laying out the basis for a discussion of epistemic autonomy in a suitably modified Kantian framework.

The relevant counterparts of the three aspects of Kantian practical autonomy seem to be:

1. $Aut_{epistemic, volitional}$: Holding the belief is under the agent's direct and unconstrained *voluntary control*.
2. $Aut_{epistemic, normative}$: The holding of the belief is not epistemically blameworthy.
3. $Aut_{epistemic, metaphysical}$: The belief is the agent's *responsibility* in the sense that it is *attributable* to the agent in the sense required for the application of deontic predicates (like "blameworthy") to the agent with respect to her holding of that belief.

Following the demise of doxastic voluntarism the Kantian knot must be untied in the epistemological case. The unified elegance of the Kantian notion of practical autonomy cannot be retained. Still, the gist of the Kantian notion can be preserved in the doxastic realm by the following definition:

I shall say that an *agent* is epistemically *agent-autonomous* (in contrast to her being role-autonomous, see below) in her holding of a belief if, and only if, her holding of that belief is *normatively and metaphysically autonomous*, i.e. an appropriate object of epistemic deontic evaluation and epistemically blameless (not epistemically blameworthy). In what follows, most of the time I shall tacitly assume the indexing of agent-autonomy to specific holdings of beliefs and simply talk of agent-autonomy.

As we have seen in Section 7.1, if a belief is the agent's *responsibility* in the present sense, she enjoyed at least some mode of doxastic control over it, even if (see Chapter 8) its content was not therefore under her direct voluntary control. Following the demise of doxastic voluntarism, this is as close an analogue to practical volitional autonomy as we get in the doxastic realm. Now further, if her belief is also epistemically blameless (normatively autonomous) it is the case that, if her belief is epistemically undesirable, even though the agent enjoyed some mode of doxastic control over it, in exercising or omitting to exercise that control she violated no intellectual obligations incumbent upon her with either a relevant foresight or a relevant blameworthy inadvertence to risk (see Chapters 12 and 13).

The notion of epistemic autonomy is notoriously elusive, and it would be severely foolhardy to submit a single analysis as adequate for all purposes. Still, I shall argue below, the conception advanced above brings with it certain advantages. First, though,

I shall point to a quite common conception of epistemic autonomy related to – but crucially different from – the present.

15.3. EPISTEMIC AGENT-AUTONOMY AND EPISTEMIC ROLE-AUTONOMY

In order to get a more secure grip on the present concept of epistemic agent-autonomy, note that the notion of epistemic autonomy has sometimes been used in a sense quite different from the one invoked above, namely as resembling the German concept of “*Lehrfreiheit*,”⁵²⁷ meaning basically the privilege of an unfettered choice of research projects and teaching curricula enjoyed by university faculty members. This concept closely resembles the concept of academic freedom invoked by Michael Polanyi:

Academic freedom consists in the right to choose one’s own problem for investigation, to conduct research free from any outside control, and to teach one’s subject in the light of one’s own opinions.⁵²⁸

From the outset this is a state realized by individuals occupying certain *roles* in a suitably arranged political system, or to put it in Polanyi’s gnomic phrasing: Freedom is an efficient form of organization.⁵²⁹ Thus, epistemic autonomy in his sense is not primarily a governing norm for individual doxastic conduct, but rather a norm for the construction of societies. Despite, or rather *because of*, their interconnectedness and the ensuing danger of equivocation I shall term the Polanyi-style concept “epistemic role-autonomy” (as, being a social privilege, it attaches to certain social roles) and contrast it with epistemic agent-autonomy, the epistemic autonomy ascribed to an agent *qua* agent. I shall say that an agent is epistemically role-autonomous in regard to a certain social role if, and only if, she occupies that role and that role does not involve any *forced relations of epistemic deference*. Academic freedom in this sense is then best seen as a kind of role-autonomy applying specifically to the role of an academic researcher.

Obscuring the difference between epistemic role-autonomy and epistemic agent-autonomy may have dire consequences. For example Lynne Arnault’s conception of autonomy as membership of a group that allows its members to express their points of view in a natural, undistorted and unrepressed way, is obviously best construed as a brand of epistemic role-autonomy.⁵³⁰ However if interpreted thus, Arnault’s conception is open to criticism. As Trudy Govier has pointed out,⁵³¹ agents, despite belonging to a group, may wish to speak and act in ways not representative of that group. Also it seems clear that agents, despite belonging to a certain privileged group, may easily end up in situations where the privileges attached to that particular group-membership are void. An example may be that of a free-press journalist ending up as a terrorist hostage. Such cases illustrate the importance of indexing role-autonomy to agents in regard to specific social roles rather than to agents simpliciter.

It might be thought now that since agents always occupy some social role or other, the distinction between agent-autonomy and role-autonomy is futile. This impression is misleading: Agent-autonomy applies to agents *whatever* role they occupy, whereas role-autonomy applies to *specific* roles occupied by the agent. However as the concept of agent-autonomy involves reference to intellectual obligations, and since these

may well be contingent upon social positions (see Section 12.3), agent-autonomy is not insensitive to social roles either. Still, agent-autonomy and role-autonomy are concepts with a very different structure: Role-autonomy puts demands on the society (or societies) of which an agent is a member. Agent-autonomy puts demands on particular agents. Of course, if agent-status should in some cases apply to a society as a whole (e.g. when a state declares war on another), the concept of agent-autonomy in such cases may apply to societies as well. Conversely, if a single person is construable as a “multiple self” containing within her doxastic life several distinct agents in a kind of quasi-social interaction, the concept of role-autonomy may apply to each such agent within the person’s cognitive economy.

A brief consideration of two historical case studies will perhaps further clarify the relevant distinction: The stoic philosopher Epictetus hailed the ideal of the calm agent, whose desires or desirability judgments are in perfect consonance with the events of the external world, even if this meant a radical policy of compliance with harsh living conditions:

That man is free, who lives as he wishes, who is proof against compulsion and hindrance and violence, whose impulses are untrammelled, who gets what he wills to get and avoids what he wills to avoid.⁵³²

Epictetus’ “calm stoic” is stipulated to be epistemically agent-autonomous: in Epictetus’ eyes she certainly leads an impeccable doxastic life. However she may well be radically epistemically role-heteronomous as she allows herself to be guided by the contingencies of her environment, possibly including an epistemically deferent role as that of a confined slave crucially reliant on her masters for vital information about the world. Plausibly it is exactly this blatant disregard for the value of role-autonomy in Epictetus’ writings, which has brought Steve Fuller to diagnose him as engaged in a project of systematic self-deception.⁵³³

Kant’s autonomous agent, on the other hand, is radically epistemically role-autonomous in her “role” as a citizen in the Kingdom of Ends,⁵³⁴ insofar as her will is allowed to stand under no external causal influences whatsoever, let alone any forced relations of epistemic deference. The radical nature of this claim is perhaps best brought out by considering the following consequence: Kant’s autonomous agent is not allowed to let the Moral Law guide her actions even partially out of respect for Kant’s writings; she must recognize its compelling force entirely on her own. Arguably this imposes overly strong demands on the average agent. If Epictetus showed no regard for the value of role-autonomy, perhaps Kant erred in the other direction. He valued role-autonomy (at least concerning moral judgements) so much, that he put a complete ban on deference in moral matters, whether it be forced or unforced.

15.4. EPISTEMIC AGENT-AUTONOMY EXAMINED

Epistemic agent-autonomy in the sense introduced in Section 15.2 captures an intuitively desirable aspect of a concept of epistemic autonomy, namely that the agent is *self-reliant* in her autonomously held beliefs, in the sense that, being responsible, yet blameless, either her autonomous belief is not epistemically undesirable, or

she has not violated any intellectual obligations relevant to its epistemic quality in a blame-inducing way (with a relevant foresight or inadvertence to risk). Either way her holding of her autonomously held belief does not give rise to a negative evaluation of her intellectual conduct.

This nicely captures the sense in which the production of epistemically autonomous students is a goal of general education: General education, among other things, arguably aims at making us blameless believers: It aims to make us reasonable believers forming beliefs in a reliable way and on adequate grounds. If our beliefs fall short of these standards, it aspires to make sure that this is not due to our vicious service to the truth-goal: It teaches us to fulfil our intellectual obligations and be alert to the risk of impeding our belief-forming dispositions.

However some may feel that the above goal is not reached until the student *habitually* holds beliefs in a blameless fashion, and that occasional or even consistent, but strainfully achieved, success in this respect is not enough to merit an ascription of agent-autonomy proper.

On this approach, epistemic agent-autonomy is a *virtue* in the sense of being a habitually engrained quality (here: the quality of forming (revising, rejecting) beliefs in a blameless fashion.⁵³⁵) I will have no objection against this understanding of agent-autonomy, which I will allow to stand side by side with the piece-meal understanding offered above. It is however crucial to notice that habituality does not do any substantial normative work in the present context, and that “virtue” is rather ill fit to work as a fundamental normative concept: On pains of parody, no virtue or habit can work without the possibility of occasional malfunctions. Human beings are regrettably so constituted that even the most habitually engrained quality may fail to surface now and then. Thus, in evaluating an agent’s merit concerning her holding of a specific belief, it is insufficient to point out that she possesses “the virtue of agent-autonomy.” Rather we will need to consider if her habit were in fact satisfactorily instantiated in the present case. Hence we are back with the piece-meal conception of agent-autonomy as introduced above. In each case, this concept takes the normative toll.

The virtue theorist might now take refuge in insisting that an adequate notion of agent-autonomy does not apply to singular holdings of beliefs at all, but rather to an agent’s doxastic life in its entirety or at least to significant portions of it. However this approach suffers from the defect that agent-autonomy has been traditionally construed (as with Frederick F. Schmitt below) as applying to the management of epistemic sources like testimony that deliver beliefs in a piece-meal fashion. Overall, I shall therefore prefer a piece-meal conception of epistemic agent-autonomy. This policy also retains a maximum of consistency with the Kantian concept of practical autonomy. Surely Kant would have balked at the suggestion that practical autonomy should be some form of *habit*: He certainly conceived of respect (*Achtung*) for the moral law as something actively manifested in *each* singular instance of autonomous action, not as something slavishly engrained.

The term “autonomy” has strong individualistic connotations. As seen, its famous use in Kant’s *Grundlegung* is markedly individualistic: On Kant’s view, an agent’s

practical autonomy need not in the least depend on her interaction with other agents: In fact the autonomous agent is not allowed to let her actions determine by them to any degree. It is not surprising, then, that talk of autonomy in epistemological contexts has centered on the management of beliefs acquired through social epistemic sources, as e.g. the testimony of other agents. The question has been whether an agent can uphold any kind of epistemic autonomy in a world where she must base most of her beliefs on social epistemic sources like testimonial evidence.⁵³⁶ Frederick F. Scmitt has invoked the concept of (intellectual) autonomy in this vein:

If our earlier argument is sound, most justified beliefs are justified on nonHumean testimony. We must ask, then, whether there is any room left for autonomy in the broader sense which encompasses both control and thinking for oneself. We do prize such autonomy. We believe people should think for themselves, be critical, monitor and correct their own cognitive activity.⁵³⁷

It is precisely these issues of doxastic control and intellectual obligation, which have occupied the bulk of this work. In terms of the present framework, it seems plausible that in the passage above (intellectual) autonomy is particularly linked to the fulfilling of special, as it were, “counter-social” intellectual obligations (of e.g. honing a critical stance towards testimonial sources). This focus on autonomy as a (prickly) individualist defence against the social⁵³⁸ has earned the concept a bad reputation in some quarters of the philosophical environment, where the Enlightenment-style “Autonomous Reasoner” has been easily satirized as a “neurotic fantasy.”⁵³⁹ Some feminist philosophers have even contended that the autonomous agent, with her “detached, impartial, neutral and self-reliant” doxastic life, could only be the figment of an oppressive male imagination.⁵⁴⁰

I believe that intellectual obligations with regard to social epistemic sources invoked by such (male) Enlightenment thinkers as Locke, Hume and Kant, are plausibly construed in such a way that the force of the feminist objection can be absorbed without abolishing the ideal of epistemic autonomy altogether, although I shall not argue this intricate point here.⁵⁴¹ Suffice it to notice that these authors, as well as later adherents of the Enlightenment tradition like W.K. Clifford, share an idea of a strong obligation imposed on any agent to avoid realizing the epistemic “cardinal vice” of credulity or gullibility in relation to social epistemic sources. In Clifford’s words:

But I cannot help doing this great wrong towards Man, that I make myself credulous. The danger to society is not merely that I should believe wrong things, though that is great enough; but that I should become credulous, and lose the habit of testing things and inquiring into them; for then it must sink back into savagery.⁵⁴²

Everything depends then on the strictures imposed on gullibility. I suspect that in order to take proper account of the psychological and social embedding of human cognition, these strictures may need to be relaxed quite a lot from those suggested by the above-mentioned Enlightenment authors and their later followers, insofar as the ideal of non-gullibility is to be realizable at all. However, given a moderate interpretation of gullibility, “neurotic fantasies” should be avoidable.

As I have suggested, being epistemically autonomous in the sense of having beliefs which one has authored, yet have not influenced in blame-inducing ways, does not

mean a departure from the social imbedding of one's cognitive system, but rather a deep-seated interaction with it. This is because fulfilling one's intellectual obligations may well demand one to take advantage of one's educational opportunities and allow one's cognitive system to be adjusted so as to honour the intellectual ideals of one's society. Thus, given the conception of epistemic agent-autonomy opted for here, we not only preserve some measure of continuity with the important Kantian concept of practical autonomy; we also maintain that an ideal of epistemic autonomy, properly understood, does not require one to live the prodigious intellectual life of a supercilious hermit.

NOTES

1. Clifford (1999) p. 78.
2. In the present study the epistemic dimension of belief-evaluation will be tied essentially to the goal of having only true beliefs (the truth-goal). See Chapter 5 for an in-depth discussion. The epistemic dimension is not the only possible dimension of belief-evaluation, although undeniably a highly important one: Beliefs may, e.g. also be evaluated for their aesthetic merits.
3. “Epistemic deontologism includes, among other things, a commitment to the idea that we can legitimately be reproached, scorned and blamed for not believing what we ought to believe and that we can legitimately be praised and the like for believing what we ought to believe.” Ryan (2003) p. 48. Ryan does not make it explicit in this passage that epistemic deontologism concerns *epistemic*, rather than say, aesthetic, dimensions of reproach, blame, praise etc. This, however, should be clear from the context.
4. The “educated racist” is a favourite of the literature on deontic epistemic normativity invoked in very similar variants in such texts as Russell (2001) and Steup (2000). I shall return to this example in several contexts below and discuss exactly under which circumstances the educated racist may be epistemically blameworthy for holding her racist belief.
5. (1975) All ER pp. 347–383. Court of Appeal Decision (1975) 1 All ER p. 8ff. This case is perhaps the single most discussed rape case of all time. Curley (1976), Husak & Thomas (1992), Archard (1999) and Baron (2001) are among the many papers commenting on the *Morgan* case.
6. The nature of the *actus reus*, the “social harm” (Dressler (2001) p. 115), covered by the term “rape” has been an issue where jurisdictions have crucially diverged and commentators have persistently disagreed. See Dressler (1998) 418–429 for a thorough review of the debate. An account based entirely on the notion of unwanted sex has been offered by some commentators (see, e.g. Husak & Thomas (1992) p. 119 n56). Others have stressed the presence of physical force (and fierce resistance). However, non-consensual sex seems a reasonable way of defining rape, even if the nature of sexual consent is in itself a very intricate issue. This was indeed how rape was defined under British jurisdiction at the time *Morgan* went to trial (see Curley (1976) p. 352).
7. (1975) All ER pp. 400–401.
8. Curley (1976) p. 325.
9. (1975) 2 All ER pp. 1060–1063. In this cruel case, a man, Leak, punished his wife by forcing her to have sex with his friend Cogan, who was severely drunk at the time. See Baron (2001) p. 12 for comments on the case.
10. The High Court of Australia had declared its formal independence of British decisions in 1963 after another highly controversial ruling by the Law Lords. On the appeal of the 1961 case *Smith v Director of Public Prosecutions* (A.C. 290), The Law Lords ruled that a defendant can be judged guilty of murder, if she is not insane or incapable of forming an intent, even if in the actual case she had no intentions to kill the victim. In 1967 the *Smith* decision was overturned in the British jurisdiction as well. However, the damage was done and the *Morgan* case resulted in an even greater estrangement. See Curley (1976) p. 336.
11. Curley (1976) p. 325. See *ibid.* for a dramatic report of the outrage, written shortly after the events.
12. Here I will assume that each of the men had a conception of the nature of sexual consent accurate enough to warrant the claim that he really believed that Mrs. Morgan consented. If he was severely mistaken as to what sexual consent is, this might arguably complicate the case further. However, the precise nature of sexual consent is perhaps the most complex topic of discussion relating to rape

issues. See Husak & Thomas (1992), Schulhofer (1992), McGregor (1996) and Archard (1999) for very divergent recent approaches to the subject.

13. In a similar overview of this scenario, David Archard prefers to coin this premise thus, using the neutral “Smith” for one of the men and “Jones” for Mrs. Morgan: “Smith would not have had sex with Jones, had he believed [that at the time of this sex, Jones did not consent to sex with Smith].” Archard (1999) p. 213. However, this formulation surely does not capture what is important in this case from a moral perspective. Suppose that this premise is not fulfilled, because, if Smith had had this belief s/he would have expressed it, and Jones would have gotten so angry, that s/he would have raped Smith! Intuitively, such counterfactual circumstances can neither add nor detract from Smith’s blameworthiness in the actual world. If Smith acted on his/her belief that Jones consented to have sex with him/her, the evaluation of Smith’s blameworthiness must take its starting point in this fact, notwithstanding how the actual world would perhaps have been if this had not been the case.
14. Smith (1983) p. 547.
15. Ibid. p. 556.
16. Even though it is hard to believe that in the *actual* Morgan case, the men did not have highly reprehensible desires.
17. Some authors have clearly missed out on this point. See especially Heller (2000), which shall be criticized later (Section 8.3.1.2).
18. Smith (1983) p. 555.
19. Montmarquet (1995) p. 46.
20. Curley (1975) p. 181.
21. Clifford (1999) p. 70.
22. Clifford also presents a case parallel to the ship-owner example, this time involving false accusations based on sincere beliefs. Here he judges that the accusers, even if acting on sincere beliefs, are “no longer to be counted honourable men” insofar as “they had no right to believe on such evidence as were before them.” Ibid. p. 72.
23. Ibid. p. 72.
24. Ibid. p. 73.
25. Haack (2001) p. 26: “Neither here nor elsewhere in the essay does Clifford ever distinguish ‘it is epistemically wrong’ from ‘it is morally wrong.’”
26. See *ibid.* p. 21. Also: “He [Clifford] offers, in other words, only arguments that could, at most, establish the correlation thesis_J. It is illuminating, as a further test of the claim that the correlation thesis_J is not true ... to show how Clifford’s attempted extrapolation fails.” Ibid. p. 26.
27. “Nevertheless, Clifford’s judgment of this case [the ‘ship-owner example’] seems correct: it is a case of morally culpable ignorance ...” Ibid. p. 26.
28. Clifford seems to change freely between talking of duties of conducting inquiry in a certain way (irrespective of the ensuing beliefs) and duties of holding beliefs in a certain way. Obviously Haack seems to have noticed only the latter kind of duties in Clifford’s writing, whereas the head-line suggests that the former is the central one. This impression is only strengthened by noting that when Clifford speaks of believing for “unworthy” or “insufficient” reasons he seems to mean holding a belief not “honestly earned in patient investigation.” See Clifford (1999) pp. 70, 74.
29. Clifford (1999) pp. 75–76.
30. To those suspicious of involving the notion of truth in these matters, remark that Clifford also assumes the existence of a specific “epistemic conscience”: “And the question which our conscience is always asking about that which we are tempted to believe is not, ‘is it comfortable and pleasant?’ but, ‘Is it true?’” Ibid. p. 83.
31. Ibid. p. 87.
32. See, e.g. Archard (1999) p. 216: “..[F]or Smith to be judged culpable, it is necessary also that his belief that Jones is consenting is, in some sense, unreasonable.” Baron (2001) p. 2: “... only reasonable mistakes should exculpate.”
33. Clifford expresses this point thus, referring to his ship-owner example: “The question of right and wrong has to do with the origin of his belief, not the matter of it; not what it was, but how he got

- it; not whether it turned out true or false, but whether he had a right to believe on such evidence as was before him.” Clifford (1999) p. 71.
34. It might be thought odd to describe the ship-owner’s belief as an honest mistake. However, remember, that Clifford expressly asks us to assume that the ship-owner sincerely believed in the soundness of his ship at the time it set sails.
 35. Dressler (2001) p. 115.
 36. See *ibid.* p. 143ff. A crime typically under strict liability is, e.g. statutory rape of young children. Here it is typically considered of public interest that an offender gets convicted if it can only be proved that s/he had sexual intercourse with the child.
 37. *Ibid.* p. 116.
 38. See Saltzburg et al. (1994) p. 196.
 39. Dressler (2001) p. 116.
 40. MPC Section 2.02 (2)(a)(i). I take it to be clear here that Smith’s conduct is an element of the offence. All references to MPC quoted from Saltzburg et al. (1994) pp. 199–200.
 41. MPC Section 2.02, (2)(b)(i).
 42. Those are exactly that in order to be guilty of an offence, the person must at least have acted with a *mens rea*, i.e. either purposely, knowingly, recklessly or negligently. MPC Section 2.02 (1).
 43. MPC Section 2.02, (2)(c).
 44. Curley (1976) p. 347. Curley explicitly quotes the MPC’s definition of recklessness.
 45. *Ibid.* p. 348.
 46. *Ibid.* p. 349.
 47. *Ibid.* p. 348.
 48. MPC 2.02 2 d.
 49. Hall (1963) p. 632.
 50. *Ibid.* p. 636.
 51. Dressler (2001) p. 116.
 52. “If the RPAS [reasonable person in the actor’s situation] knows only what the defendant knows, however, there is never any negligence either...; The RPAS will always act as the defendant acted where the defendant is not conscious of the risk, and will act differently only where the defendant is conscious of the risk, i.e. reckless.” Alexander (1990) p. 99. As Smith was supposedly not conscious of the risk of raping Mrs. Morgan, he did not perceive a reason to consider the matter further.
 53. For thorough comments on cognitive deficiency as an epistemic excuse, see Section 14.3.2.
 54. Dressler (2001) p. 141.
 55. Alexander (1990) p. 98.
 56. 4 Wash. App. 908, 484 P.2d 1167 (1971). See Alexander (1990) pp. 96–98.
 57. Supposedly, negligence comes in degrees, with only “criminal negligence” being strong enough to establish criminal liability for an offence, whereas “civil” negligence may establish civil liability in tort cases. I shall not discuss this distinction here. See Saltzburg et al. (1994) pp. 212–213.
 58. *Ibid.* p. 97.
 59. 52 Ala. 308 (1875). References to this case are made on the basis of Saltzburg et al. (1994) p. 214.
 60. Alston (1993) p. 533.
 61. Alston (1989) p. 85.
 62. Goldman (1988) p. 53. It is worth noticing that Goldman alters his conception substantially later in the text: “This means that the weak notion of justification that interests me is not precisely that of blamelessness, but the narrower notion of *mere* blamelessness. That is, the weak notion is that of *ill-formed-but-blameless* belief.” *Ibid.* p. 56.
 63. “I want to focus on the matter of partial reliance on one’s own observations, which for simplicity I will call *partial autonomy*.” Schmitt (1987) p. 62.
 64. *Ibid.* p. 63.
 65. *Ibid.* p. 69.
 66. *Ibid.* p. 69.
 67. *Ibid.* p. 63.

68. Alexander (1990) pp. 102–103. Alexander uses this example as a preliminary to completely “deconstructing” the notion of criminal negligence by arguing roughly that the connection between past blameworthy belief-influencing action and present blameworthy action based on mistaken beliefs cannot be established in actual cases. I shall seek reasons to modify this point below. See in particular Section 13.4.
69. If this does not seem clear, consider that in the very least the belief was formed by a highly unreliable belief-forming mechanism for forming beliefs about children’s health matters: The Williams’s deficient cognitive dispositions. Being thus formed is clearly an epistemically undesirable quality of a belief. See more in Chapter 5 below.
70. See particularly Owens (2000) pp. 82–84.
71. Code (1987) p. 93. Code is at least partially excused for this relaxed attitude towards the notion of belief due to her over-all “responsibilist” approach to epistemology (see *ibid.* p. 36–67). However even if the focus of epistemology is removed from the justification of beliefs to wider epistemic obligations (responsibilities), an account of exactly *what* we do when we “believe responsibly,” may still be needed.
72. Concerning the possibility of such “group beliefs,” see, e.g. Schmitt (1994).
73. Montmarquet (1993) p. 90.
74. Kant’s three degrees of *Fürwahrhalten* “holding-for-true,” being *Meinen* (opining), *Glaube* “belief” and *Wissen* (knowledge) are clearly meant to be mutually exclusive. (*Kritik der reinen Vernunft*. Akademieausgabe 3.532.32–3.522.16 (A822/B850). See Kant (1933) p. 645–652 for translations of the key terms.). Further his notion of *Überzeugung* (conviction) is not normatively neutral but contrasted with the normatively inferior attitude of *Überredung* (persuasion). Seemingly closest to the modern notion of belief is his *doktrinale Glaube* (doctrinal belief), which is a mode of *Fürwahrhalten* not susceptible to pragmatic concerns (Kant claims that belief in the existence of God or in life on other planets fall under this category. *Kr.d.r.V. Akademieausg.* 3.534.29–3.535.31 (A825–826/B853–854). However, as a mode of *Glaube*, doctrinal belief in a proposition excludes knowledge, and hence does not capture the content of the modern concept.
75. Williamson (2000) p. 10.
76. Montmarquet (1993) p. 90.
77. Cf. Dancy & Sosa (eds.) (1992) p. 396, which offers an almost parallel understanding of the notion of propositional knowledge.
78. Dennett (1969) p. 179–185.
79. Incorrect, since it is unclear what the letter “p” stands for in such phrases as “she believes that p.” Certainly it cannot immediately stand for a proposition since the subphrase expressing the proposition would begin with “that.” It would be better to write “she believes **p**,” where **p** is an unbound propositional variable. However, this is far from common usage in the epistemological literature.
80. Sellars (1989) p. 126.
81. Cohen (1989) p. 368.
82. It should be noted, that, e.g. William Alston have tied the term “acceptance” much closer to belief: “I understand ‘accepting’ a proposition as an activity that gives rise to a belief.” Alston (1989) p. 121n. This, emphatically, is not the present understanding. It should be clear from the present paragraph why I find useful a less doxastically committed notion of acceptance. Robert Stalnaker has used “acceptance” as a generic notion, treating belief as one among several “acceptance concepts” Stalnaker (2002) p. 159. Others, like Patrick Maher, have tied “acceptance” narrowly to acts of sincere assertion (see *ibid.*). Whatever the legitimacy of these alternative uses, they do not comply with the present.
83. “If we allow that belief is a dispositional notion, I would think that we would want to grant the same status for acceptance, . . .” Montmarquet (1993) p. 93.
84. Cohen (1989) p. 368.
85. Amusingly, Montmarquet, a stern defender of the fundamentality of the notion of belief, writes: “These are acts of acceptance . . . , which are (or which would be) motivated by a desire to hold true *beliefs*.” (my italics). Montmarquet (1993) p. 90. It is hard to see how such a motivation should work,

- unless the relevant agent comprehend herself as engaged in some complex act of self-suggestion by inducing beliefs through acceptances; a rather bizarre sort of case.
86. Kaplan (1981) p. 138. It may well be the case that Kaplan is here after a conception of belief, rather than a conception of acceptance in Cohen's sense. However, as a conception of belief, Kaplan's analysis is surely inadequate: Many agents may have beliefs without any ability to defend a claim. One may here think of small children.
 87. Sellars (1989) p. 128. Sellars' example of a statement is: "this red and rectangular item is the surface."
 88. Contrary to Cohen's, whose conception of belief, is markedly less useful than his account of acceptance. Cohen holds that "Belief . . . is a disposition to feel it true that p, whether or not one goes along with the proposition as a premiss." (i.e. whether or not one *accepts* that p). Cohen (1989) p. 368.
 89. Cf. van Fraassen (1984) p. 252. Of course, sometimes on their surface, expressions of feelings appear very similar to expressions of beliefs in this regard: For example the phrase "he is a terrible person" may be employed to express hate. Under such circumstances, however, I strongly suspect that the speech-act does not qualify as an emotional expression unless accompanied by some kind of non-verbal expression of the speaker's level of internal tension (by gestures, vocal pitch etc.), which need not accompany a successful expression of belief. I have however, no space to argue this point at length here.
 90. Cohen (1989) p. 369.
 91. Owens (2000) p. 40–41.
 92. *Ibid.*
 93. See *ibid.*
 94. See, e.g. Rorty (1983) for a thorough treatment of the subject. Rorty introduces the notion of akratic belief as follows: "A person believes akratically when he believes that p, being implicitly aware that p conflicts with the preponderance of serious evidence or with a range a principles to which he is committed." *Ibid.* p. 175. I shall leave it open throughout the present study whether the term "epistemic akrasia" is really well chosen as a characterization of such an "incontinent" belief state, as the term "akrasia" may well too strongly suggest a commitment to doxastic voluntarism (see Section 8.1).
 95. Montmarquet (1993) p. 87.
 96. See Code (1987) p. 1.
 97. *Ibid.* p. 75.
 98. *Ibid.* p. 74.
 99. Hardwig (1991) p. 702.
 100. As does Cohen. See Cohen (1989) p. 368.
 101. Cohen (1989) p. 385.
 102. Most notably William Alston in Alston (1993).
 103. See Kenner (1967) for a thorough discussion of various uses of the term "blame."
 104. I strongly suspect that several instances of such distinctions in the literature are better captured by distinguishing between deontic and non-deontic modes of epistemic evaluation. At least that seems to be the case with a distinction of this type tentatively drawn in Owens (2000). Owens, e.g. invokes the case of a [knowingly skilful] mathematician that runs through a fallacious proof, which, however, to her seems perfectly valid. She thus comes to form the belief that the conclusion of the proof is true. Now, unknown to the mathematician, she is under the influence of a mind-warping drug. In this case, Owens states: "The subject has an unjustified belief, but it does not follow that she is unjustified in having those beliefs." *Ibid.* p. 141. It would seem that Owens wants to say something like: The subject is blameless in holding the belief, but in fact her holding of that belief falls short of certain ideal epistemic standards. If this is indeed the case, I believe the latter phrasing would have been more fortunate.
 105. An possible exception to this terminological point is the predicate "autonomous." There is indeed a perfectly legitimate sense in which, e.g. "knowledge" in some sense of the word may be termed "autonomous" without thereby evaluating any *knowers*. For example, W.W. Bartley, III states: "Objective knowledge is autonomous in just the sense that Marx had in mind: It outgrows our

- control, crosses our expectations, and nullifies our calculations.” Bartley (1987) p. 435. Bartley offers as an example of such autonomy the unpredictable consequences following the “creation” of the sequence of real numbers. However, the sense of “knowledge” invoked here is hardly that of a doxastic state. Thus this “Marxian” sense of autonomy does not immediately apply to beliefs.
106. Steup (1988) p. 67.
 107. At least Steup gives as his only example of an obligation in the relevant context, an obligation that an agent can be *blamed* for not fulfilling: “If I have given you a promise, I have the obligation, or duty, to keep it, and if I don’t keep it, then I do something I can be *blamed* for.” Steup (1988) p. 67 (my italics).
 108. *Ibid.*
 109. Sellars distinguishes norms prescribing what someone *ought to do* “rules of actions” from rules pertaining to how something or someone *ought to be* “rules of criticism.” Sellars (1969) p. 508. It seems plausible that Sellars is here after a distinction between deontic and non-deontic norms in my preferred terminology, even if distinguishing “ought-to-do”s from “ought-to-be”s does not quite do the job; e.g. a human agent may *ought to be* blameless: This import is conveyed by Sellars’ emphasis that only rules of action demand the subject of the rules to possess appropriate “recognitional capacities” (*ibid.* p. 115. Scarequotes. Sellars in fact submits that the term is “a menace to sound philosophy”). It seems plausible that Sellars imposes this demand in order to make sure that the agents, whose conduct fall under rules of action, might in fact be held *blameworthy* for failing to follow such rules. This supposition is further supported by the observation that he terms subjects of rules of criticism “*subject-matters*,” whereas subjects of rules of action are termed “agent-subjects.”
 110. Steup (1988) p. 67: “I shall take the concepts of *obligation* and *permission* as the primary deontic terms.” Alston (1989) p. 115: “We may think of *requirement*, *prohibition*, and *permission* as the basic deontological terms”
 111. *Ibid.* p. 116.
 112. *Ibid.*
 113. *Ibid.* p. 115.
 114. *Ibid.* p. 119.
 115. *Ibid.* p. 137. Alston (1993) p. 533.
 116. Alston (1993) p. 530.
 117. For example Code (1987). Code seemingly takes “blamelessness” to be the central sense of “responsibility,” a notion that she never explicitly defines. However she gives decisive cues that responsibility is for her an evaluative concept by contrasting it with irresponsibility as well as important hints that responsibility is in her sense a deontic notion by explicitly stating that, e.g. a computer cannot be “responsible.” *Ibid.* p. 51.
 118. “A responsibility to know is at issue.” Code (1987) p. 2.
 119. “I want to claim that individuals are in many instances *responsible* for being epistemically virtuous, at least to some reasonable degree.” Montmarquet (1993) p. 28 (italics in original). Arguably, though, in most contexts Montmarquet uses the term “responsible” in the way suggested in the present study.
 120. Scanlon (1998) p. 248. Scanlon also persistently uses “responsibility” to mean “duty” or “obligation.”
 121. *Ibid.*
 122. See Oshana (1997) pp. 76–77.
 123. Fischer & Ravizza (1993a).
 124. Stating exactly what determinism amounts to is no easy matter. Fischer & Ravizza follow Peter van Inwagen in suggesting that an event *p* is determined in the relevant sense if, and only if, no one has, or ever has had any choice whether *p*. *Ibid.* p. 9. As we shall soon see below (Chapter 8), this definition renders most belief-contents determined, since no one literally *chooses* them. This is hardly the conclusion we want from a definition of determinism, since often these contents are in fact affected by human activity.
 125. Fischer & Ravizza (1993a) p. 32.
 126. *Ibid.* p. 11. According to Fischer and Ravizza (*ibid.*), such a view “justifies our responsibility-reactions by appeal to the useful consequences that follows from these reactions.”

127. Leaving aside esoteric concerns about backwards causation, which hardly applies on the macro-level relevant to the deontic evaluation of belief.
128. This “defiant” stance towards determinism has been taken by J.J.C. Smart, who has argued that if “pure chance” were to play a role in the history of the world, threats and punishment could not be wholly effective measures. See Fischer & Ravizza (1993a) p. 12 n22. In fact though, this line of reasoning relies on the highly implausible claim that unless anything is entirely determined, unpredictable “chance” rules supreme: It suffices to notice that even if human action is not completely determined, it is hardly always unpredictable either.
129. At least one author, namely J.J.C. Smart, has presented a version of the social-regulation view that could be taken to suggest a substantial analysis of the notion of blameworthiness: “[T]he ascription of responsibility and the non-ascription of responsibility, have therefore a clear pragmatic justification which is quite consistent with a whole-hearted belief in determinism.” See *ibid* p. 12. According to Fischer & Ravizza, Smart’s “social regulation view” is to count as a substantial analysis of the concept of responsibility. “A third and perhaps most significant appeal of the social-regulation view is that it arguably meets the conditions for an adequate theory of responsibility” (*ibid* p. 13). Cf. Section 3.1 it is not altogether clear what is meant by “responsibility” in the present context, but it could be taken to imply blameworthiness. Also Peter Strawson has it that the prospectivist optimist “undertakes to show that the truth of determinism would not shake the foundations of the concept of moral responsibility.” Strawson (1993) p. 62. Perhaps it would be most charitable to read Smart in a weaker way. However, if I am really fighting a straw man, at least it seems like a popular one!.
130. See Fischer & Ravizza (1993a) p. 7.
131. The term “structuralism” is used by Fischer & Ravizza (1993a) p. 32 to denote a-temporalist theories. However, this term is dangerous: there is no reason why structures could not be temporal in character.
132. *Ibid*.
133. Zimmermann (1988) p. 38.
134. *Ibid* p. 16.
135. It should be noted that I here employ the notion “Strawsonian” to denote the view held by Peter Strawson according to Fischer & Ravizza. They, however use “Strawsonian” to denote a view held by Fischer, who tries to improve on the (blatantly implausible) view that they attribute to Strawson (see Section 3.3.3). I believe that my choice of terminology creates the least confusion.
136. *Ibid* p. 15.
137. *Ibid* p. 16.
138. *Ibid* p. 17.
139. Strawson (1993).
140. *Ibid* p. 14.
141. *Ibid* p. 55.
142. Fischer & Ravizza (1993a) p. 16.
143. Strawson (1993) p. 55.
144. *Ibid* p. 58.
145. Fischer & Ravizza (1993a) p. 17.
146. See *ibid* p. 18. Fischer terms such a theory “Strawsonian,” because he takes it to preserve the gist of Strawson’s supposed analysis of moral responsibility. It should be clear by now, why this terminology is highly unfortunate. Strawson never suggested anything of the sort.
147. *Ibid* p. 17.
148. It should be noticed from the outset that I take some of the properties making for the epistemic undesirability of an agent’s belief to be *relational* properties of that belief, most generally the relational property that the belief is held by the relevant agent in a certain way. For this reason one might think it more natural that I made it explicit in the definition that I take the primary object of deontic evaluation to be, not the state of affairs that the agent believes a proposition with some particular content, but rather the agent’s particular belief-state with its particular properties, including relational ones. However, I believe this to be an unnecessary over-loading of the definition, since I take it to be clear from the phrasing that what is evaluated (blamed) is an agent’s particular belief-state.

149. Zimmermann (1997) p. 229.
150. *Ibid.* p. 231.
151. "I can claim that accuses are possible, as long as it is accepted that acting freely in the belief that one is doing objective moral wrong is sufficient for being morally blameworthy for one's behaviour (but not sufficient for actually doing objective moral wrong). This, I believe, is precisely what should be accepted." *Ibid.* p. 234. I shall leave aside Zimmermann's distinction between subjective moral wrong (the wrong an agent is aware of) and objective moral wrong, as the objective concept seems to me by far the more natural. I shall not try to disambiguate Zimmermann's curious notion of "acting in a belief," which could mean both "acting on a belief" and "acting while having the belief."
152. *Ibid.* p. 235.
153. *Ibid.*
154. Russell (2001) p. 35.
155. At least this is how the example is constructed. Of course many details would need to be filled in order to secure that no relevant excusing conditions obtain. Here I shall simply go with Russell's conclusion.
156. The exact sense of objectivity here is a much-disputed matter that plays a central role in, e.g. the untiring debates between internalists and externalists concerning epistemic justification. This is an issue I shall hope to avoid in the present study.
157. "She will be blameless just in case she has no adequate reason to believe that what she did was objectively wrong or has a legitimate excuse for doing it. Again it does not matter whether it was in fact objectively wrong, as the example of the physician shows..." *Ibid.*
158. According to Alston, this position is even held by "most epistemologists who have attempted to explicate justification." Alston (1989) p. 85. See *ibid.* n5 p. 85 for a short list of confessing "deontologists." Goldman (1988) introduces a sophistication in that he accepts a deontic notion of justification, but contrast such "weak" deontic justification with "strong" justification based on reliability. See Section 1.5.
159. Alston (1978) p. 280: see, let us use the term 'epistemization' for whatever transforms a true belief that *p* into knowledge that *p*." In the general understanding, justification lost out on this opportunity when confronted by the famous "Gettier-examples": cases in which an agent holds a justified true belief short of qualifying as knowledge.
160. At least this seems to be the import of Steup (2001) pp. 134–141.
161. Alston (1993) pp. 528–530.
162. Alston thus accommodates for a "reliable indicator" as well as a "reliable process" version of reliabilism (cf. Dancy & Sosa (eds) (1992) p. 433. See also Section 5.4.
163. Actually, if one presumes truth to be *the* defining epistemic goal (as I do in the present study), I suspect that vindicating the narrowly *epistemic* desirability of the last three desiderata on Alston's list is not easy at all. Of course reflecting on the epistemic status of one's belief may be "eminently worthwhile" and even "a noble calling" (Alston (1993) p. 531). Also constructing coherent belief systems may actually be what makes for the "satisfaction of the intellectual side of our nature" (*Ibid.*). What is perhaps not so clear is how these noble pursuits further the truth-goal, even if intuitively they typically do. For example aiming for a very coherent or reflected belief-system may arguably sometimes stand in the way of satisfying the desideratum of truth-conducivity.
164. Alston (1989) p. 83.
165. *Ibid.* (my italics).
166. *Ibid.*
167. For example Linda Zagzebski submits that "most contemporary epistemology" is committed to the view that "the person who is greatest in knowledge is the one who has amassed in his mind the highest number of true propositions that pass whatever test for warrant the theory has proposed." Zagzebski (1996) p. 45. In a similar vein, James Montmarquet has submitted that broader epistemic goals "cannot be defined simply in terms of learning many truths (or many truths and few falsehoods)" (Montmarquet (1993) p. 33) and seemingly holds this to be the main-stream view concerning the source of epistemic value. I find it incredibly how hypostatizing the truth-goal should commit one to

such bizarre value judgements. Rather, I take the main-stream position to be that the question of how many truths and falsehoods an agent should believe is entirely outside the province of epistemology proper, which is only concerned with the epistemic quality of *particular* actual or possible beliefs or perhaps the truth:falsity ratio in an agent's mass of beliefs; never their sheer quantity.

168. Jones (1997) p. 424.
169. See in particular David (2001) and DePaul (2001).
170. Oshana (1997) p. 77. This does not exclude, of course, that overriding reasons may render any form of punishment illegitimate. For blameworthy beliefs, the further complication arises that the decisive belief-influencing actions may be long past and hence not correctible in any straightforward sense. However, there is a relatively clear sense in which epistemically blameworthy agents may be "punished" for their doxastic transgressions by relegation or diminished trust in their fitness for promotion within the cognitive network of their epistemic community.
171. Of course this claim is controversial. However, many philosophers who have argued for the social function of knowledge, have done so by endorsing some version of instrumentalism: The factor "epistemizing" true belief into knowledge has been held to be reliable formation or some other property, whose desirability is derivable from the truth-goal. An interesting exception is Edward Craig, who has suggested that social function is in fact at the core of the concept of knowledge. To him knowledge is simply the "property X" of an agent who fills the social role of a "good informant" in addition to having true belief (see Jones (1997) p. 432). To me this solution seems more like a way of defining oneself out of trouble, since it is hard to see why we should take the "property X" to resemble anything normally understood as knowledge. A more promising anti-instrumentalist line of approach to the value of knowledge is suggested by Ward E. Jones, who argues that only a knower may *understand* why his true beliefs are true (although often even knowledge does not suffice for understanding of this sort). *Ibid.* However, Jones concedes that understanding, as a contingent property of knowledge, is only valuable from the first-person perspective (*ibid.*) and he does not try to render individual understanding valuable from a social perspective. I personally doubt whether this could be satisfactorily done, although the issue is obviously complex and deserves further treatment.
172. I shall use this abbreviation, well aware that some may insist that belief-formations etc. may also be adequately explained by explanations of an intrinsically non-causal kind. Although I disagree to this, I shall not engage with this position here.
173. Whereas having good reasons for holding a belief is generally considered a *necessary* condition for being rational in holding it, some authors have opted for stronger *sufficiency* requirements. For example it may be claimed that a belief cannot be rational, if the agent has been in a certain sense *indolent* with respect to the reasons she has for holding it, i.e. roughly: if it may be legitimately claimed that her evidence, even though adequate, was of such a nature that she should have checked it better. Foley & Fumerton (1982) (correctly, I think) rejects this suggestion. Such a thesis simply forces a deontic role on the notion of rationality or reasonableness that it is entirely unfit to play. However, the same authors argue that an agent must be *rational* in believing she has good reasons to hold a belief, in order to be rational in holding that belief. *Ibid.* p. 38. In my opinion, this higher-order requirement cannot but lead to a vicious regress. See Schmitt (1983) for further critique.
174. This mismatch between first and second-order doxastic attitudes is close to what is generally known as epistemic akrasia or doxastic incontinence. However most accounts of these notions presuppose some kind of doxastic voluntarism, a position I shall reject (see Chapter 8). I shall therefore not employ these familiar terms.
175. This is acknowledged by Alston (1993) p. 539 n16.
176. See, e.g. Achinstein (1983) for a thorough discussion.
177. Achinstein (1983) p. 146ff. Discussed in Graham (1997) pp. 227–228. My paraphrase is consciously rough, but the subtleties left out are not immediately relevant.
178. Graham (1997) p. 228.
179. Russell (2001) p. 38. Russell holds that a belief is "objectively justified" if, and only if, the believer has good reasons for holding it (see *ibid.* p. 39).

180. Robert Audi may be hinting at this point, when in the run of his argument against doxastic voluntarism he argues: “A reason for action is always expressible in a phrase of the form ‘in order to A’ where ‘A’ ranges over action-types, though a reason for action may also be expressed in other ways. A reason for belief is never so expressible.” Audi (2001) p. 99. Now, at best Audi points to the unnaturalness of talking about instrumental reasons for believing. However, simply assuming that such reasons do not exist, severely begs the question against doxastic voluntarism.
181. In particular a confusion between instrumental and basing reasons maybe accounts for the claim defended in various versions by, e.g. Winters (1979) and Scott-Kakures (1993) that it is impossible to hold a belief while aware that one adopted it voluntarily and still hold it “for that reason” (Winters (1979) p. 253): If one has somehow succeeded in voluntarily adopting a belief, certainly one does not have a *basing* reason for holding it in the present sense, as the occurrence of this belief is now best explained in terms of a belief-forming intention rather than in terms of awareness of evidence. However, the *instrumental* reason one has for holding the belief is not concerned in this: One does not after the belief-acquisition *base* one’s belief on reasons, which one does not take to support its truth, even if it was adopted for non-truth-oriented *instrumental* reasons. It is hard to see why one’s awareness of the latter fact should threaten the stability of the acquired belief. See Section 8.2.1.3 for a fuller treatment of these matters.
182. Alston (1993) p. 528.
183. *Ibid.* p. 539.
184. The notion of rationality or reasonableness is notoriously elusive. I only claim to have captured here one important sense, in which a belief might be reasonably held. Another obvious sense in which a belief may be “rationally” or “reasonably” held, is if the agent has good *instrumental* reasons for holding it. In that sense it may, e.g. be “reasonable” or “rational” for a person living in a society with severe persecutions of atheists to believe that God exists.
185. See Toribio (2002), especially p. 50.
186. *Ibid.* p. 40.
187. *Ibid.* p. 39.
188. See, e.g. Wilson et al. (1995).
189. See Dancy & Sosa (eds.) (1992) p. 433.
190. Goldman (1988) p. 53. Goldman here holds that, for a belief to be *fully* or *strongly* justified, it must be formed by adequate (reliable) processes and adequate (reliable) methods, insofar as methods were used in its formation (*ibid.*).
191. *Ibid.* p. 55.
192. *Ibid.* p. 60.
193. *Ibid.*
194. Dancy & Sosa (eds.) (1992) p. 432. Goldman (1988) p. 54.
195. Goldman (1988) p. 61.
196. *Ibid.* p. 62.
197. See Goldman (1993), especially p. 103.
198. Jones (1997) attributes this defence of value-monism to David Armstrong and Colin McGinn respectively and (correctly, I think) rules it inadequate. *Ibid.* p. 428.
199. The “or” here is of course a standard-logical inclusive “or.” I take it that non-reasonableness may unproblematically be equated with unreasonableness.
200. This claim also seems to fit with our everyday practices: Certainly an average person would protest at being excused for an undesirable incidence, in which she did nothing wrong at all. Consider, e.g. a person who had made an explicit and ratified arrangement not to meet at her job on some particular date. However her boss forgets about this arrangement. As a result her company faces some serious problems. Now, this person would hardly like to hear that she is appropriately *excused* for these problems: She would rather feel that she *needs* no excuse at all, as she did nothing wrong by staying home that day. (I owe this example to Marcia Baron in private conversation). Consider also the more extreme case, where an agent is told that the weather is going to bad tomorrow, but need not worry, since, really, she is appropriately excused for this. If not simply dismissing this comment as a

joke, such an agent would rightly feel that she does not owe an excuse for the future weather, since tomorrow's weather is not something she could be required to control in any sense.

201. Zagzebski (1996) p. 49.
202. Ibid. Perhaps Zagzebski is justified in taking this notion as pre-understood, since according to her "In post-Aristotelian ancient philosophy and throughout most of the modern age, certainty has been given more attention than understanding, whereas in Plato's and Aristotle's concept of *episteme*, in the philosophies they influenced in the Middle Ages, and in such modern philosophers as Spinoza and Locke, it was the reverse." Ibid. p. 46. I shall not quarrel with this sweeping historical claim.
203. Ibid. p. 50.
204. Ibid. p. 44.
205. Zagzebski (1996) p. 44.
206. Russell (2001) p. 39.
207. Ibid. p. 35. (my italics).
208. I trust my reader to do the logical footwork for him or herself. Only the resources of first-order propositional logic are called upon. The clumsy appearance of the result is my responsibility.
209. Ibid. p. 39.
210. Ibid. p. 40. I believe this passage strongly suggests the reading of Russell I have adopted: He seems to at least suggest here that the blameworthiness of agents must be explained *within* the requirements he has imposed, thus embarrassing the interpretation that he only opts for sufficient, but unnecessary conditions of epistemic blameworthiness.
211. If this sounds too odd, remember that we are talking metaphorically here. A more exact formulation, employing the notions introduced above would be: His awareness of certain prejudiced testimony is no longer the best causal explanation of the fact that he holds the belief.
212. Note, that this demand is considerably stronger than that imposed for *doxastic incontinence* by John Heil, who has argued that "One is at fault if one harbours a belief, i.e. inconsistent with all that one takes to be relevant to the warrant of that belief. . . (. . .) . . . The incontinent believer is not simply making a mistake. He believes what he *recognizes* to clash with what he holds true." Heil (1984) p. 67. Here it is not demanded that the second-order reasons are good ones. Further, even though Heil certainly considers doxastic incontinence, so conceived, an epistemic vice (ibid. p. 70), it would be uncharitable to read the above as a full-blown account of epistemic blameworthiness.
213. In Owens (2000) and Alston (1993) respectively. Alston's stand on the issue is rather complicated. In Alston (1993) he declares a deontological conception of epistemic justification "untenable" (ibid. p. 533). However, he presupposes here that proponents of this conception must require beliefs to be under "effective voluntary control" (ibid.). In earlier papers (Alston (1989) Chapters 4 and 5) Alston has developed a concept of deontic justification based on indirect control (see Section 10.1), but concludes after considering some practical examples that this concept "does not give us, what we expect of epistemic justification" (ibid. p. 95) and even that "there is nothing to be said for the deontological conception as a fundamental concept for epistemology" (ibid. p. 152). However, Alston also maintains that deontological justification remains an "important and interesting concept" (ibid.), even in contexts of training or education. Neither does he deny the relevance of deontic terms such as "epistemic obligation." In Alston (1993) he abandons the search for a fundamental concept of justification in favour of a pluralist approach. One should perhaps expect that, given his previous acknowledgement of its pragmatic importance, he would now find room for at least one deontic conception of justification alongside a fan of evaluative concepts, but somewhat surprisingly he takes refuge in a tightened conception of the kind of doxastic control required for epistemic blameworthiness in order to keep deontic concepts out in the cold.
214. Owens (2000) p. 121. The "non-moral things" are clearly meant to include beliefs. The issues of the normative status of beliefs and their control occupy the bulk of Owens' book.
215. Hume (1975) p. 322. Quoted Owens (2000) p. 121.
216. Ibid. p. 118: "Where it [my road rage] expresses a psychological state itself subject to rational assessment, I am culpable."
217. Ibid. p. 118.

218. Feldman (2001) p. 88.
219. Ibid.
220. I denied the basic status of terms like “permission” and “prohibition” in the context of epistemic deontology in Section 3.1 above.
221. Ibid. p. 89.
222. Alston (1989) p. 143.
223. The formulation is kept very loose at this early stage of the discussion. In particular, as shall emerge, one cannot say that the blameworthiness of a belief requires that the agent could have prevented its formation.
224. That is (in other words), the instantiation of the properties (relational and other) making for the epistemic undesirability of her belief that p.
225. It should be noticed that the fact that a notion of inexcusability features prominently in this clause does not make the analysis flatly circular, as the notion of inexcusability featuring in (2) now applies to exercises of doxastic control rather than to holdings of beliefs. However, there is a legitimate regress worry here, since one may try to excuse an agent for her failure to serve the truth-goal in exercising her doxastic control by referring to her holding of certain background beliefs, which might also be undesirable from an epistemic perspective. If it is argued now that this excuse is inappropriate, since the agent lacks an appropriate epistemic excuse for holding these undesirable background beliefs, a regress gets going. In the proper later context (Section 14.4) I shall argue that this regress is not vicious to the objectives of the present study.
226. My reasons for not coining this condition in counterfactual terms shall be made clear in Section 10.4. If one eschews the notion of a causally effective omission, it should be quite easy to skip the parts on my discussion turning on this notion. However, I take it that events caused by omissions are quite commonplace, e.g. it seems entirely natural to say that my plants withered, (in part) *because* I failed to water them.
227. “As a minimum,” because \mathcal{M}^* may of course gratuitously be supplied with modes of doxastic control not relevant to any possible ascriptions of epistemic blameworthiness.
228. See, e.g. Stocker (1982) p. 399.
229. For a discussion of dimensions of epistemic desirability, see Chapter 5 above.
230. Ryan (2003) p. 66.
231. A difficult point may be: If the music-listener’s staying in the room is somehow (morally) desirable, is she then *praiseworthy* for staying in the room, even if she could not have left it? Here I find it hard to praise her for staying in the room. Rather, if she did nothing to leave the room (banged on the door, etc.) she might be praised for this or we may simply praise her for *intending* not to leave the room. This and similar complications account for the fact that I shall have little to say about the notion of praiseworthiness in the present study.
232. Feldman (2001) p. 81.
233. Ibid. pp. 82–83.
234. Ibid. p. 83.
235. Feldman actually sets up such a scenario: If a person gets tortured unless I believe in 30 sec that the lights are turned on, I had better exercise my Feldman-control over this belief for moral reasons. Ibid. p. 82.
236. That is: the properties making for the belief’s epistemic desirability or undesirability in the present sense.
237. Most likely, this is a truism. See below.
238. Audi (2001) p. 95.
239. Davidson (1980) p. 50. The passage continues: “. . . , and this in turn requires, I think, that what the agent does is known to him under some description.” I consider this further claim more like a special version of the activity-guiding requirement needed for deontological adequacy than a metaphysical necessity: Surely under normal circumstances an agent can hardly be responsible for something that he does not know he is doing under any description, but I think it is less clear that it is therefore not an action at all. Consider, e.g. the exotic case, where the agent has knowingly administered to himself

a drug that prevents him from knowing what he is doing. Now, his doings while under the influence of the drug seems to be something he can be held responsible for, and perhaps they even are actions. Another view could have it that these doings are really blameworthy consequences of the drug-taking action and not actions themselves. Personally I find it hard to decide between these options.

240. I take the Davidsonian stance that *all* action is really “primitive” physical (or mental) action under some suitable description.
241. Perhaps most importantly, I ignore the question about whether, e.g. intentionally giving one’s money to a religious sect due to severe brainwashing, is in fact an action at all, and if so, an action for which one is responsible. In other words, I ignore the issue of *deep* voluntary control, i.e. action from freely formed intentions (see, e.g. Steup (2000) pp. 31–32).
242. *Ibid.* p. 51
243. *Ibid.* p. 54.
244. A very similar point was already made by Heil (1983) p. 359 before the current surge in doxastic voluntarism: “Second, one convinced that the adoption of a belief is a non-basic act is obliged to produce some accounts of the basic acts thought to generate (in whatever way) the act of belief-adoption. It is not clear what these actions *could* be.” I am perfectly sympathetic to this point. Only I follow Davidson in denying that there is any interesting difference between basic (or primitive) and non-basic (or non-primitive) actions at all. Actions may be mere bodily movements or complicated technical achievements relative to different descriptions equally legitimate in different contexts. In a provocative (and somewhat dangerous) phrasing, basic actions are all the actions that are! See Davidson (1980) pp. 59–60. For an provocatively eloquent re-formulation of Heil’s point see Alston (1989) p. 129: “What button would I push?” etc.
245. See Heil (1983). I discuss Heil’s contribution to the debate over the importance of such exercises of direct doxastic control to evaluations of epistemic blameworthiness in Section 9.2.
246. The deontological importance of practical property control only becomes visible in bizarre Frankfurt-cases like the “locked room” example invoked by Sharon Ryan (see Section 7.2). Here the agent cannot control the “content” of her action, namely that of remaining in the room. However, she may well control certain properties of her staying that are important to the deontic status of her remaining in the room, insofar as her so remaining had undesirable consequences. Imagine, e.g. that just outside the room a man died, while the agent was in the room. Here, although the locked-up agent could not possibly have helped the man, our attitude toward her may still be affected by further circumstances. For example if she knowingly ignored the man’s lament, not realizing that the room was locked, this naturally elicits a different response from the case where she realized that the room was locked and therefore gave up any attempt at rescuing him. It should be noted that such cases are not immediately analogous to typical instances of believing: here we may in fact sometimes be able to influence the content of our beliefs, although, as I shall argue at length below, this control is rarely deontologically adequate.
247. *Ibid.* pp. 119–142 ff.
248. *Ibid.* p. 119.
249. *Ibid.* p. 137.
250. *Ibid.* p. 128.
251. *Ibid.* p. 128.
252. See. *ibid.* note 21 for an extensive list. Among the texts figuring on Alston’s list, Steup (1988) is particularly noticeable. The crucial quotation must be the following: “If I cannot directly choose a belief, I have in order to choose a belief, to do something else than the choosing” *Ibid.* p. 71. However, Steup’s position on Alston’s list seems entirely unwarranted, at least if based on quotations like the above: Steup’s prime examples of “choices” are not “basic” choices but rather complex ones like choosing right eating habits! See *ibid.*
253. Alston (1989) p. 134.
254. *Ibid.*
255. *Ibid.* pp. 137–138.
256. *Ibid.* p. 143.

257. Ibid. p. 137.
258. Due to the notion of basing evidence (or reason) employed in the present study (see Section 5.3), “seems” should not be given an overly psychologistic reading. Rather a piece of evidence “seems sufficient” to a cognitive system, if the system reacts to exposure to that evidence by forming a belief, insofar as certain over-riding causal mechanisms are not active (like, e.g. an emotional pressure).
259. Logik. Akademieausgabe 9.73.27–32.
260. Exactly what doxastic voluntarism amounts to has rarely been spelt out in detail. However, at least one author, Robert Audi, has presented an understanding not explicitly wedded to epistemic deontology. See Audi (2001) p. 93. However, Audi quickly goes on to draw the connection to “the ethics of belief,” and thus follows the general trend of more or less implicitly taking the importance of the issue of direct content control to be its relevance to epistemic deontology.
261. Nevertheless, some doxastic voluntarists of a compatibilist inclination verge dangerously on the brink of a deontologically impotent conception of doxastic freedom. See my comments on Ryan (2003) in Section 8.3.1.1.
262. This is van Fraassen’s preferred term. Van Fraassen (1984) p. 250.
263. Ibid. p. 254.
264. Ibid. p. 255. Of course this interpretation rests upon not taking the phrase “entered a contract with myself” too literally. I suspect this is the most charitable reading, since van Fraassen hardly needs a more literal reading to make the point he needs for his argument.
265. Van Fraassen points to “cognitive commitment” involved in belief in order to solve a problem facing classical Bayesian accounts of rationality, according to which an agent should be willing to accept as fair any bet on the truth of a proposition with odds exceeding her degree of trust in it. The problem is that this commitment may force the agent to accept a so-called “Dutch Strategy”; a set of bets, which individually seem fair, but taken together inevitably results in a loss. Van Fraassen demonstrates that the “Dutch Strategy” may be set up, once the agent’s beliefs do not adhere to a certain “Reflection” or “Self-confidence” principle (Ibid. p. 244). Roughly, according to this principle the agent must have full confidence in her future “credence function,” i.e. if the agent believes that she will at a future time hold a proposition 60% likely to be true, she must now believe that the proposition is 60% likely to be true. Certainly self-confidence to such a degree runs counter to basic fallibilist intuitions and can hardly be demanded for rationality. Now, Van Fraassen counter-balances this demand for rationality with the “committed” nature of beliefs: Given my commitment to fallibilism I cannot simply be required to give up my fallibilism, if confronted by a bookmaker using a “Dutch Strategy.” Thus, even if I fall victim to a Dutch Strategy, I may remain rational in a natural sense. I wholly endorse van Fraassen’s point. However, the irrelevance of such “voluntarist” commitments to the present discussion should be obvious. For a very different opinion see Wansing (2004) p. 420, where Van Fraassen is represented as an arch-voluntarist.
266. “On a strong interpretation it [doxastic voluntarism] expresses the view that believing itself is an *action-type* having some tokens that are directly voluntary. On a weaker interpretation it expresses the view that *forming* a belief is sometimes such an action-type. Call the first view the *behavioural version* of doxastic voluntarism and the second the *genetic version*.” Audi (2001) p. 94. It should be remarked that strictly speaking Audi’s “genetic version” are narrower than mine, since it on only concerns belief-*formations*. However, the extension to cover also the prevention, suspension and sustention of beliefs seems natural. Heinrich Wansing (Wansing (2004)) has pointed to an almost endless number of further distinctions one may draw within genetic doxastic voluntarism (and could draw within behavioural voluntarism as well): One may make clear whether one’s claims concern actual or possible agents, actual or possible beliefs, at least one agent or all agents, at least one belief or all beliefs, conscious or unconscious beliefs, conscious or unconscious belief-formation, etc. In the present context it seems irrelevant to enforce all of these distinctions at all times. Where nothing else is mentioned, I am talking about the actual conscious beliefs of actual agents (past, present, future) and take behavioural voluntarism in its basic version to claim that at least one such belief is/was consciously formed at will by an actual agent.
267. Steup (1988) p. 76, (my italics).

268. *Ibid.* p. 72.
269. See Steup (2000) p. 32. According to Steup, such belief-formation as well as arm-raising are subject to “categorical voluntary control.” The very formulation of this claim poses significant problems. For this reason I shall postpone its treatment to the later discussion of “the general analogy argument,” in which it shall figure prominently.
270. James (1981) p. 424.
271. *Ibid.* p. 1167.
272. Gale (1999) p. 71.
273. James (1981) p. 401.
274. *Ibid.* p. 425.
275. I shall leave James for here, and refer to the very informative Gale (1999) for further analysis.
276. Some may be inclined to argue against behavioural doxastic voluntarism by holding that beliefs are states, for which reason they cannot be events. A fortiori, thus, they cannot be actions, since actions are events. See in particular O’Shaughnessy (1980) p. 26 for an argument along these lines. However, it is far from clear that a general distinction between states and events obtain. It is not enough to argue that states are somehow unchanging, whereas events involve changes, for if events and states are conceived of as particulars any state will also change in some respect or other and vice versa.
277. Surprisingly, no actual doxastic voluntarists in the literature have openly acknowledged the need for a hypothesis of this type.
278. Steup (2000) p. 32.
279. *Ibid.* p. 25.
280. Steup never makes clear by examples or otherwise, what exactly *disbelieving* a proposition amounts to. He seems however, in the above-quoted passage to endorse at least the conceptual possibility that an agent can believe *p* and then, as it were, “seamlessly” come to disbelieve it by decision.
281. In which case, given $\text{Blame}_{\text{epist, control}}$, she would of course also have *prevented* that she holds the belief that *p* in an epistemically undesirable way.
282. In which case, given $\text{Blame}_{\text{epist, control}}$, she would of course also have *suspended* the occurrence that she holds the belief that *p* in an epistemically undesirable way.
283. A remarkable exception to this tendency is Code (1987) who starts out in the second mode, blame attributed for *not* knowing in which side of the road to drive. *Ibid.* pp. 1–3. However in this particular case, the psychological possibility of preventing a relevant belief by believing the contradictory proposition seems open: Basic traffic rules are not something about which a normal agent is likely to harbour contradictory beliefs.
284. Matthias Steup, perhaps the most persistent recent advocate of doxastic voluntarism, is an exception. In Steup (1988) he discusses the case of an agent who arguably should have prevented “withheld” the formation of an undesirable belief (*ibid.* pp. 79–80) and in Steup (2000), the notion of suspension of belief figures prominently.
285. Ryan (2003) p. 68. See also Pojman (1985) p. 38. At least for Augustine, the attribution of doxastic voluntarism seems entirely warranted and widely agreed upon (see, e.g. Van Fraassen (1984) p. 250). However, I will not enter into an extensive exegetical discussion in the present context, where I shall deal with doxastic voluntarism on a more general level.
286. Scott-Kakures (1993) p. 77.
287. O’Shaughnessy (1980) p. 21.
288. Scott-Kakures (1993) p. 77.
289. Williams (1973).
290. Pojman (1985) states the logical impossibility of a rational person acquiring a belief by will in full consciousness and irrespective of truth considerations. He remains agnostic or even hostile concerning more global versions of CIC. *Ibid.* p. 52. Pojman’s central argument turns on declaring fully conscious beliefs that {*p* and I believe *p* for other than truth considerations} “logically odd” (*ibid.* p. 48). However, as nicely pointed out in Wansing (2004), the alleged “logical oddity” of such beliefs seems to matter very little to CIC concerning the belief that *p*, since they are strictly speaking about one’s belief that *p* rather than its wilful acquisition.

291. O'Shaughnessy (1980).
292. Scott-Kakures (1993).
293. Barbara Winters is a more difficult case. It is far from clear, which kind of necessity is suggested by her continued use of the modal operator "Necessarily..." in Winters (1979). There she discusses her thesis that "Necessarily, $\neg(\exists x)(\exists p): (x \text{ believes in full consciousness } [x \text{ believes } p \ \& \ x\text{'s belief of } p \text{ is not sustained by any truth consideration}])$ " on an equal footing with Bernard Williams' version of CIC (which she rejects), but in defending this "general claim," she explicitly relies upon empirical premises.
294. Scott-Kakures (1993) pp. 77–78. It is not clear from this quotation in isolation that Scott-Kakures is talking about conceptual impossibility. However, as should be clear from the quotations offered above, this is made very clear earlier in the article.
295. A paradigm case of "causal luck" would be the case in which an agent with a paralysed arm might still succeed in "raising" it through her intention to do so, if, e.g. the emotional stress of her incapacity causes a jism to thrust upwards the paralysed limb. I take it that the term "unmediated" in Scott-Kakures formulation of the conceptual impossibility claim is meant to exclude the doxastic parallel of such cases. At least he later makes it clear that he takes actions like arm-raising as paradigmatic of the conception of belief-formation, which he wants to rule out. Scott-Kakures (1993) p. 91.
296. At this point, Winters presents a problem, since she explicitly argues for the impossibility of "sustaining" a belief under certain circumstances (see note above). However, she ventures that "sustained at will" applies to a belief originally acquired at will and still held for that reason." Winters (1979) p. 253.
297. O'Shaughnessy (1980) pp. 26–27.
298. Bennett (1990) p. 104.
299. Of course the possibility obtains that Bennett thinks of the "bio-feedback" method as only one way among several in which a belief could (hypothetically) be acquired at will. However, there is no trace in the text that Bennett should not hold "bio-feedback virtuosity" necessary in order to achieve this feat.
300. O'Shaughnessy (1980) p. 27.
301. O'Shaughnessy introduces the term "b-believing" in the following passage: "Now I shall begin this discussion by making the (almost certainly counter-factual) assumption that there might exist such an act as the immediate willing of belief, (modelled upon the familiar basic act that is the immediate willing of arm rise). I shall call it 'b-believing'...." Ibid. p. 22. One might well wonder whether O'Shaughnessy refers to a behavioural or a genetic version of doxastic voluntarism here, but immediately above he has made it clear that the genetic version is the intended one: "My question is altogether different. It is: can there be an act that is the bringing about of belief?" Ibid.
302. Williams (1973) p. 149. The left-out passage concerns so-called "B-states," weaker surrogates of belief states that a machine may be in. Ibid p. 145. The issue of B-states is not narrowly relevant to the present context, where we shall be dealing with full-blow beliefs throughout.
303. Govier (1976) p. 648; Bennett (1990) p. 95; Scott-Kakures (1993) p. 84.
304. Govier (1976) *ibid.* See also Bennett (1990).
305. As argued by Scott-Kakures (1993) *ibid.*
306. O'Shaughnessy (1980) p. 25.
307. Williams (1973) p. 148.
308. I take it that Williams' intentions are not violated by replacing "know" with "believe" throughout. Heinrich Wansing, in his careful analysis of Williams' argument, has taken Williams' "know" quite literally. Since knowledge is a factive state, this has the unwelcome consequence of allowing a rather trivial reading of the argument, since, to an adherent of CIC like Williams, any ascription of a belief formed at will is a logical falsehood, hence trivially unknowable. (This is recognized by Wansing. See especially Wansing (2002) p. 25). It seems clear from the context that, rather than a circular argument based on the factivity of knowledge, Williams is after basing CIC on alleged (conceptual) necessities concerning human psychological capabilities, especially such as concerns beliefs and "suspicions." See Engel (1999) for a very positive stance towards Williams' strategy.
309. Williams (1973) p. 136.

310. Williams (1973) p. 137.
311. *Ibid.*
312. *Ibid.*
313. Govier (1976) p. 647. A problem concerning Govier's phrasing might be this: A belief that *p* is not automatically a belief that *p* is true; Children may, e.g. hold beliefs and still lack a concept of truth. However, this difficulty does not seem to affect Govier's point, which only concerns "most people."
314. Williams (1973) p. 148.
315. A perfectly parallel claim is made in O'Shaughnessy (1980) p. 25 where O'Shaughnessy argues that "only the very mad" come *near* consciously to hold such incompatible belief, and that such beliefs would hardly qualify as "genuine" beliefs at all.
316. As any false mathematical proposition, I take it, is necessarily false the conjunction of the proposition with itself is also necessarily false. Further an agent who believes a mathematical proposition *p* plausibly may also be attributed the belief *p* & *p* and vice versa.
317. Williams (1973) p. 148.
318. Winters (1979) p. 254.
319. Winters (1979) p. 254; Scott-Kakures (1993) pp. 82 and 94; and Bennett (1990) p. 93. Bennett sets up the point by an elaborate thought example involving self-deceptive agents called "Credamites": "When a Credamite gets a belief in this way (by fiat) he forgets that this is how he came by it" (*ibid.*).
320. Scott-Kakures (1993) p. 99.
321. Montmarquet (1986) p. 50. I shall leave it open whether such "incontinence" really deserves the coveted but controversial title of "epistemic akrasia."
322. Dana Radcliffe seems to have gone for a *prospective* version of the stronger negative requirement: "There is no incompatibility between recognizing that I have not yet come to a belief that Tom is guilty and the absence of a second-order epistemic assessment of that possible belief ... (...) ... Does my lacking an opinion so far about the epistemic status of that belief prevent me from intending or willing to believe it? Surely not!" Radcliffe (1997) p. 149. Now, as seen above, I take the prospective arguments to either beg the question or rely on the retrospective claim. Thus, I see no reason for pursuing such prospective concerns in detail.
323. Winters consider two cases of *capricious* belief in the present sense: Beliefs about future matters of fact held by some self-acclaimed psychic and religious beliefs supposedly self-induced by some mysterious "leap of faith." Winters (1979) p. 250–251. Here, I take it, only the latter category meets the specification. Surely, a psychic would normally quote the "fact" that her belief about the future was produced by her psychic powers as a good rationalizing reason for holding it. Further, we should hardly expect capricious beliefs to remain capricious for long. I take it that most persons are prone to immediately take *something* to constitute positive evidence for even their most farfetched beliefs, once formed.
324. Scott-Kakures (1993) p. 88.
325. Scott-Kakures (1993) p. 91.
326. Scott-Kakures (1993) p. 95.
327. These are bizarre cases, where simply holding the belief in itself generates a good rationalizing reason for holding it. Scott-Kakures mentions the case, where someone credibly offers you \$1,000,000 for forming the belief that you are a millionaire. If somehow you succeed in believing this, you instantly have a good rationalizing reason for believing it. I believe we may safely ignore the relevance of this rare phenomenon to the wider issue of doxastic voluntarism.
328. Scott-Kakures (1993) p. 94.
329. Wansing (2004) p. 424. See Wansing (2002) and Wansing (2004) for a fully adequate and technically sound introduction to Wansing's thought on doxastic control. Wansing (2006) presents a more accessible up-to-date introduction. I am grateful towards Henirich Wansing for aiding me in presenting his views in a reasonably fair, albeit relatively non-technical manner.
330. Alston (1989) p. 123.
331. Ryan (2003) p. 66.

332. "Once we have judged that our evidence supports believing that p , we believe that p automatically and instantaneously. No effort is required and we are certainly not aware of any act such as "executing" our decision. However, these differences do not undermine my main point: that one has executed a decision to φ because one has concluded that one's reasons or one's evidence supports φ -ing." Steup (2000) p. 35.
333. Below (see, e.g. Section 9.4), I shall in fact endorse Steup's view against Alston that an agent can be blameworthy even for "pressing" empirical beliefs.
334. Ginet (2001) p. 70.
335. Ibid. p. 124.
336. Ibid. p. 125.
337. Ibid. p. 126.
338. Ibid. p. 123.
339. In Section 2.1 I have offered reasons to hold strictly apart the notions of acceptance and belief.
340. Ibid. p. 122.
341. I am not aware of any serious psychological research regarding doxastic actions such as the intentional formation of belief.
342. Cf., a.o., Feldman (2001) p. 83: "we are at the mercy of our evidence (or perhaps our evidence together with what other non-evidential factors cause beliefs)." Or in Robert Audi's suggestive words: "belief is more like a response to external grounds than a result of an internal volitive thrust." Audi (2001) p. 98. I am however, not sure what good the internal/external distinction does here. John Heil has submitted that "believers are largely at the mercy of their belief-forming equipment." Heil (1983) p. 357.
343. Trudy Govier has aptly named such a doxastic prodigy "Supermind." Govier (1976) p. 649.
344. Ryan (2003) p. 63.
345. Ibid. p. 62.
346. Ibid. p. 48.
347. "Rather, what we should compare the belief case with is an action that is decisively opposed by practical reasons, such as giving the first person you see on the street a hefty blow to the nose." Steup (2001) p. 16.
348. "It's true that I cannot decide whether to believe a car is coming at me when I see one bearing down on me or whether I have hands, arms and legs. But neither can I decide to stick a knife in my eye for no reason, nor to kill a baby or rob a bank." Russell (2001) p. 42.
349. Ryan (2003) pp. 70–71.
350. Stocker (1982) p. 410. A parallel claim is made in Zagzebski (1996) pp. 66–67.
351. Arguably, neither author is strictly committed to this claim, although Stocker takes the analogy to establish at least that an agent may be responsible for her holding of beliefs (ibid.) and Ryan, as seen above, clearly holds free/intentional actions to be proper objects of blame and takes all formations of beliefs to be instances of such free or intentional action. If I am setting up one or more straw men here, it is only for the purpose of showing that the argument poses no threat to my rejection of doxastic voluntarism.
352. See Section 7.2.
353. Alston (1989) p. 133.
354. Davidson (1980) p. 53.
355. See in particular ibid. p. 58.
356. Heller (2001) p. 133.
357. Ibid. p. 131.
358. Ibid. p. 133.
359. Ibid. p. 136.
360. Ibid. p. 135.
361. Ibid.
362. Ibid. p. 136.
363. Ibid. p. 138.

364. Ibid. p. 139.
365. Baron (2001) p. 8.
366. Ginet (2001) p. 64.
367. It seems clear that he uses the term “decision” to mean simply *intention* in the above context. This impression is only strengthened later in the paper, where Ginet shifts freely between the terms “decision” and “intention” in parallel contexts.
368. Ibid. p. 70.
369. Adopted from *ibid.* p. 65. Ginet presents “counting on its being the case” as a necessary and sufficient condition for belief in the run of an example involving the belief that a coin-flip lands heads. However, he immediately proceeds to employ this observation on divergent instances of belief. Thus, it seems plausible to presume that Ginet intends to offer “counting on its being the case” as a general definition of “belief.”
370. Ibid.
371. Ibid. p. 66.
372. Ibid.
373. Remark, though, that Ginet seems to eschew this defence by submitting above that “to not prepare for the truth of not-p is itself to stake something on p.”
374. See Alston (1989) p. 123.
375. Ginet (2001) p. 66.
376. Ibid. p. 64. On the same page, Ginet offers two examples with “suppression” strategies for forming beliefs, parallel to the above example of the man with the injured wife. Further he offers an example with a deliberating poker player. This example, in any important respect, is parallel to the above-quoted courtroom example.
377. This is not to deny, of course, that action may in many cases occur without precedent value-decisions.
378. Steup (2000) p. 30.
379. Ibid.
380. Ibid. p. 31.
381. Ibid.
382. Ibid. p. 32.
383. Ibid. p. 31.
384. Ibid. p. 32.
385. Ibid. p. 45.
386. See especially *ibid.* pp. 35–36 and p. 46 respectively.
387. See the quotation at the beginning of Chapter 8.
388. Ibid. p. 35.
389. See, e.g. Bratman (1999) p. 6. Donald Davidson has championed a theory of intention closely related to S-intention: “To intend to perform an action is, on my account, to hold that it is desirable to perform an action of a certain sort in the light of what one believes is and will be the case.” Davidson (1980) p. 100.
390. Ibid. p. 28.
391. Ibid. p. 14.
392. Heil (1983) p. 363.
393. However, standard doxastic voluntarism opting at all instances for either the first or the second strategy (see Section 8.1) available to it for accounting for an agent’s blameworthiness for holding an undesirable belief, *entails* property voluntarism in the following sense: If the agent is epistemically blameworthy because she performed the action of *forming* a particular undesirable belief, she also brought it about that an epistemically undesirable belief was formed, and could be equally blamed on this basis. If she is epistemically blameworthy because she should have performed, but did not perform, the doxastic action of *preventing* herself from holding a particular undesirable belief, she also should have brought it about, but did not bring it about, that an undesirable belief was not formed. She could thus be equally blamed for both omissions.
394. Heil (1983) p. 357.

395. *Ibid.* p. 363.
396. *Ibid.* p. 360.
397. *Ibid.*
398. *Ibid.* p. 361.
399. In the second sentence of Heil (1983) he states that “we occasionally blame a person for believing what we imagine he ought not believe.” *Ibid.* p. 355. He nowhere rebuts the appropriateness of this practice.
400. Montmarquet (1993) p. 46.
401. *Ibid.*
402. *Ibid.* p. 33.
403. *Ibid.*
404. See *ibid.* pp. 30–32.
405. *Ibid.* p. viii.
406. “My question, then is this with respect to such [evil] acts: If we cannot assign culpability for holding such [nazi] beliefs, how can we assign culpability for acts premised on their (assumed) truth?” *Ibid.* p. 2. Montmarquet clearly thinks, as do I (see Section 1.3), that such acts may be (morally) blameworthy, even if “honestly mistaken.”
407. It should be remarked though, that elsewhere in Montmarquet (1993), Montmarquet seems to lapse back into a classical doxastic voluntarist position, despite his earlier explicit remarks quoted above that “epistemic responsibility” does not hinge upon control over particular “given” beliefs: He defends the view that a belief may be “weakly voluntary” “to the extent that it is formed or held under circumstances (a) allowing for, but not dictating, its epistemically virtuous formation or retention; and that (b) had the subject not been epistemically virtuous, this belief would not have been held, or continued to be held with the same degree of conviction.” It is condition (b) of this definition, that more than smacks of doxastic voluntarism. As shall be seen in Section 10.4, genetic property control need make no such requirements: There are so-called “Frankfurt cases” where (b) is not satisfied and yet the belief is still under a mode of doxastic control relevant for ascriptions of epistemic blameworthiness. Thus the more charitable option is going with Montmarquet’s property voluntarist remarks, as I have done above. In Montmarquet (1999) he seems to have opted for this reading himself, where he roots epistemic blameworthiness entirely in an agent’s “direct control with respect to [her] degree of care.” *Ibid.* p. 843.
408. *Ibid.* pp. 47–48.
409. Arguably Hitler, as a well-known historical figure, distorts our relevant intuitions in this context. To avoid this, we could substitute him with a less famous Nazi henchman.
410. See *ibid.* p. 2.
411. See, e.g. Zagzebski (1996) pp. 84–88.
412. *Ibid.* p. 142.
413. *Ibid.* p. 137.
414. *Ibid.* p. 141.
415. *Ibid.* p. 152.
416. *Ibid.*
417. I am of course offering a totally ludicrous reading of the term “belief forming habit” here. My point is, however, that Alston does nothing to rule out such a reading. He only submits that the “belief forming habit” clause of his analysis is relevant in cases, where “the belief was acquired by the activation of a habit we would not have possessed had we fulfilled our intellectual obligations.” Anna’s “habit” of forming beliefs about Benny fits this picture perfectly.
418. Cf. Zimmermann (1997).
419. I shall not attempt to spell out exactly which kind of dispositions adequately fill in Alston’s picture.
420. Alston (1989) p. 142.
421. *Ibid.* p. 138.
422. *Ibid.*
423. Govier (1976) p. 643.

424. Some may now wonder, why I took so great care in refining Alston's analysis. I did this because of the attention it has enjoyed as the only explicit doxastic Pascalianist position in the published literature. It had to be made absolutely clear that even an extensive refinement as that undertaken above will not save Alston from the objections I shall offer.
425. Owens (2000) p. 83. The "juridical theory of responsibility" is here simply the claim that "responsibility requires freedom and control" (ibid. p. 21).
426. I shall discuss in Section 13.2 the intricate question of what "required to foresee" could mean in the present context.
427. Frankfurt (1969).
428. Fischer & Ravizza (1993b) p. 327. The authors emphasize that of course Frankfurt is not committed to such extensions of his original point, which did not regard responsibility for consequences. However, to me as to Fischer & Ravizza the extension seems perfectly natural. The supposition that Elizabeth "has had" the device implanted in her is ambiguous. I take it to mean that the scientist implanted the device in her brain without her consent.
429. Notice that the slight time-gap between the reliable and the unreliable testimony in the above example is not essential. One might imagine that the demagogue and the criminologist express the statistical proposition simultaneously.
430. Stocker (1982) makes much of this point: "In regard to at least many [formations of] serious beliefs, we simply do not have, and would not realistically be able to have, foresight." Ibid. p. 403.
431. Alston (1989) p. 135.
432. See in particular Graham (1997) throughout and Pappas (2000) pp. 118–119.
433. This conception is arguably more relaxed than any conception expressly endorsed in the literature and most likely also inadequate for many purposes. For example the proposition intended to be expressed need not be the one perceived to be expressed, in order for an utterance to qualify as a piece of testimony on my account. However, for the present purpose, this very wide conception should suffice.
434. "Working-indicator-properties, which are those properties actually used in a given practice to indicate rational authority, and which may or may not be so reliable. The norm of credibility governs, who is picked out as a good informant: it tells us to attribute rational authority to all and only those informants who possess relevant indicator-properties." Fricker (1998) p. 168. "Rational authority" here is simply Fricker's term for "credibility."
435. P. Faulkner (2002) p. 357.
436. Ibid.
437. See, e.g. Hardwig (1991) p. 693 for this common observation.
438. That is the one corresponding to the notion of substantive responsibility employed by T.M. Scanlon in Scanlon (1998).
439. Heil (1983) p. 362.
440. Kornblith (1983) p. 38.
441. Ibid. p. 34.
442. Scare-quotes due to the immense ambiguity of the term. See Section 3.2.
443. Ibid. p. 33.
444. Ibid. p. 34.
445. Ibid. p. 38.
446. In fact this does not seem to follow from his above-quoted definition: Even though an agent has not done "all that he should in order to bring it about that he have true beliefs," he might still (at least occasionally) hold an epistemically desirable belief.
447. It should be noted that the term "education" is here taken in a very broad sense as anything that might systematically improve an agent's cognitive dispositions relevant to belief, i.e. anything that may systematically improve her propensity to serve the truth-goal. I shall of course not assume that the value of education uniquely derives from this propensity.
448. Ibid. p. 38.
449. Ginet (2001) p. 64.

450. Alston (1989) pp. 137–138. See also Kornblith (1983) p. 35: “Being justified requires more than simply reasoning properly; it requires that one gather evidence properly as well.”
451. Talking of the non-performance and performance of an omission is awkward. Arguably, however the English language has no term that covers both actions and omissions in a natural way. Further, it should be noted that any action may also be described as an omission (of not performing the action) and vice versa. Whether to invoke one term or the other is often a hard decision, guided only on ideas of what is the most “natural” phrasing in a given context.
452. I take it that I am in consonance with Alston’s usage of the term, although he appears very unwilling to offer any substantial account of the notion of an intellectual obligation apart from the unrevealing indirect statement that the fulfilling of such an obligation is what it is “reasonably to expect of a person.” Alston (1989) p. 145.
453. See, e.g. Alston (1989) p. 86: “S is J_d [deontologically justified] in believing that p *iff* in believing that p S is not violating any epistemic obligations.” See also Castaneda (1988) throughout and Steup (1988) p. 67.
454. See especially Haack (2001) pp. 26–28. As pointed out in Section 1.3, Haack gets a very uncharitable and implausible interpretation out of Clifford by trying to fit him into the procrustean bed of the modern notion of epistemic justification. Especially Haack takes Clifford to argue that all epistemically unjustified beliefs are also somehow “morally culpable,” a thesis without any warrant in Clifford’s writings at all. Rather, I will aim to make clear below, Clifford’s main concern was our obligations “duties” to perform certain belief-influencing actions of roughly the types outlined in Section 11.3, and the influence which our failure to fulfil such duties might have on our service to the Enlightenment tradition of the Western world. (Clifford (1999) has some very bad things to say about religious dogmatism, where Muslim culture functions as an easily identifiable dummy for an outright attack on Christian fundamentalism, see especially *ibid.* pp. 80–81). Haack’s interpretation of Clifford is certainly “revisionary,” but not that “friendly” after all! (“This will suggest a friendly, if revisionary interpretation of, what is plausible in Clifford’s ... talk of “the Ethics of belief.” Haack (2001) p. 22.
455. *Ibid.* p. 25.
456. Smith (1983) p. 547.
457. “Frequently the benighting act is an omission.” *Ibid.* p. 547.
458. “But most discussions fail to emphasize (or even note) that the benighting act must be more than objectively wrong; it must also be one for which the agent is culpable. If the agent had an excuse for the benighting act, then the subsequent wrongful act is obviously blameless as well.” *Ibid.* p. 548. I wholeheartedly agree to this point.
459. *Ibid.* p. 543.
460. *Ibid.* p. 545.
461. *Ibid.*
462. Whether, when such routines are not specifically pointed out, a doctor should in criminal contexts be held to higher standards of care than a laymen when operating within her professional setting, is a matter of some controversy. See, e.g. Saltzburg et al. (1994) p. 213.
463. This to intercept the (unwarranted) objection that the “ought” in, e.g. the sentence “An agent *ought* to fulfil her institutional intellectual obligations” is only a “role-ought” in Richard Feldman’s sense. See, e.g. Feldman (2001) pp. 87–88. As pointed out by Feldman, in my opinion correctly, a “role-ought” is not a deontic “ought” in the sense that it does not bear on evaluations of blameworthiness (*ibid.*).
464. *Ibid.* p. 556. In Section 1.3 I argued against the *necessity* of reprehensible desires and aversions for morally blameworthy action. Here their *sufficiency* is also questioned.
465. “For example, the near-sighted driver may thoroughly repent her willingness to leave the house without her glasses; given the same options now, she would reject the one she chose. Or, even if the faulty desire persists, it may not play any direct role in producing the unwitting act.” *Ibid.* p. 559. In fact, Smith’s example is somewhat unfortunate, insofar as we would normally expect a near-sighted driver to stand under some (perhaps vaguely defined) *legal* duty to wear her glasses when driving.

That is: we would expect a violation of an *institutional* intellectual obligation to obtain, whose source is not that hard to point out (the relevant traffic rules). For the sake of argument, I shall however follow through with Smith's example on her terms.

466. Kornblith (1983) p. 34.
 467. *Ibid.* p. 36.
 468. Alston (1989) p. 146.
 469. Clifford (1999) p. 76.
 470. *Ibid.* pp. 74–75.
 471. *Ibid.* p. 91.
 472. *Ibid.*
 473. *Ibid.* p. 74.
 474. Notice that Clifford does not make distinctions between relevant and irrelevant beliefs or truths: His famous slogan goes: “It is wrong always, everywhere and for anyone to believe anything upon insufficient evidence.” *Ibid.* p. 77.
 475. Montmarquet (1993) p. 29.
 476. If this conception of non-institutional intellectual obligation strikes my reader as too lean, remember that I am here (cf. Section 12.1) invoking the notion in a sense narrowly concerning belief-influencing actions relevant to ascriptions of epistemic blameworthiness. Of course, in other contexts, a much richer concept of intellectual obligation tied to broader epistemic goals as, e.g. wisdom or understanding would be perfectly legitimate (see also Section 5.5).
 477. *Ibid.* and Zagzebski (1996).
 478. Nicomachean Ethics 1094b 20–23.
 479. In fact the literature on the deontological relevance of foresight has been almost singularly obsessed with the relation between foresight and intention. See, e.g. Baldwin (1979) and Aulisio (1996).
 480. Zimmermann (1986) p. 206.
 481. Since there is no sign that Zimmermann employs a similar distinction, I take it that this is more of a clarification within my own preferred terminology.
 482. I assume that this is possible on a folksy version of the physics of chaotic systems having it that a butterfly flapping its wings in South America can cause a storm in Australia.
 483. I am grateful to Benjamin Schnieder for drawing my attention to this point, which was not treated with sufficient care in Nottelmann (2004).
 484. I shall assume here that acceptances are “occurrent” or “conscious” attitudes in the sense intended by, e.g. Zimmermann. If we shall have to allow unconscious acceptances, arguably this would complicate immensely the present discussion. However, I do not find it clear how much sense can be made of the claim that in performing an action, an agent unconsciously used a certain proposition as a premise in her practical reasoning.
 485. This phrase is somewhat ambiguous: notice that it is what the agent foresees will happen as a result of her action of omission that is important, not what, given that she performs a certain action or omission, the agent foresees.
 486. To not further complicate the present context, I have here ignored the extremely unlikely possibility that the agent may have openly contradictory beliefs about the matter.
 487. What qualifies as “being of the same kind” here is extremely hard to spell out in further details. In the case of the “educated racist” it seems enough to render her epistemically blameworthy for holding her racist belief that blacks are generally inferior to whites, if she believed at the time of her skipping class that a significant risk obtained, that as a result of her inattentiveness, she would form undesirable beliefs with some kind of racist content in the future. Also if she accepted this about future beliefs about other people, it strikes at least this author as being enough. However, if all she accepted was that, as a result of her skipping class, she would form undesirable beliefs about, say, train engines in the future, this does not seem the right basis for rendering her epistemically blameworthy for holding the undesirable racist belief, unless other of her foresights or blameworthy inadvertences are more relevant to her later undesirable racist belief.

488. Ibid. p. 102.
489. Alexander (1990) p. 103.
490. See, e.g. Oakley & Cocking (1994) pp. 205–206: “According to a widely-accepted account, the concept of moral responsibility involves two conditions; avoidability and foreseeability. That is, I am morally responsible for consequences C if, and only if, they are among the foreseeable results of what I could reasonably have avoided doing.”
491. The example given, of course, is an attempt to turn to one extreme “the foresight button.” I acknowledge that this involves certain implausibilities. For example we would hardly expect a harshly convicted person to attend some educational program just before going to jail. Also we would normally expect certain belief-influencing actions or omissions undertaken after the release to be of relevance. I realize that these implausibilities may well disturb our frail intuitions. However, a more mundane example might well be hard to find.
492. Cf. Smith (1983).
493. Rorty (1989) p. 75.
494. Alston (1989) p. 145.
495. Kuhn (1962). Before Kuhn’s widely discussed book, e.g. the Austrian sociologist and Vienna Circle Member Otto Neurath poignantly referred to the natural science tradition as a “modern scientific folklore.” See Neurath (1941). Whether Kuhn’s picture of physics as a discipline characterized by long periods of calm hegemony brutally separated by violent conceptual revolutions, is generally adequate, is of course a highly controversial matter. I take it here, that such hegemony may at least characterize the life of particular national schools within a scientific discipline for limited periods of time.
496. Steup (1988) p. 78.
497. Alston, in responding to Steup’s criticisms, makes this point thus: “I think that Steup is displaying an insensitivity to cultural difference. He supposes that there are standards recognized in all cultures that determine what is adequate evidence or good enough grounds, for one or another kind of belief. That does not seem to me to be the case.” Alston (1989) p. 146.
498. Alston (1989) p. 145.
499. Nicomachean Ethics 1113b line 32–34. In contrast to most modern legislations, the Greeks took intoxication in the moment of crime as a severely aggravating circumstance, since the offender was then guilty of making himself ignorant *as well as* of causing a further social harm.
500. Alston (1989) p. 145.
501. Ibid. p. 146.
502. Ibid. p. 146.
503. Ibid.
504. Steup (1988) pp. 79–80.
505. Ibid. p. 80.
506. Alston coins this response to Steup’s rejoinder thus: “The case I have in mind is one in which the student is quite confident in his reading; this is definitely the way it strikes him, and he has no tendency to doubt it (at least not prior to seeing the grade he gets on the final exam).” Alston (1989) p. 147.
507. Steup (1988) p. 79. See the passage quoted at length above.
508. This is clearly a more charitably reading than assuming that Alston conceives of an agent’s “cognitive deficiency” as her disability to form an adequately based belief after having been confronted with evidence, from which a more cognitively optimal agent *would have formed an adequately based belief*. Certainly in the college student case, a more refined student, would most likely have come up with an adequately based belief about the main thesis of Book IV of Locke’s *Essay*. However, we might well imagine a case, where a strikingly cognitively deficient agent forms an inadequately based belief in a situation, where a more cognitively optimal agent would not have formed a belief at all.
509. Montmarquet (1993) p. viii.
510. See *ibid.* pp. 45–48.

511. Ibid. p. 46. I take it that Montmarquet's use of the term "accept" here is simply a slip of the pen, since elsewhere in Montmarquet (1993) he strongly emphasizes that it is *beliefs* rather than acceptances with which he is normatively concerned. See in particular *ibid.* pp. 90–92.
512. Montmarquet (1999) p. 844.
513. Montmarquet (1993) p. 47.
514. I shall follow Korsgaard (1999) in using the predicate "autonomous" in the present sense as applicable directly to actions. Of course actions in this context must be tacitly understood as individuated by their motivations; it is the right kind of *motivation*, namely respect (Achtung) for the categorical imperative or moral law, that makes for autonomy, not the action's physical aspects.
515. Kant persistently employs the metaphor of a mechanical spring (Triebfeder) in order to convey the influence of the agent's respect for maxims on her actions: "Auch haben wir oben gezeigt, wie weder Furcht, noch Neigung, sondern lediglich Achtung fürs Gesetz diejenige Triebfeder sei, die der Handlung einen moralischen Werth geben kann." *Grundlegung*. Akademieausgabe 4.440.5–7. This metaphor strongly suggests a causal interpretation, however this seems ruled out by his doctrine of the unconstrained operation of the free will (see below). We are facing here one of the deep problems of Kantian exegesis. I will not further explore this matter in the present context.
516. *Grundlegung*. Akademieausgabe 4.446.7–10.
517. Ibid. Akademieausgabe 4.441.3–7.
518. Ibid. Akademieausgabe 4.437.6–13.
519. This follows trivially from the observation that the ideal of autonomy is an ideal for acts of the will.
520. Korsgaard (1999) p. 1.
521. Ibid. p. 14.
522. Ibid. p. 14–16 ff.
523. The terms "responsible" and "responsibility" have been used in a confusing multitude of interrelated senses as has been discussed in Section 3.2. Here I presuppose the following usage: An agent is responsible for *b* if, and only if, she has *authored* *b* in the sense relevant to the application of deontological predicates to the agent with respect to *b*, *b* being a doxastic state or (as here) an action.
524. The acute reader may object that I have given no argument for the joint sufficiency of the three condition. I am not going to give such an argument, since in the present context, properly speaking the conditions all express the same Kantian claim in different phrasings. In that sense each condition, properly understood, is in itself sufficient.
525. Owens (2000) p. 5.
526. For an interesting discussion see Richardson (2001). As one should perhaps expect, the Kantian concept does not measure up well to Richardson's rich stock of practical examples. However, Richardson's discussion suffers from his unwillingness to distinguish properly between the metaphysical, volitional and normative aspects of Kantian autonomy as out-lined above. Thus, he discusses threats to volitional and metaphysical autonomy as, e.g. heroin addiction on a par with threats to normative autonomy as, e.g. evil-conducive character traits.
527. See, e.g. Fuller (1993) p. 209.
528. Polanyi (1951) p. 33.
529. Ibid. p. 34.
530. Arnault (1989).
531. Govier (1993) p. 103.
532. Epictetus (1968) p. 213. The hardships that an agent must sometimes suffer to realize this state of "freedom" is, e.g. brought out in the following passage: "This is the point where we must begin, this is where the citadel must be destroyed, and the tyrants cast forth; we must give up our body, and all that belongs to it – faculties, property, reputation, offices, honours, children, brothers, friends – all these we must regard as having no concern to us." Ibid. p. 223.
533. Fuller (1993) p. 208. Fuller imagines that Epictetus, a slave set free, developed his idiosyncratic conception of freedom in his days of slavery in order to cope with his strongly constrained social condition. Fuller also explicitly represents Epictetus as a model of Nietzsche's negatively charged concept of "slave morality" (*ibid.*). However, in fact Nietzsche treats Epictetus' stern doctrines with

- reverence and contrasts them with the “weak” slave morality of Christianity. See, e.g. Nietzsche (1966) I, 1267.
534. Reich der Zwecke. Grundlegung. Akademieausgabe 4.438.18–21: “Demnach muss ein jedes vernünftige Wesen so handeln, als ob es durch seine Maximen jederzeit ein gesetzgebendes Glied im allgemeinen Reiche der Zwecke wäre.” Speaking of role-autonomy in this context is somewhat inappropriate, as certainly Kant conceived of the Kingdom of Ends more like a regulative ideal than a society actually realized. The prize that Kant puts on a lack of epistemic deference concerning moral matters should, however, be obvious. Perhaps Kant’s invocation of a “Kingdom of Ends” may be diagnosed as a grandiose effort to establish as a norm for individual conduct even a good such as epistemic non-deference, which is inherently linked to a certain political organisation.
535. See Montmarquet (1993) p. 27 for further reflections of virtues as habits.
536. The massive focus in “social epistemology” on testimony as a social epistemic source owes largely to the contemporary discussion’s roots in Hume’s Enquiry (Hume (1975)) Sect. X (On Miracles). In many ways this focus is unfortunate since adoption of beliefs through testimony is arguably only one of several kinds of epistemic deference, which may also manifest itself as, e.g. imitation of doxastic practices or a transfer of tacit knowledge.
537. Schmitt (1987) p. 61. It should be noted that John Hardwig invokes the related notions of “epistemic self-reliance” and “epistemic self-sufficiency” in Hardwig (1991) p. 701. For a thorough discussion of Hume’s so-called “reductionist epistemology of testimony” see, e.g. D. Faulkner (1998).
538. For example according to Robert Paul Wolff, the gist of autonomy is a markedly asocial “refusal to be ruled.” Wolff (1970) p. 18: “The primary obligation of man is autonomy, the refusal to be ruled.” To Wolff this is only realized when man “fulfils his obligations to make himself the author of his decisions.” (ibid.). Here no room for intersubjective or wider social authorship is made at all.
539. Govier (1993) p. 104.
540. Ibid. pp. 102–103.
541. For a thorough discussion of Locke’s and Hume’s ideals concerning the management of testimony, see, e.g. Schmitt (1987).
542. Clifford (1999) p. 76.

REFERENCES

- Achinstein, P. 1983, "Concepts of Evidence," in Achinstein, P. (ed.): *The Concept of Evidence*, Oxford University Press, Oxford, pp. 145–174.
- Alexander, L. 1990, "Reconsidering the Relationship Among Voluntary Acts, Strict Liability, and Negligence in Criminal Law," *Social Philosophy & Policy* 7, 84–104.
- Alston, W.P. 1978, "Meta-Ethics and Meta-Epistemology," in Goldman, A.I. & Kim, J. (eds): *Values and Morals*, Reidel, Dordrecht, pp. 275–297.
- Alston, W.P. 1989, *Epistemic Justification. Essays in the Theory of Knowledge*, Cornell University Press, London.
- Alston, W.P. 1993, "Epistemic Desiderata," *Philosophy and Phenomenological Research* 53, 527–551.
- Archard, D. 1999, "The *Mens Rea* of Rape: Reasonableness and Culpable Mistakes," in Burgess-Jackson, K. (ed.): *"A Most Detestable Crime": New Philosophical Essays on Rape*, Oxford University Press, New York, pp. 213–229.
- Arnault, L. 1989, "The Radical Future of a Classic Moral Theory," in Jaggar, A. & Bordo, S.R. (eds): *Gender/Body/Knowledge: Feminist Reconstructions of Being and Knowing*, Rutgers University Press, New Brunswick, pp. 188–207.
- Audi, R. 2001, "Doxastic Voluntarism and the Ethics of Belief," in Steup (ed.) [2001], pp. 93–111.
- Aulisio, M.P. 1996, "On the Importance of the Intention/Foresight Distinction," *American Catholic Philosophical Quarterly* 70, 189–205.
- Baldwin, T. 1979, "Foresight and Responsibility," *Philosophy* 54, 347–360.
- Baron, M.W. 2001, "I Thought She Consented," *Philosophical Issues* 11, 1–32.
- Bartley III, W.W. 1987, "Alienation Alienated: The Economics of Knowledge versus the Psychology and Sociology of Knowledge," in Radnitzky, G. & Bartley III, W.W. (eds): *Evolutionary Epistemology, Rationality and the Sociology of Knowledge*, Open Court, LaSalle, pp. 423–452.
- Bennett, J. 1990, "Why is Belief Involuntary?" *Analysis* 50, 87–107.
- Bratman, M.E. 1999, *Intention, Plans, and Practical Reason*, CSLI Publications, Stanford.
- Castaneda, Hector-Neri 1988, "Knowledge and Epistemic Obligation," in Tomberlin, J.E. (ed.): *Philosophical Perspectives 2. Epistemology*, Ridgeview, Atascadero, pp. 211–233.
- Clifford, W.K. 1999, *The Ethics of Belief and Other Essays*, Prometheus Books, Amherst.
- Code, L. 1987, *Epistemic Responsibility*, University Press of New England, Hanover.
- Cohen, L.J. 1989, "Belief and Acceptance," *Mind* 98, no. 391, 367–389.
- Curley, E.M. 1975, "Descartes, Spinoza, and the Ethics of Belief," in Mandelbaum, M. & Freeman, E. (eds): *Spinoza: Essays in Interpretation*, Open Court, LaSalle, pp. 169–189.
- Curley, E.M. 1976, "Excusing Rape," *Philosophy and Public Affairs* 5, 325–360.
- Dancy, J. & Sosa, E. (eds): 1992, *A Companion to Epistemology*, Blackwell, London.
- David, M. 2001, "Truth as the Epistemic Goal," in Steup (ed.) [2001], pp. 151–169.
- Davidson, D. 1980, *Essays on Actions and Events*, Oxford University Press, Oxford.
- Dennett, D.C. 1969, *Content and Consciousness*, Routledge & Kegan Paul, London.
- DePaul, M.R. 2001, "Value Monism in Epistemology," Steup (ed.) [2001], pp. 170–186.
- Descartes, R. 1993, *Discourse on Method and Meditations on First Philosophy*, 3rd edition, translated by Donald A.C., Hackett, Indianapolis.
- Dressler, J. 1998, "Where We Have Been, and Where We Might Be Going: Some Cautionary Reflections on Rape Law Reform," *Cleveland State Law Review* 46, 410–442.
- Dressler, J. 2001, *Understanding Criminal Law*, 3rd edition, Lexis Publishing, New York.

- Engel, P. 1999, "Volitionism and Voluntarism about Belief," in Meijers, A. (ed.): *Belief, Cognition and the Will*, Tillburg University Press, Tillburg, pp. 9–25.
- Epictetus. 1968, *The Discourses of Epictetus*, translated by Matheson, P.E. The Heritage Press, New York.
- Faulkner, D. 1998, "David Hume's Reductionist Epistemology of Testimony," *Pacific Philosophical Quarterly* 79, 302–313.
- Faulkner, P. 2002, "On the Rationality of Our Response to Testimony," *Synthese* 131, 353–370.
- Feldman, R. 2001, "Voluntary Belief and Epistemic Evaluation," Steup (ed.) [2001], pp. 77–92.
- Fischer, J.M. & Ravizza, M. (eds) 1993, *Perspectives on Moral Responsibility*, Cornell University Press, London.
- Fischer, J.M. & Ravizza, M. 1993a, "Introduction," Fischer & Ravizza (eds) [1993], pp. 1–44.
- Fischer, J.M. & Ravizza, M. 1993b, "Responsibility for Consequences," in Fischer & Ravizza (eds) [1993], pp. 322–347.
- Foley, R. & Fumerton, R. 1982, "Epistemic Indolence," *Mind* 151, 38–56.
- Frankfurt, H.G. 1969, "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 45, 829–839.
- Fricker, M. 1998, "Rational Authority and Social Power: Towards a truly Social Epistemology," *Proceedings of the Aristotelian Society* 98, 159–177.
- Fuller, S. 1993, *Philosophy of Science and its Discontents*, 2nd Edition, The Guilford Press, New York.
- Gale, R.M. 1999, "William James and the Willfulness of Belief," *Philosophy and Phenomenological Research* 59, 71–91.
- Ginet, C. 2001, "Deciding to Believe," in Steup (ed.) [2001], pp. 63–76.
- Goldman, A.I. 1988, "Strong and Weak Justification," in Tomberlin, J.E. (ed.): *Philosophical Perspectives* 2. *Epistemology*, Ridgeview, Atascadero, pp. 51–69.
- Goldman, A.I. 1993, "Epistemic Folkways and Scientific Epistemology," in Goldman, A.I. (ed.): *Readings in Philosophy and Cognitive Science*, MIT Press, Cambridge, pp. 95–116.
- Govier, T. 1976, "Belief, Values, and the Will," *Dialogue* 15, 642–663.
- Govier, T. 1993, "Self-Trust, Autonomy, and Self-Esteem," *Hypatia* 8, 99–120.
- Graham, P.J. 1997, "What is Testimony?," *The Philosophical Quarterly* 47 (187), 227–232.
- Haack, S. 2001, "The Ethics of Belief" Reconsidered," in Steup (ed.) [2001] pp. 21–33.
- Hall, J. 1963, "Negligent Behavior Should be Excluded from Penal Liability," *Columbia Law Review* 63, 632–644.
- Hardwig, J. 1991, "The Role of Trust in Knowledge," *The Journal of Philosophy* 88, 693–708.
- Heil, J. 1983, "Doxastic Agency," *Philosophical Studies* 43, 355–364.
- Heil, J. 1984, "Doxastic Incontinence," *Mind* 153, 56–70.
- Heller, M. 2001, "Hobartian Voluntarism: Grounding a Deontological Conception of Epistemic Justification," *Pacific Philosophical Quarterly* 81, 130–141.
- Hume, D. 1975, *Enquiries concerning Human Understanding and Concerning the Principles of Morals*, 3rd edition, with text revised and notes by Niddich, P.H. Clarendon Press, Oxford.
- Husak, D. & Thomas III, G.C. 1992, "Date Rape, Social Convention, and Reasonable Mistakes," *Law and Philosophy* 11, 95–126.
- James, W. 1981, *The Principles of Psychology*, in Frederick H.B. (ed.) *The Works of William James*, Vol. 8–9, Harvard University Press, Cambridge.
- Jones, W.E. 1997, "Why Do We Value Knowledge?," *American Philosophical Quarterly* 34, 423–439.
- Kant, I. 1933, *Critique of Pure Reason*, translated by Norma K.S., Macmillan, London.
- Kaplan, M. 1981, "Rational Acceptance," *Philosophical Studies* 40, 129–145.
- Kenner, L. 1967, "On Blaming," *Mind, New Series* 76, 238–249.
- Kornblith, H. 1983, "Justified Belief and Epistemically Responsible Action," *The Philosophical Review* 92, 33–48.
- Korsgaard, C.M. 1999, "Self-Constitution in the Ethics of Plato and Kant," *The Journal of Ethics* 3, 1–29.
- Kuhn, T.S. 1962, *The Structure of Scientific Revolutions. International Encyclopedia of Unified Science Volume 2, Number 2*, University of Chicago Press, Chicago.

- McGregor, J. 1996, "Why When She Says No She Doesn't Mean Maybe and Doesn't Mean Yes: A Critical Reconstruction of Consent, Sex, and the Law," *Legal Theory* 2, 175–208.
- Montmarquet, J.A. 1986, "The Voluntariness of Belief," *Analysis* 46, 49–53.
- Montmarquet, J.A. 1993, *Epistemic Virtue and Doxastic Responsibility*, Rowman & Littlefield, Lanham.
- Montmarquet, J.A. 1995, "Culpable Ignorance and Excuses," *Philosophical Studies* 80, 41–49.
- Montmarquet, J.A. 1999, "Zimmermann on Culpable Ignorance," *Ethics* 109, 842–845.
- Neurath, O. 1941, "Universal Jargon and Terminology," *Proceedings of the Aristotelian Society, New Series* 41, 127–148.
- Nietzsche, F. 1966, *Werke in Drei Bänden*, Hanser Verlag, Munich.
- Nottelmann, N. 2004, "Foresight and Blameworthiness for Action Consequences," *Danish Yearbook of Philosophy* 39, 67–74.
- Oakley, J. & Cocking, D. 1994, "Consequentialism, Moral Responsibility, and the Intention/Foresight Distinction," *Utilitas* 6, pp. 201–216.
- Oshana, M.A.L. 1997, "Ascriptions of Responsibility," *American Philosophical Quarterly* 34, 71–82.
- O'Shaughnessy, B. 1980, *The Will: A Dual Aspect Theory*, Cambridge University Press, Cambridge.
- Owens, D. 2000, *Reason Without Freedom: The Problem of Epistemic Normativity*, Routledge, London.
- Pappas, G.S. 2000, "Epistemic Deference," *Acta Analytica* 15, 113–126.
- Pojman, L.P. 1985, "Believing and Willing," *Canadian Journal of Philosophy* 15, 37–55.
- Polanyi, M. 1951, *The Logic of Liberty: Reflections and Rejoinders*, Routledge, London.
- Radcliffe, D. 1997, "Scott-Kakures on Believing at Will," *Philosophy and Phenomenological Research* 57, 145–151.
- Rahman, S., Symons, J., Gabbay, D., Van Bendegem, J.-P. (eds): 2004, *Logic, Epistemology and the Unity of Science*, Kluwer, Dordrecht.
- Richardson, H.S. 2001, "Autonomy's Many Normative Presuppositions," *American Philosophical Quarterly* 38, 287–303.
- Rorty, A. 1983, "Akratic Believers," *American Philosophical Quarterly* 20, 175–183.
- Rorty, R. 1989, *Contingency, irony, and solidarity*, Cambridge University Press, Cambridge.
- Russell, B. 2001, "Epistemic and Moral Duty," in Steup (ed.) [2001], pp. 34–62.
- Ryan, S. 2003: "Doxastic Compatibilism and the Ethics of Belief," *Philosophical Studies* 114, 47–79.
- Saltzburg, S.A., Diamond, J.L., Kinports, K., & Morawetz, T.H. 1994, *Criminal Law. Cases and Material*, The Michie Company Law Publishers, Charlottesville.
- Scanlon, T.M. 1998, *What We Owe to Each Other*, The Belknap Press of Harvard University Press, Cambridge.
- Schmitt, F.F. 1983, "Against Epistemic Indolence," *Mind, New Series* 92, 424–429.
- Schmitt, F.F. 1987, "Justification, Sociality, and Autonomy," *Synthese* 73, 43–85.
- Schmitt, F.F. 1994, "The Justification of Group Beliefs," in Schmitt, F.F. (ed.): *Socializing Epistemology: The Social Dimensions of Knowledge*, Rowman & Littlefield, Lanham, pp. 257–287.
- Schulhofer, S.J. 1992, "Taking Sexual Autonomy Seriously: Rape Law and Beyond," *Law and Philosophy* 11, 35–94.
- Scott-Kakures, D. 1993, "On Belief and the Captivity of the Will," *Philosophy and Phenomenological Research* 53, 77–103.
- Sellars, W. 1969, "Language as Thought and as Communication," *Philosophy and Phenomenological Research* 29, 506–527.
- Sellars, W. 1989, *The Metaphysics of Epistemology. Lectures by Wilfrid Sellars*, edited by Pedro V.A., & Ridgeview, A (eds).
- Smith, H. 1983, "Culpable Ignorance," *The Philosophical Review* 92, 543–571.
- Stalnaker, R. 2002, "Epistemic Consequentialism II," *Proceedings of the Aristotelian Society, Supplementary Volume* 76, 153–168.
- Steup, M. 1988, "The Deontic Conception of Epistemic Justification," *Philosophical Studies* 53, 65–84.
- Steup, M. 2000, "Doxastic Voluntarism and Epistemic Deontology," *Acta Analytica* 15, 25–56.
- Steup, M. 2001, "Epistemic Duty, Evidence and Internality," in Steup (ed.) [2001], pp. 134–150.

- Steup, M. (ed.) 2001, *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, Oxford University Press, Oxford.
- Stocker, M. 1982, "Responsibility Especially for Beliefs," *Mind* 151, 398–417.
- Strawson, P. 1993, "Freedom and Resentment," in Fischer & Ravizza (eds) [1993], 45–66.
- Toribio, J. 2002, "Semantic Responsibility," *Philosophical Explorations* 5, 39–58.
- Van Fraassen, B.C. 1984, "Belief and the Will," *The Journal of Philosophy* 81, 235–256.
- Wansing, H. 2002, "Seeing to it that an agent forms a belief," *Logic and Logical Philosophy* 10, 185–197.
- Wansing, H. 2004, "Action-theoretic Aspects of Theory Choice," in Rahman, S., et al. (eds) [2004], 419–436.
- Wansing, H. 2006, "Doxastic Decisions, Epistemic Justification, and the Logic of Agency," *Philosophical Studies* 128, 201–227.
- Williams, B. 1973, "Deciding to Believe," in Williams, Bernard: *Problems of the Self. Philosophical Papers 1956–1972*, Cambridge University Press, London pp. 136–151.
- Williamson, T. 2000, *Knowledge and its Limits*, Oxford University Press, Oxford.
- Wilson, T.D., Hodges, S.D., & LaFleur, S.J. 1995, "Effects of introspecting about reasons: Inferring attitudes from accessible thoughts," *Journal of Personality and Social Psychology* 69, pp. 16–28.
- Winters, B. 1979, "Believing at Will," *The Journal of Philosophy* 76, 243–250.
- Wolff, R. P. 1970, *In Defence of Anarchism*, University of California Press, Berkeley.
- Zagzebski, L.T. 1996, *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*, Cambridge University Press, Cambridge.
- Zimmermann, M.J. 1986, "Negligence and Moral Responsibility," *Nous* 20, 199–218.
- Zimmermann, M.J. 1988, *An Essay on Moral Responsibility*, Rowan & Littlefield, Totowa.
- Zimmermann, M.J. 1997, "A Plea for Accuses," *American Philosophical Quarterly* 34, 229–243.

NAME INDEX

- Achinstein P., 59, 241, 259
 Alexander, Larry 14, 15, 18, 189, 197–200,
 235, 236, 256, 259
 Alston, William 16, 35, 36, 53–55, 62, 79, 81,
 87, 92–94, 96, 120–122, 124, 126, 129,
 134, 136, 142, 153–157, 161–164, 174,
 184, 205, 208, 209, 211, 213–217,
 235–238, 240–245, 250–256, 259
 Aquinas, Thomas 104
 Archard, David 233, 234, 259
 Aristotle 40, 188, 212, 243
 Arnault, Lynne 227, 257, 259
 Audi, Robert 100, 242, 244, 246, 250, 259
 Aulisio, Mark P. 255, 259
- Baldwin, Thomas 255, 259
 Baron, Marcia 130, 233, 234, 242, 251, 259
 Bartley III, W.W. 237, 259
 Bennett, Jonathan 41, 42, 106, 107, 111, 248,
 249, 259
 Bratman, Michael 251, 259
- Castaneda, Hector-Neri 254, 259
 Clifford, William Kingdon 1, 7–10, 178,
 185, 186, 213, 230, 233–235, 254,
 255, 258, 259
 Cocking, Dean 256, 261
 Code, Lorraine 1, 10, 11, 29, 30, 236–238,
 247, 259
 Cohen, Jonathan 25, 26, 31, 236, 237, 259
 Curley, E.M. 6–8, 12, 233–235, 259
- Dancy, Jonathan 236, 240, 242, 259
 David, Marian 19, 26, 79, 129, 158, 225, 234,
 241, 242, 259–261
 Davidson, Donald 88, 126, 127, 244, 245,
 250, 251, 259
 Dennett, Daniel C. 24, 236, 259
 DePaul, Michael R. 241, 259
 Descartes, René 56, 259
 Dressler, Joshua 13, 233, 235, 259
- Einstein, Albert 186
 Engel, Pascal 248, 260
 Epictetus 228, 258, 260
- Faulkner, David 258, 260
 Faulkner, Paul 166, 253, 260
 Feldman, Richard 73, 81, 85, 86, 126, 128,
 244, 250, 254, 260
 Fischer, John Martin 38, 41–45, 161, 238,
 239, 253, 260, 262
 Foley, Richard 241, 260
 Frankfurt, Harry G. 85, 153, 161, 162, 204,
 245, 252, 253, 260
 Fricker, Miranda 166, 253, 260
 Fuller, Steve 228, 257, 258, 260
 Fumerton, Richard 241, 260
- Gale, Richard M. 101, 247, 260
 Ginet, Carl 121, 131–134, 136, 173, 250,
 251, 254, 260
 Goldman, Alvin I. 16, 65–67, 235, 240,
 242, 259, 260
 Govier, Trudy 107, 109, 157, 227,
 248–250, 253, 257,
 258, 260
 Graham, Peter J. 241, 253, 260
- Haack, Susan 7, 8, 178, 185, 234, 254, 260
 Hall, Jerome 13, 235, 260
 Hardwig, John 30, 237, 253, 258, 260
 Heil, John 89, 96, 143–145, 150,
 169, 170, 172, 243, 245,
 250–253, 260
 Heller, Mark 128–130, 234, 250, 260
 Hitler, Adolf 148, 149, 151, 252
 Hume, David 79, 136, 230, 243, 258, 260
- James, William 23, 25, 27, 96, 101, 113,
 143, 144, 146, 186, 187, 203, 217,
 240, 247, 259–261
 Jones, Ward E. 241, 242, 260

- Kant, Immanuel 24, 75, 99, 138, 223–225, 228–230, 236, 257, 258, 260
- Kaplan, Mark 25, 237, 260
- Kenner, Lionel 237, 260
- Kierkegaard, Søren 104
- Kornblith, Hillary 169–173, 182, 183, 253–255, 260
- Korsgaard, Christine M. 224, 257, 260
- Kuhn, Thomas S. 209, 256, 260
- McGregor, Joan 234, 261
- Montmarquet, James A. 6, 23, 25, 27, 96, 113, 143, 144, 146–149, 187, 203, 204, 217–221, 234, 236–238, 240, 249, 252, 255, 257, 258, 261
- Neurath, Otto 256, 261
- Nietzsche, Friedrich 258, 261
- Oakley, Justin 256, 261
- Oshana, Marina A.L. 238, 241, 261
- Owens, David 19, 26–28, 79–81, 129, 130, 158, 159, 225, 236, 237, 243, 253, 257, 261
- Pappas, George S. 253, 261
- Pascal, Blaise 104, 157
- Pojman, Louis P. 105, 247, 261
- Polanyi, Michael 227, 257, 261
- Radcliffe, Dana 249, 261
- Ravizza, Mark 38, 41–45, 161, 238, 239, 253, 260, 262
- Richardson, Henry S. 257, 261
- Rorty, Amelie 237, 261
- Rorty, Richard 206, 256, 261
- Russell, Bruce 20, 50, 51, 59, 63, 64, 73–77, 124, 233, 240, 241, 243, 250, 261
- Ryan, Sharon 84–86, 104, 121, 124–128, 233, 244–247, 250, 261
- Scanlon, T.M. 37, 238, 253, 261
- Schmitt, Frederick F. 17, 18, 229, 235, 236, 241, 258, 261
- Schulhofer, Stephen J. 234, 261
- Scott-Kakures, Dion 105, 107, 110, 111, 113, 115, 116, 242, 247–249, 261
- Sellars, Wilfrid 24, 26, 34, 236–238, 261
- Smith, Holly 234, 254, 256
- Sosa, Ernest 236, 240, 242, 259
- Stalnaker, Robert 236, 261
- Steup, Matthias 34, 35, 38, 53, 100, 103, 123, 124, 136–141, 209–211, 214, 215, 217, 233, 238, 240, 245, 247, 250, 251, 254, 256, 259–262
- Stocker, Michael 125–128, 244, 250, 253, 262
- Strawson, Peter 41–43, 45, 239, 262
- Thomas III, George C. 233, 234, 260
- Tolstoy, Leo 104
- Toribio, Josefa 64, 242, 262
- Van Fraassen, Bas C. 99, 246, 262
- Wansing, Heinrich 116, 117, 246, 248, 250, 262
- Williams, Bernard 15, 18, 99, 105–113, 115, 116, 197, 199, 206, 236, 247–249, 262
- Williamson, Timothy 24, 236, 262
- Wilson, Timothy D. 242, 262
- Winters, Barbara 111, 242, 248, 249, 262
- Wolff, Robert Paul 258, 262
- Zagzebski, Linda Trinkaus 71, 72, 187, 240, 243, 250, 252, 255, 262
- Zimmermann, Michael J. 41, 48–50, 190–192, 239, 240, 253, 255, 261, 262

SUBJECT INDEX

- acceptance, the propositional attitude of 1, 20-31, 123, 189-192, 194-197, 204, 212, 220, 221, 236, 237, 250, 255, 257
- actus reus* requirement for criminal liability, the 4, 5, 10, 13, 233
- accuses 48-51, 154, 240
- autonomy 17, 18, 21, 223-231, 238, 257, 258
- belief, the propositional attitude of 1-9, 11-13, 15-20, 23-31, 33-39, 41, 44, 45, 47-49, 51, 53-77, 79, 81-85, 87-91, 93-96, 99-122, 124-136, 138-151, 153-175, 177-187, 189-192, 194, 196-198, 200, 201, 203-207, 209-221, 223, 225, 226, 229, 233-256
- blameworthiness
- epistemic 1-3, 6, 10, 11, 13-21, 23, 25, 28, 31, 33, 36-45, 47, 48, 51, 53, 55, 56, 61, 63-65, 68-70, 72-77, 79, 81-86, 90, 94-96, 99, 100, 103, 104, 112, 118-124, 129-131, 142-151, 153-155, 157-165, 167-175, 177, 178, 180-184, 186, 187, 189, 194, 196-201, 203-209, 211-215, 217-221, 226, 233, 241, 243-245, 251, 252, 255
 - moral 1-3, 5-11, 13, 16, 19, 30, 38, 39, 41, 44, 48, 49, 69, 75, 81, 91, 125, 129, 131, 149, 161, 181, 189, 190, 197, 199-201, 206, 208, 225, 240, 255
- Cogan*, the case of *Regina versus* 4, 233
- cognitive deficiency 14, 149-151, 213-217, 235, 236, 256
- culpable ignorance 8, 179, 212, 234
- cultural isolation 40, 47, 203, 209, 211-215
- deontological adequacy 82, 83, 85, 89, 91, 94-97, 99, 100, 102-104, 123, 126, 128, 130, 143, 150, 151, 153, 158, 159, 163, 165, 169, 170, 197, 204, 244, 245
- doxastic agency 20, 61, 99, 102-105, 112-114, 118-120, 122-124, 126, 131, 132, 141, 142, 144, 158, 250, 252
- doxastic control
- content-directed 90-94, 99-142, 143, 153-168, 204, 246
 - direct 28, 35, 87-94, 99-151, 158, 159, 162, 167, 169, 180, 204, 207, 216-219, 225, 226, 245, 246, 252, 254
 - indirect 104, 128, 146, 151, 153-175, 180, 204, 216-219, 243
 - long-range 93, 94, 164
 - property-directed 90, 91, 94, 95, 104, 143-152, 147-156; 169-175, 180, 204, 245, 252
- doxastic voluntarism 20, 60, 61, 87, 88, 92, 96, 99-142; 143-147, 150, 153, 167, 169-171, 187, 190, 204, 215, 218, 219, 223, 226, 227, 237, 241, 242, 245-252
- the conceptual impossibility of 105-118
 - the psychological impossibility of 118-142
- educated racist, the example of the 3, 60, 84, 102, 104, 120, 121, 147-149, 151, 160, 162, 164, 167, 169, 173, 174, 184, 187, 189, 196, 197, 200, 201, 207, 208, 220, 221, 233, 255
- epistemic Frankfurt-cases 85, 153, 161, 162, 204, 245, 252, 253, 260
- epistemic undesirability 17, 18, 34, 47, 49, 51, 53, 54, 57, 61, 62, 68-70, 72, 84, 86, 90, 91, 95, 140-142, 144, 146, 150, 151, 158, 187, 240, 244, 253
- evidence 1, 4, 15, 19, 25, 27, 50, 54, 57-64, 69-71, 75, 85, 86, 93-95, 113, 122, 124, 128, 131, 133-135, 139, 155-157, 160, 164, 180, 185, 188, 191, 200, 204, 207-213, 215-217, 219, 220, 230, 234, 235, 237, 241, 242, 246, 249, 250, 254-257

- excuses from epistemic blameworthiness 2, 9, 10, 14, 15, 19–21, 34, 40, 47, 48, 50, 58, 64, 68–71, 73–77, 81, 82, 84, 103, 104, 120, 125, 144, 145, 148, 150, 151, 156, 157, 160, 164, 167, 170, 172–175, 177, 187, 191, 193, 194, 203, 205–209, 211–217, 220, 221, 235, 236, 240, 242, 244, 254
- foresight 190–201
- freedom 99, 129, 131, 138, 140, 225, 227, 246, 253, 257
- God 24, 74, 77, 104, 121, 122, 157, 158, 164, 236, 242, 245, 249, 254
- Gordon*, the case of *State versus* 15
- honest mistakes 1, 4–7, 9–16, 30, 64, 69, 180, 194, 197–201, 206, 219–221, 235, 252
- intellectual obligation 177–188
- intention 17, 35, 88, 92, 100, 105, 108, 110, 113, 115, 116, 122, 125, 127, 128, 132, 134–136, 138–141, 242, 248, 251, 255, 166, 167
- justification, epistemic 1, 3, 8, 16–18, 24, 27, 28, 33–36, 43, 45, 53, 54, 59, 62, 65, 68, 71, 73, 121, 153, 154, 171, 178, 182, 185, 199, 209, 218, 230, 235, 236, 239–243, 254
- knowledge 11, 13, 15, 24, 27–29, 53, 54, 56, 68, 72, 80, 109, 133, 144, 174, 210, 214, 221, 236, 237, 240, 241, 248, 258, 259
- mens rea* requirement for criminal liability, the 10–14, 16, 235
- Model Penal Code (MPC), the 10–15, 235
- Morgan, the case of Director of Public Persecutions versus 3–6, 9–14, 17, 64–66, 71, 167, 180, 181, 187, 198, 206, 207, 233–235
- negligence 1, 2, 11–15, 198, 235, 236, 259
- omissions 13, 30, 82, 99, 104, 144, 145, 165, 175, 177–179, 182, 184, 186, 187, 190–195, 197, 199, 203, 205, 217, 244, 254, 255
- permission 16, 33, 35, 36, 81, 92, 238, 244
- prohibition 33, 35, 36, 238, 244
- recklessness 11, 12, 198, 235
- reliabilism 39, 40, 53, 54, 57, 58, 62–70, 76, 89, 95, 151, 162, 164, 167, 170, 171, 173, 204, 209, 210, 229, 240–242, 253
- religion 24, 25, 74, 77, 104, 121, 122, 157, 158, 164, 185, 236, 242, 245, 249, 254
- resentment 3, 5, 33, 39, 41–43, 45, 79, 80, 206, 248
- responsibility 8, 23, 33, 34, 36–38, 41, 43, 44, 66, 79, 84, 88, 94, 113, 120, 121, 123, 125, 129, 130, 136, 143–147, 153, 158, 159, 161, 171, 179, 180, 182, 183, 186, 212, 217, 219, 223, 225, 226, 229, 238, 239, 244, 245, 250, 252, 256, 257
- substantial 37
- science 91, 146, 147, 186, 209, 211, 214, 255, 256
- self-deception 7, 111, 112, 114, 118, 119, 228, 249
- ship-owner example, the 7, 8, 10, 234, 235
- truth 54, 63, 69, 259, 262
- virtue epistemology 17, 18, 34, 146–149, 186, 214, 238, 252
- Williams*, the case of *State Versus* 15, 18, 19, 197–199, 206, 236

BRIEF SUMMARY

The present study aims to provide a working analysis of the epistemic dimension of blameworthiness that may attach to an agent for holding a particular belief under particular circumstances. As such it constitutes a defence of a central tenet of epistemic deontology, as this position has been ordinarily understood in the recent literature: An agent's holding of a belief may be an appropriate object of deontic attitudes, an epistemic dimension of blame in particular.

I argue that an epistemically blameworthy belief is a belief that is undesirable from an epistemic perspective (epistemically undesirable) and for whose epistemic undesirability the believer has no adequate excuse. I maintain that a believer lacks an adequate excuse of the relevant kind if, and only if, the epistemic undesirability of her belief results from an exercise of doxastic control constituting a violation of an intellectual obligation performed with a relevant type of foresight or blameworthy inadvertence to risk. Moreover, the believer must have no adequate excuse for violating her intellectual obligation under the relevant circumstances.

In brief, I understand the key concepts invoked in this analysis thus: I take a belief to be epistemically undesirable, if it fails to satisfy one or more epistemic *desiderata* for belief: Key epistemic *desiderata* I take to be reasonableness, adequate basing, and reliable formation. An intellectual obligation I take to be an obligation to perform or omit a belief-influencing action (an exercise of doxastic control). I argue that such obligations may arise under many circumstances, most conspicuously within professional contexts. The relevant notion of foresight I take to concern an agent's acceptance that an increase in the risk of forming epistemically undesirable beliefs results from her actions or omissions. Inadvertence to risk concerns an agent's failure to perceive such an increase in risk.

The relation between the notions of doxastic control and epistemic blameworthiness is an entangled issue. I reject all versions of the prominent theory known as "doxastic voluntarism," according to which the doxastic control relevant to ascriptions of epistemic blameworthiness directly manipulates the contents of particular beliefs. Rather, I argue, the relevant modes of doxastic control are typically indirect, and command no decisive power over particular belief-contents except in exotic cases (and here the significant mode of doxastic control is always indirect). In typical cases, epistemic blameworthiness arises from an agent's transgressive exercise of her *indirect* influence on the *epistemic quality* of her beliefs. The propositional contents of her beliefs she may rarely influence to any relevant extent.

Several authors have taken blamelessness from an epistemic perspective to constitute at least one core element in the ever-controversial notion of epistemic justification. To this concern, the present analysis is of obvious relevance. However, an

adequate understanding of the conditions for epistemic blameworthiness is a matter of much broader significance. I argue that my analysis provides a method of justifying many of our ingrained reactive attitudes toward one another and may even make sense of certain intuitively compelling, but theoretically puzzling, cases of moral blameworthiness and criminal liability.

Thus, an accurate grasp of the conditions for epistemic blameworthiness may substantially inform our social interactions with other agents, not least the circumstances under which we are ready to submit them to punishments ranging from mild resentment to harsh terms of incarceration.

SYNTHESE LIBRARY

1. J. M. Bochénski, *A Precip of Mathematical Logic*. Translated from French and German by O. Bird. 1959 ISBN 90-277-0073-7
2. P. Guiraud, *Problèmes et méthodes de la statistique linguistique*. 1959 ISBN 90-277-0025-7
3. H. Freudenthal (ed.), *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences*. 1961 ISBN 90-277-0017-6
4. E. W. Beth, *Formal Methods*. An Introduction to Symbolic Logic and to the Study of Effective Operations in Arithmetic and Logic. 1962 ISBN 90-277-0069-9
5. B. H. Kazemier and D. Vuysje (eds.), *Logic and Language*. Studies dedicated to Professor Rudolf Carnap on the Occasion of His 70th Birthday. 1962 ISBN 90-277-0019-2
6. M. W. Wartofsky (ed.), *Proceedings of the Boston Colloquium for the Philosophy of Science, 1961–1962*. [Boston Studies in the Philosophy of Science, Vol. I] 1963 ISBN 90-277-0021-4
7. A. A. Zinov'ev, *Philosophical Problems of Many-valued Logic*. A revised edition, edited and translated (from Russian) by G. Küng and D.D. Comey. 1963 ISBN 90-277-0091-5
8. G. Gurvitch, *The Spectrum of Social Time*. Translated from French and edited by M. Korenbaum and P. Bosserman. 1964 ISBN 90-277-0006-0
9. P. Lorenzen, *Formal Logic*. Translated from German by F.J. Crosson. 1965 ISBN 90-277-0080-X
10. R. S. Cohen and M. W. Wartofsky (eds.), *Proceedings of the Boston Colloquium for the Philosophy of Science, 1962–1964*. In Honor of Philipp Frank. [Boston Studies in the Philosophy of Science, Vol. II] 1965 ISBN 90-277-9004-0
11. E. W. Beth, *Mathematical Thought*. An Introduction to the Philosophy of Mathematics. 1965 ISBN 90-277-0070-2
12. E. W. Beth and J. Piaget, *Mathematical Epistemology and Psychology*. Translated from French by W. Mays. 1966 ISBN 90-277-0071-0
13. G. Küng, *Ontology and the Logistic Analysis of Language*. An Enquiry into the Contemporary Views on Universals. Revised ed., translated from German. 1967 ISBN 90-277-0028-1
14. R. S. Cohen and M. W. Wartofsky (eds.), *Proceedings of the Boston Colloquium for the Philosophy of Sciences, 1964–1966*. In Memory of Norwood Russell Hanson. [Boston Studies in the Philosophy of Science, Vol. III] 1967 ISBN 90-277-0013-3
15. C. D. Broad, *Induction, Probability, and Causation*. Selected Papers. 1968 ISBN 90-277-0012-5
16. G. Patzig, *Aristotle's Theory of the Syllogism*. A Logical-philosophical Study of *Book A of the Prior Analytics*. Translated from German by J. Barnes. 1968 ISBN 90-277-0030-3
17. N. Rescher, *Topics in Philosophical Logic*. 1968 ISBN 90-277-0084-2
18. R. S. Cohen and M. W. Wartofsky (eds.), *Proceedings of the Boston Colloquium for the Philosophy of Science, 1966–1968, Part I*. [Boston Studies in the Philosophy of Science, Vol. IV] 1969 ISBN 90-277-0014-1
19. R. S. Cohen and M. W. Wartofsky (eds.), *Proceedings of the Boston Colloquium for the Philosophy of Science, 1966–1968, Part II*. [Boston Studies in the Philosophy of Science, Vol. V] 1969 ISBN 90-277-0015-X
20. J. W. Davis, D. J. Hockney and W. K. Wilson (eds.), *Philosophical Logic*. 1969 ISBN 90-277-0075-3
21. D. Davidson and J. Hintikka (eds.), *Words and Objections*. Essays on the Work of W. V. Quine. 1969, rev. ed. 1975 ISBN 90-277-0074-5; Pb 90-277-0602-6
22. P. Suppes, *Studies in the Methodology and Foundations of Science. Selected Papers from 1951 to 1969*. 1969 ISBN 90-277-0020-6
23. J. Hintikka, *Models for Modalities*. Selected Essays. 1969 ISBN 90-277-0078-8; Pb 90-277-0598-4
24. N. Rescher et al. (eds.), *Essays in Honor of Carl G. Hempel*. A Tribute on the Occasion of His 65th Birthday. 1969 ISBN 90-277-0085-0

SYNTHESE LIBRARY

25. P. V. Tavanec (ed.), *Problems of the Logic of Scientific Knowledge*. Translated from Russian. 1970
ISBN 90-277-0087-7
26. M. Swain (ed.), *Induction, Acceptance, and Rational Belief*. 1970
ISBN 90-277-0086-9
27. R. S. Cohen and R. J. Seeger (eds.), *Ernst Mach: Physicist and Philosopher*. [Boston Studies in the Philosophy of Science, Vol. VI]. 1970
ISBN 90-277-0016-8
28. J. Hintikka and P. Suppes, *Information and Inference*. 1970
ISBN 90-277-0155-5
29. K. Lambert, *Philosophical Problems in Logic*. Some Recent Developments. 1970
ISBN 90-277-0079-6
30. R. A. Eberle, *Nominalistic Systems*. 1970
ISBN 90-277-0161-X
31. P. Weingartner and G. Zecha (eds.), *Induction, Physics, and Ethics*. 1970
ISBN 90-277-0158-X
32. E. W. Beth, *Aspects of Modern Logic*. Translated from Dutch. 1970
ISBN 90-277-0173-3
33. R. Hilpinen (ed.), *Deontic Logic*. Introductory and Systematic Readings. 1971 *See also* No. 152.
ISBN Pb (1981 rev.) 90-277-1302-2
34. J.-L. Krivine, *Introduction to Axiomatic Set Theory*. Translated from French. 1971
ISBN 90-277-0169-5; Pb 90-277-0411-2
35. J. D. Sneed, *The Logical Structure of Mathematical Physics*. 2nd rev. ed., 1979
ISBN 90-277-1056-2; Pb 90-277-1059-7
36. C. R. Kordig, *The Justification of Scientific Change*. 1971
ISBN 90-277-0181-4; Pb 90-277-0475-9
37. M. Čapek, *Bergson and Modern Physics*. A Reinterpretation and Re-evaluation. [Boston Studies in the Philosophy of Science, Vol. VII] 1971
ISBN 90-277-0186-5
38. N. R. Hanson, *What I Do Not Believe, and Other Essays*. Ed. by S. Toulmin and H. Woolf. 1971
ISBN 90-277-0191-1
39. R. C. Buck and R. S. Cohen (eds.), *PSA 1970*. Proceedings of the Second Biennial Meeting of the Philosophy of Science Association, Boston, Fall 1970. In Memory of Rudolf Carnap. [Boston Studies in the Philosophy of Science, Vol. VIII] 1971
ISBN 90-277-0187-3; Pb 90-277-0309-4
40. D. Davidson and G. Harman (eds.), *Semantics of Natural Language*. 1972
ISBN 90-277-0304-3; Pb 90-277-0310-8
41. Y. Bar-Hillel (ed.), *Pragmatics of Natural Languages*. 1971
ISBN 90-277-0194-6; Pb 90-277-0599-2
42. S. Stenlund, *Combinators, γ Terms and Proof Theory*. 1972
ISBN 90-277-0305-1
43. M. Strauss, *Modern Physics and Its Philosophy*. Selected Paper in the Logic, History, and Philosophy of Science. 1972
ISBN 90-277-0230-6
44. M. Bunge, *Method, Model and Matter*. 1973
ISBN 90-277-0252-7
45. M. Bunge, *Philosophy of Physics*. 1973
ISBN 90-277-0253-5
46. A. A. Zinov'ev, *Foundations of the Logical Theory of Scientific Knowledge (Complex Logic)*. Revised and enlarged English edition with an appendix by G. A. Smirnov, E. A. Sidorenka, A. M. Fedina and L. A. Bobrova. [Boston Studies in the Philosophy of Science, Vol. IX] 1973
ISBN 90-277-0193-8; Pb 90-277-0324-8
47. L. Tondl, *Scientific Procedures*. A Contribution concerning the Methodological Problems of Scientific Concepts and Scientific Explanation. Translated from Czech by D. Short. Edited by R.S. Cohen and M.W. Wartofsky. [Boston Studies in the Philosophy of Science, Vol. X] 1973
ISBN 90-277-0147-4; Pb 90-277-0323-X
48. N. R. Hanson, *Constellations and Conjectures*. 1973
ISBN 90-277-0192-X
49. K. J. J. Hintikka, J. M. E. Moravcsik and P. Suppes (eds.), *Approaches to Natural Language*. 1973
ISBN 90-277-0220-9; Pb 90-277-0233-0
50. M. Bunge (ed.), *Exact Philosophy*. Problems, Tools and Goals. 1973
ISBN 90-277-0251-9
51. R. J. Bogdan and I. Niiniluoto (eds.), *Logic, Language and Probability*. 1973
ISBN 90-277-0312-4

SYNTHESE LIBRARY

52. G. Pearce and P. Maynard (eds.), *Conceptual Change*. 1973
ISBN 90-277-0287-X; Pb 90-277-0339-6
53. I. Niiniluoto and R. Tuomela, *Theoretical Concepts and Hypothetico-inductive Inference*. 1973
ISBN 90-277-0343-4
54. R. Fraïssé, *Course of Mathematical Logic – Volume 1: Relation and Logical Formula*. Translated from French. 1973
ISBN 90-277-0268-3; Pb 90-277-0403-1
(For *Volume 2* see under No. 69).
55. A. Grünbaum, *Philosophical Problems of Space and Time*. Edited by R.S. Cohen and M.W. Wartofsky. 2nd enlarged ed. [Boston Studies in the Philosophy of Science, Vol. XII] 1973
ISBN 90-277-0357-4; Pb 90-277-0358-2
56. P. Suppes (ed.), *Space, Time and Geometry*. 1973
ISBN 90-277-0386-8; Pb 90-277-0442-2
57. H. Kelsen, *Essays in Legal and Moral Philosophy*. Selected and introduced by O. Weinberger. Translated from German by P. Heath. 1973
ISBN 90-277-0388-4
58. R. J. Seeger and R. S. Cohen (eds.), *Philosophical Foundations of Science*. [Boston Studies in the Philosophy of Science, Vol. XI] 1974
ISBN 90-277-0390-6; Pb 90-277-0376-0
59. R. S. Cohen and M.W. Wartofsky (eds.), *Logical and Epistemological Studies in Contemporary Physics*. [Boston Studies in the Philosophy of Science, Vol. XIII] 1973
ISBN 90-277-0391-4; Pb 90-277-0377-9
60. R. S. Cohen and M. W. Wartofsky (eds.), *Methodological and Historical Essays in the Natural and Social Sciences. Proceedings of the Boston Colloquium for the Philosophy of Science, 1969–1972*. [Boston Studies in the Philosophy of Science, Vol. XIV] 1974
ISBN 90-277-0392-2; Pb 90-277-0378-7
61. R. S. Cohen, J. J. Stachel and M. W. Wartofsky (eds.), *For Dirk Struik. Scientific, Historical and Political Essays*. [Boston Studies in the Philosophy of Science, Vol. XV] 1974
ISBN 90-277-0393-0; Pb 90-277-0379-5
62. K. Ajdukiewicz, *Pragmatic Logic*. Translated from Polish by O. Wojtasiewicz. 1974
ISBN 90-277-0326-4
63. S. Stenlund (ed.), *Logical Theory and Semantic Analysis*. Essays dedicated to Stig Kanger on His 50th Birthday. 1974
ISBN 90-277-0438-4
64. K. F. Schaffner and R. S. Cohen (eds.), *PSA 1972. Proceedings of the Third Biennial Meeting of the Philosophy of Science Association*. [Boston Studies in the Philosophy of Science, Vol. XX] 1974
ISBN 90-277-0408-2; Pb 90-277-0409-0
65. H. E. Kyburg, Jr., *The Logical Foundations of Statistical Inference*. 1974
ISBN 90-277-0330-2; Pb 90-277-0430-9
66. M. Grene, *The Understanding of Nature*. Essays in the Philosophy of Biology. [Boston Studies in the Philosophy of Science, Vol. XXIII] 1974
ISBN 90-277-0462-7; Pb 90-277-0463-5
67. J. M. Broekman, *Structuralism: Moscow, Prague, Paris*. Translated from German. 1974
ISBN 90-277-0478-3
68. N. Geschwind, *Selected Papers on Language and the Brain*. [Boston Studies in the Philosophy of Science, Vol. XVI] 1974
ISBN 90-277-0262-4; Pb 90-277-0263-2
69. R. Fraïssé, *Course of Mathematical Logic – Volume 2: Model Theory*. Translated from French. 1974
(For *Volume 1* see under No. 54)
ISBN 90-277-0269-1; Pb 90-277-0510-0
70. A. Grzegorzcyk, *An Outline of Mathematical Logic*. Fundamental Results and Notions explained with all Details. Translated from Polish. 1974
ISBN 90-277-0359-0; Pb 90-277-0447-3
71. F. von Kutschera, *Philosophy of Language*. 1975
ISBN 90-277-0591-7
72. J. Manninen and R. Tuomela (eds.), *Essays on Explanation and Understanding*. Studies in the Foundations of Humanities and Social Sciences. 1976
ISBN 90-277-0592-5
73. J. Hintikka (ed.), *Rudolf Carnap, Logical Empiricist*. Materials and Perspectives. 1975
ISBN 90-277-0583-6

SYNTHESE LIBRARY

74. M. Čapek (ed.), *The Concepts of Space and Time. Their Structure and Their Development.* [Boston Studies in the Philosophy of Science, Vol. XXII] 1976 ISBN 90-277-0355-8; Pb 90-277-0375-2
75. J. Hintikka and U. Remes, *The Method of Analysis. Its Geometrical Origin and Its General Significance.* [Boston Studies in the Philosophy of Science, Vol. XXV] 1974
ISBN 90-277-0532-1; Pb 90-277-0543-7
76. J. E. Murdoch and E. D. Sylla (eds.), *The Cultural Context of Medieval Learning.* [Boston Studies in the Philosophy of Science, Vol. XXVI] 1975 ISBN 90-277-0560-7; Pb 90-277-0587-9
77. S. Amsterdamski, *Between Experience and Metaphysics.* Philosophical Problems of the Evolution of Science. [Boston Studies in the Philosophy of Science, Vol. XXXV] 1975
ISBN 90-277-0568-2; Pb 90-277-0580-1
78. P. Suppes (ed.), *Logic and Probability in Quantum Mechanics.* 1976
ISBN 90-277-0570-4; Pb 90-277-1200-X
79. H. von Helmholtz: *Epistemological Writings. The Paul Hertz / Moritz Schlick Centenary Edition of 1921 with Notes and Commentary by the Editors.* Newly translated from German by M. F. Lowe. Edited, with an Introduction and Bibliography, by R. S. Cohen and Y. Elkana. [Boston Studies in the Philosophy of Science, Vol. XXXVII] 1975 ISBN 90-277-0290-X; Pb 90-277-0582-8
80. J. Agassi, *Science in Flux.* [Boston Studies in the Philosophy of Science, Vol. XXVIII] 1975
ISBN 90-277-0584-4; Pb 90-277-0612-2
81. S. G. Harding (ed.), *Can Theories Be Refuted? Essays on the Duhem-Quine Thesis.* 1976
ISBN 90-277-0629-8; Pb 90-277-0630-1
82. S. Nowak, *Methodology of Sociological Research.* General Problems. 1977
ISBN 90-277-0486-4
83. J. Piaget, J.-B. Grize, A. Szeminńska and V. Bang, *Epistemology and Psychology of Functions.* Translated from French. 1977 ISBN 90-277-0804-5
84. M. Grene and E. Mendelsohn (eds.), *Topics in the Philosophy of Biology.* [Boston Studies in the Philosophy of Science, Vol. XXVII] 1976 ISBN 90-277-0595-X; Pb 90-277-0596-8
85. E. Fischbein, *The Intuitive Sources of Probabilistic Thinking in Children.* 1975
ISBN 90-277-0626-3; Pb 90-277-1190-9
86. E. W. Adams, *The Logic of Conditionals.* An Application of Probability to Deductive Logic. 1975
ISBN 90-277-0631-X
87. M. Przełęcki and R. Wójcicki (eds.), *Twenty-Five Years of Logical Methodology in Poland.* Translated from Polish. 1976 ISBN 90-277-0601-8
88. J. Popolski, *The Methodology of History.* Translated from Polish by O. Wojtasiewicz. 1976
ISBN 90-277-0550-X
89. A. Kasher (ed.), *Language in Focus: Foundations, Methods and Systems.* Essays dedicated to Yehoshua Bar-Hillel. [Boston Studies in the Philosophy of Science, Vol. XLIII] 1976
ISBN 90-277-0644-1; Pb 90-277-0645-X
90. J. Hintikka, *The Intentions of Intentionality and Other New Models for Modalities.* 1975
ISBN 90-277-0633-6; Pb 90-277-0634-4
91. W. Stegmüller, *Collected Papers on Epistemology, Philosophy of Science and History of Philosophy.* 2 Volumes. 1977 Set ISBN 90-277-0767-7
92. D. M. Gabbay, *Investigations in Modal and Tense Logics with Applications to Problems in Philosophy and Linguistics.* 1976 ISBN 90-277-0656-5
93. R. J. Bogdan, *Local Induction.* 1976 ISBN 90-277-0649-2
94. S. Nowak, *Understanding and Prediction.* Essays in the Methodology of Social and Behavioral Theories. 1976 ISBN 90-277-0558-5; Pb 90-277-1199-2
95. P. Mittelstaedt, *Philosophical Problems of Modern Physics.* [Boston Studies in the Philosophy of Science, Vol. XVIII] 1976 ISBN 90-277-0285-3; Pb 90-277-0506-2

SYNTHESE LIBRARY

96. G. Holton and W. A. Blanpied (eds.), *Science and Its Public: The Changing Relationship*. [Boston Studies in the Philosophy of Science, Vol. XXXIII] 1976 ISBN 90-277-0657-3; Pb 90-277-0658-1
97. M. Brand and D. Walton (eds.), *Action Theory*. 1976 ISBN 90-277-0671-9
98. P. Gochet, *Outline of a Nominalist Theory of Propositions*. An Essay in the Theory of Meaning and in the Philosophy of Logic. 1980 ISBN 90-277-1031-7
99. R. S. Cohen, P. K. Feyerabend, and M.W.Wartofsky (eds.), *Essays in Memory of Imre Lakatos*. [Boston Studies in the Philosophy of Science, Vol. XXXIX] 1976
ISBN 90-277-0654-9; Pb 90-277-0655-7
100. R. S. Cohen and J. J. Stachel (eds.), *Selected Papers of Léon Rosenfield*. [Boston Studies in the Philosophy of Science, Vol. XXI] 1979 ISBN 90-277-0651-4; Pb 90-277-0652-2
101. R. S. Cohen, C. A. Hooker, A. C. Michalos and J. W. van Evra (eds.), *PSA 1974. Proceedings of the 1974 Biennial Meeting of the Philosophy of Science Association*. [Boston Studies in the Philosophy of Science, Vol. XXXII] 1976 ISBN 90-277-0647-6; Pb 90-277-0648-4
102. Y. Fried and J. Agassi, *Paranoia*. A Study in Diagnosis. [Boston Studies in the Philosophy of Science, Vol. L] 1976 ISBN 90-277-0704-9; Pb 90-277-0705-7
103. M. Przełęcki, K. Szaniawski and R. Wójcicki (eds.), *Formal Methods in the Methodology of Empirical Sciences*. 1976 ISBN 90-277-0698-0
104. J. M. Vickers, *Belief and Probability*. 1976 ISBN 90-277-0744-8
105. K. H. Wolff, *Surrender and Catch*. Experience and Inquiry Today. [Boston Studies in the Philosophy of Science, Vol. LI] 1976 ISBN 90-277-0758-8; Pb 90-277-0765-0
106. K. Kosik, *Dialectics of the Concrete*. A Study on Problems of Man and World. [Boston Studies in the Philosophy of Science, Vol. LII] 1976 ISBN 90-277-0761-8; Pb 90-277-0764-2
107. N. Goodman, *The Structure of Appearance*. 3rd ed. with an Introduction by G. Hellman. [Boston Studies in the Philosophy of Science, Vol. LIII] 1977 ISBN 90-277-0773-1; Pb 90-277-0774-X
108. K. Ajdukiewicz, *The Scientific World-Perspective and Other Essays, 1931–1963*. Translated from Polish. Edited and with an Introduction by J. Giedymin. 1978 ISBN 90-277-0527-5
109. R. L. Causey, *Unity of Science*. 1977 ISBN 90-277-0779-0
110. R. E. Grandy, *Advanced Logic for Applications*. 1977 ISBN 90-277-0781-2
111. R. P. McArthur, *Tense Logic*. 1976 ISBN 90-277-0697-2
112. L. Lindahl, *Position and Change*. A Study in Law and Logic. Translated from Swedish by P. Needham. 1977 ISBN 90-277-0787-1
113. R. Tuomela, *Dispositions*. 1978 ISBN 90-277-0810-X
114. H. A. Simon, *Models of Discovery and Other Topics in the Methods of Science*. [Boston Studies in the Philosophy of Science, Vol. LIV] 1977 ISBN 90-277-0812-6; Pb 90-277-0858-4
115. R. D. Rosenkrantz, *Inference, Method and Decision*. Towards a Bayesian Philosophy of Science. 1977 ISBN 90-277-0817-7; Pb 90-277-0818-5
116. R. Tuomela, *Human Action and Its Explanation*. A Study on the Philosophical Foundations of Psychology. 1977 ISBN 90-277-0824-X
117. M. Lazerowitz, *The Language of Philosophy*. Freud and Wittgenstein. [Boston Studies in the Philosophy of Science, Vol. LV] 1977 ISBN 90-277-0826-6; Pb 90-277-0862-2
118. Not published.
119. J. Pelc (ed.), *Semiotics in Poland, 1894–1969*. Translated from Polish. 1979
ISBN 90-277-0811-8
120. I. Pörn, *Action Theory and Social Science*. Some Formal Models. 1977 ISBN 90-277-0846-0
121. J. Margolis, *Persons and Mind*. The Prospects of Nonreductive Materialism. [Boston Studies in the Philosophy of Science, Vol. LVII] 1977 ISBN 90-277-0854-1; Pb 90-277-0863-0
122. J. Hintikka, I. Niiniluoto, and E. Saarinen (eds.), *Essays on Mathematical and Philosophical Logic*. 1979 ISBN 90-277-0879-7

SYNTHESE LIBRARY

123. T. A. F. Kuipers, *Studies in Inductive Probability and Rational Expectation*. 1978
ISBN 90-277-0882-7
124. E. Saarinen, R. Hilpinen, I. Niiniluoto and M. P. Hintikka (eds.), *Essays in Honour of Jaakko Hintikka on the Occasion of His 50th Birthday*. 1979
ISBN 90-277-0916-5
125. G. Radnitzky and G. Andersson (eds.), *Progress and Rationality in Science*. [Boston Studies in the Philosophy of Science, Vol. LVIII] 1978
ISBN 90-277-0921-1; Pb 90-277-0922-X
126. P. Mittelstaedt, *Quantum Logic*. 1978
ISBN 90-277-0925-4
127. K. A. Bowen, *Model Theory for Modal Logic*. Kripke Models for Modal Predicate Calculi. 1979
ISBN 90-277-0929-7
128. H. A. Bursen, *Dismantling the Memory Machine*. A Philosophical Investigation of Machine Theories of Memory. 1978
ISBN 90-277-0933-5
129. M. W. Wartofsky, *Models*. Representation and the Scientific Understanding. [Boston Studies in the Philosophy of Science, Vol. XLVIII] 1979
ISBN 90-277-0736-7; Pb 90-277-0947-5
130. D. Ihde, *Technics and Praxis*. A Philosophy of Technology. [Boston Studies in the Philosophy of Science, Vol. XXIV] 1979
ISBN 90-277-0953-X; Pb 90-277-0954-8
131. J. J. Wiatr (ed.), *Polish Essays in the Methodology of the Social Sciences*. [Boston Studies in the Philosophy of Science, Vol. XXIX] 1979
ISBN 90-277-0723-5; Pb 90-277-0956-4
132. W. C. Salmon (ed.), *Hans Reichenbach: Logical Empiricist*. 1979
ISBN 90-277-0958-0
133. P. Bieri, R.-P. Horstmann and L. Krüger (eds.), *Transcendental Arguments in Science*. Essays in Epistemology. 1979
ISBN 90-277-0963-7; Pb 90-277-0964-5
134. M. Marković and G. Petrović (eds.), *Praxis*. Yugoslav Essays in the Philosophy and Methodology of the Social Sciences. [Boston Studies in the Philosophy of Science, Vol. XXXVI] 1979
ISBN 90-277-0727-8; Pb 90-277-0968-8
135. R. Wójcicki, *Topics in the Formal Methodology of Empirical Sciences*. Translated from Polish. 1979
ISBN 90-277-1004-X
136. G. Radnitzky and G. Andersson (eds.), *The Structure and Development of Science*. [Boston Studies in the Philosophy of Science, Vol. LIX] 1979
ISBN 90-277-0994-7; Pb 90-277-0995-5
137. J. C. Webb, *Mechanism, Mentalism and Metamathematics*. An Essay on Finitism. 1980
ISBN 90-277-1046-5
138. D. F. Gustafson and B. L. Tapscott (eds.), *Body, Mind and Method*. Essays in Honor of Virgil C. Aldrich. 1979
ISBN 90-277-1013-9
139. L. Nowak, *The Structure of Idealization*. Towards a Systematic Interpretation of the Marxian Idea of Science. 1980
ISBN 90-277-1014-7
140. C. Perelman, *The New Rhetoric and the Humanities*. Essays on Rhetoric and Its Applications. Translated from French and German. With an Introduction by H. Zyskind. 1979
ISBN 90-277-1018-X; Pb 90-277-1019-8
141. W. Rabinowicz, *Universalizability*. A Study in Morals and Metaphysics. 1979
ISBN 90-277-1020-2
142. C. Perelman, *Justice, Law and Argument*. Essays on Moral and Legal Reasoning. Translated from French and German. With an Introduction by H.J. Berman. 1980
ISBN 90-277-1089-9; Pb 90-277-1090-2
143. S. Kanger and S. Öhman (eds.), *Philosophy and Grammar*. Papers on the Occasion of the Quincentennial of Uppsala University. 1981
ISBN 90-277-1091-0
144. T. Pawlowski, *Concept Formation in the Humanities and the Social Sciences*. 1980
ISBN 90-277-1096-1
145. J. Hintikka, D. Gruender and E. Agazzi (eds.), *Theory Change, Ancient Axiomatics and Galileo's Methodology*. Proceedings of the 1978 Pisa Conference on the History and Philosophy of Science, Volume I. 1981
ISBN 90-277-1126-7

SYNTHESE LIBRARY

146. J. Hintikka, D. Gruender and E. Agazzi (eds.), *Probabilistic Thinking, Thermodynamics, and the Interaction of the History and Philosophy of Science*. Proceedings of the 1978 Pisa Conference on the History and Philosophy of Science, Volume II. 1981 ISBN 90-277-1127-5
147. U. Mönnich (ed.), *Aspects of Philosophical Logic*. Some Logical Forays into Central Notions of Linguistics and Philosophy. 1981 ISBN 90-277-1201-8
148. D. M. Gabbay, *Semantical Investigations in Heyting's Intuitionistic Logic*. 1981 ISBN 90-277-1202-6
149. E. Agazzi (ed.), *Modern Logic – A Survey*. Historical, Philosophical, and Mathematical Aspects of Modern Logic and Its Applications. 1981 ISBN 90-277-1137-2
150. A. F. Parker-Rhodes, *The Theory of Indistinguishables*. A Search for Explanatory Principles below the Level of Physics. 1981 ISBN 90-277-1214-X
151. J. C. Pitt, *Pictures, Images, and Conceptual Change*. An Analysis of Wilfrid Sellars' Philosophy of Science. 1981 ISBN 90-277-1276-X; Pb 90-277-1277-8
152. R. Hilpinen (ed.), *New Studies in Deontic Logic*. Norms, Actions, and the Foundations of Ethics. 1981 ISBN 90-277-1278-6; Pb 90-277-1346-4
153. C. Dilworth, *Scientific Progress*. A Study Concerning the Nature of the Relation between Successive Scientific Theories. 3rd rev. ed., 1994 ISBN 0-7923-2487-0; Pb 0-7923-2488-9
154. D. Woodruff Smith and R. McIntyre, *Husserl and Intentionality*. A Study of Mind, Meaning, and Language. 1982 ISBN 90-277-1392-8; Pb 90-277-1730-3
155. R. J. Nelson, *The Logic of Mind*. 2nd. ed., 1989 ISBN 90-277-2819-4; Pb 90-277-2822-4
156. J. F. A. K. van Benthem, *The Logic of Time*. A Model-Theoretic Investigation into the Varieties of Temporal Ontology, and Temporal Discourse. 1983; 2nd ed., 1991 ISBN 0-7923-1081-0
157. R. Swinburne (ed.), *Space, Time and Causality*. 1983 ISBN 90-277-1437-1
158. E. T. Jaynes, *Papers on Probability, Statistics and Statistical Physics*. Ed. by R. D. Rozenkrantz. 1983 ISBN 90-277-1448-7; Pb (1989) 0-7923-0213-3
159. T. Chapman, *Time: A Philosophical Analysis*. 1982 ISBN 90-277-1465-7
160. E. N. Zalta, *Abstract Objects*. An Introduction to Axiomatic Metaphysics. 1983 ISBN 90-277-1474-6
161. S. Harding and M. B. Hintikka (eds.), *Discovering Reality*. Feminist Perspectives on Epistemology, Metaphysics, Methodology, and Philosophy of Science. 1983 ISBN 90-277-1496-7; Pb 90-277-1538-6
162. M. A. Stewart (ed.), *Law, Morality and Rights*. 1983 ISBN 90-277-1519-X
163. D. Mayr and G. Süssmann (eds.), *Space, Time, and Mechanics*. Basic Structures of a Physical Theory. 1983 ISBN 90-277-1525-4
164. D. Gabbay and F. Guentner (eds.), *Handbook of Philosophical Logic*. Vol. I: Elements of Classical Logic. 1983 ISBN 90-277-1542-4
165. D. Gabbay and F. Guentner (eds.), *Handbook of Philosophical Logic*. Vol. II: Extensions of Classical Logic. 1984 ISBN 90-277-1604-8
166. D. Gabbay and F. Guentner (eds.), *Handbook of Philosophical Logic*. Vol. III: Alternative to Classical Logic. 1986 ISBN 90-277-1605-6
167. D. Gabbay and F. Guentner (eds.), *Handbook of Philosophical Logic*. Vol. IV: Topics in the Philosophy of Language. 1989 ISBN 90-277-1606-4
168. A. J. I. Jones, *Communication and Meaning*. An Essay in Applied Modal Logic. 1983 ISBN 90-277-1543-2
169. M. Fitting, *Proof Methods for Modal and Intuitionistic Logics*. 1983 ISBN 90-277-1573-4
170. J. Margolis, *Culture and Cultural Entities*. Toward a New Unity of Science. 1984 ISBN 90-277-1574-2
171. R. Tuomela, *A Theory of Social Action*. 1984 ISBN 90-277-1703-6

SYNTHESE LIBRARY

172. J. J. E. Gracia, E. Rabossi, E. Villanueva and M. Dascal (eds.), *Philosophical Analysis in Latin America*. 1984 ISBN 90-277-1749-4
173. P. Ziff, *Epistemic Analysis. A Coherence Theory of Knowledge*. 1984 ISBN 90-277-1751-7
174. P. Ziff, *Antiaesthetics. An Appreciation of the Cow with the Subtile Nose*. 1984 ISBN 90-277-1773-7
175. W. Balzer, D. A. Pearce, and H.-J. Schmidt (eds.), *Reduction in Science. Structure, Examples, Philosophical Problems*. 1984 ISBN 90-277-1811-3
176. A. Peczenik, L. Lindahl and B. van Roermund (eds.), *Theory of Legal Science. Proceedings of the Conference on Legal Theory and Philosophy of Science (Lund, Sweden, December 1983)*. 1984 ISBN 90-277-1834-2
177. I. Niiniluoto, *Is Science Progressive?* 1984 ISBN 90-277-1835-0
178. B. K. Matilal and J. L. Shaw (eds.), *Analytical Philosophy in Comparative Perspective. Exploratory Essays in Current Theories and Classical Indian Theories of Meaning and Reference*. 1985 ISBN 90-277-1870-9
179. P. Kroes, *Time: Its Structure and Role in Physical Theories*. 1985 ISBN 90-277-1894-6
180. J. H. Fetzer, *Sociobiology and Epistemology*. 1985 ISBN 90-277-2005-3; Pb 90-277-2006-1
181. L. Haaparanta and J. Hintikka (eds.), *Frege Synthesized. Essays on the Philosophical and Foundational Work of Gottlob Frege*. 1986 ISBN 90-277-2126-2
182. M. Detlefsen, *Hilbert's Program. An Essay on Mathematical Instrumentalism*. 1986 ISBN 90-277-2151-3
183. J. L. Golden and J. J. Pilotta (eds.), *Practical Reasoning in Human Affairs. Studies in Honor of Chaim Perelman*. 1986 ISBN 90-277-2255-2
184. H. Zandvoort, *Models of Scientific Development and the Case of Nuclear Magnetic Resonance*. 1986 ISBN 90-277-2351-6
185. I. Niiniluoto, *Truthlikeness*. 1987 ISBN 90-277-2354-0
186. W. Balzer, C. U. Moulines and J. D. Sneed, *An Architectonic for Science. The Structuralist Program*. 1987 ISBN 90-277-2403-2
187. D. Pearce, *Roads to Commensurability*. 1987 ISBN 90-277-2414-8
188. L. M. Vaina (ed.), *Matters of Intelligence. Conceptual Structures in Cognitive Neuroscience*. 1987 ISBN 90-277-2460-1
189. H. Siegel, *Relativism Refuted. A Critique of Contemporary Epistemological Relativism*. 1987 ISBN 90-277-2469-5
190. W. Callebaut and R. Pinxten, *Evolutionary Epistemology. A Multiparadigm Program, with a Complete Evolutionary Epistemology Bibliograph*. 1987 ISBN 90-277-2582-9
191. J. Kmita, *Problems in Historical Epistemology*. 1988 ISBN 90-277-2199-8
192. J. H. Fetzer (ed.), *Probability and Causality. Essays in Honor of Wesley C. Salmon, with an Annotated Bibliography*. 1988 ISBN 90-277-2607-8; Pb 1-5560-8052-2
193. A. Donovan, L. Laudan and R. Laudan (eds.), *Scrutinizing Science. Empirical Studies of Scientific Change*. 1988 ISBN 90-277-2608-6
194. H.R. Otto and J.A. Tuedio (eds.), *Perspectives on Mind*. 1988 ISBN 90-277-2640-X
195. D. Batens and J.P. van Bendegem (eds.), *Theory and Experiment. Recent Insights and New Perspectives on Their Relation*. 1988 ISBN 90-277-2645-0
196. J. Österberg, *Self and Others. A Study of Ethical Egoism*. 1988 ISBN 90-277-2648-5
197. D.H. Helman (ed.), *Analogical Reasoning. Perspectives of Artificial Intelligence, Cognitive Science, and Philosophy*. 1988 ISBN 90-277-2711-2
198. J. Wolenski, *Logic and Philosophy in the Lvov-Warsaw School*. 1989 ISBN 90-277-2749-X
199. R. Wójcicki, *Theory of Logical Calculi. Basic Theory of Consequence Operations*. 1988 ISBN 90-277-2785-6

SYNTHESE LIBRARY

200. J. Hintikka and M.B. Hintikka, *The Logic of Epistemology and the Epistemology of Logic*. Selected Essays. 1989 ISBN 0-7923-0040-8; Pb 0-7923-0041-6
201. E. Agazzi (ed.), *Probability in the Sciences*. 1988 ISBN 90-277-2808-9
202. M. Meyer (ed.), *From Metaphysics to Rhetoric*. 1989 ISBN 90-277-2814-3
203. R.L. Tieszen, *Mathematical Intuition*. Phenomenology and Mathematical Knowledge. 1989 ISBN 0-7923-0131-5
204. A. Melnick, *Space, Time, and Thought in Kant*. 1989 ISBN 0-7923-0135-8
205. D.W. Smith, *The Circle of Acquaintance*. Perception, Consciousness, and Empathy. 1989 ISBN 0-7923-0252-4
206. M.H. Salmon (ed.), *The Philosophy of Logical Mechanism*. Essays in Honor of Arthur W. Burks. With his Responses, and with a Bibliography of Burk's Work. 1990 ISBN 0-7923-0325-3
207. M. Kusch, *Language as Calculus vs. Language as Universal Medium*. A Study in Husserl, Heidegger, and Gadamer. 1989 ISBN 0-7923-0333-4
208. T.C. Meyering, *Historical Roots of Cognitive Science*. The Rise of a Cognitive Theory of Perception from Antiquity to the Nineteenth Century. 1989 ISBN 0-7923-0349-0
209. P. Kosso, *Observability and Observation in Physical Science*. 1989 ISBN 0-7923-0389-X
210. J. Kmita, *Essays on the Theory of Scientific Cognition*. 1990 ISBN 0-7923-0441-1
211. W. Sieg (ed.), *Acting and Reflecting*. The Interdisciplinary Turn in Philosophy. 1990 ISBN 0-7923-0512-4
212. J. Karpinński, *Causality in Sociological Research*. 1990 ISBN 0-7923-0546-9
213. H.A. Lewis (ed.), *Peter Geach: Philosophical Encounters*. 1991 ISBN 0-7923-0823-9
214. M. Ter Hark, *Beyond the Inner and the Outer*. Wittgenstein's Philosophy of Psychology. 1990 ISBN 0-7923-0850-6
215. M. Gosselin, *Nominalism and Contemporary Nominalism*. Ontological and Epistemological Implications of the Work of W.V.O. Quine and of N. Goodman. 1990 ISBN 0-7923-0904-9
216. J.H. Fetzer, D. Shatz and G. Schlesinger (eds.), *Definitions and Definability*. Philosophical Perspectives. 1991 ISBN 0-7923-1046-2
217. E. Agazzi and A. Cordero (eds.), *Philosophy and the Origin and Evolution of the Universe*. 1991 ISBN 0-7923-1322-4
218. M. Kusch, *Foucault's Strata and Fields*. An Investigation into Archaeological and Genealogical Science Studies. 1991 ISBN 0-7923-1462-X
219. C.J. Posy, *Kant's Philosophy of Mathematics*. Modern Essays. 1992 ISBN 0-7923-1495-6
220. G. Van de Vijver, *New Perspectives on Cybernetics*. Self-Organization, Autonomy and Connectionism. 1992 ISBN 0-7923-1519-7
221. J.C. Nyíri, *Tradition and Individuality*. Essays. 1992 ISBN 0-7923-1566-9
222. R. Howell, *Kant's Transcendental Deduction*. An Analysis of Main Themes in His Critical Philosophy. 1992 ISBN 0-7923-1571-5
223. A. García de la Sienra, *The Logical Foundations of the Marxian Theory of Value*. 1992 ISBN 0-7923-1778-5
224. D.S. Shwayder, *Statement and Referent*. An Inquiry into the Foundations of Our Conceptual Order. 1992 ISBN 0-7923-1803-X
225. M. Rosen, *Problems of the Hegelian Dialectic*. Dialectic Reconstructed as a Logic of Human Reality. 1993 ISBN 0-7923-2047-6
226. P. Suppes, *Models and Methods in the Philosophy of Science: Selected Essays*. 1993 ISBN 0-7923-2211-8
227. R. M. Dancy (ed.), *Kant and Critique: New Essays in Honor of W. H. Werkmeister*. 1993 ISBN 0-7923-2244-4
228. J. Woleński (ed.), *Philosophical Logic in Poland*. 1993 ISBN 0-7923-2293-2

SYNTHESE LIBRARY

229. M. De Rijke (ed.), *Diamonds and Defaults*. Studies in Pure and Applied Intensional Logic. 1993
ISBN 0-7923-2342-4
230. B.K. Matilal and A. Chakrabarti (eds.), *Knowing from Words*. Western and Indian Philosophical Analysis of Understanding and Testimony. 1994
ISBN 0-7923-2345-9
231. S.A. Kleiner, *The Logic of Discovery*. A Theory of the Rationality of Scientific Research. 1993
ISBN 0-7923-2371-8
232. R. Festa, *Optimum Inductive Methods*. A Study in Inductive Probability, Bayesian Statistics, and Verisimilitude. 1993
ISBN 0-7923-2460-9
233. P. Humphreys (ed.), *Patrick Suppes: Scientific Philosopher*. Vol. 1: Probability and Probabilistic Causality. 1994
ISBN 0-7923-2552-4
234. P. Humphreys (ed.), *Patrick Suppes: Scientific Philosopher*. Vol. 2: Philosophy of Physics, Theory Structure, and Measurement Theory. 1994
ISBN 0-7923-2553-2
235. P. Humphreys (ed.), *Patrick Suppes: Scientific Philosopher*. Vol. 3: Language, Logic, and Psychology. 1994
ISBN 0-7923-2862-0
Set ISBN (Vols 233–235) 0-7923-2554-0
236. D. Prawitz and D. Westerståhl (eds.), *Logic and Philosophy of Science in Uppsala*. Papers from the 9th International Congress of Logic, Methodology, and Philosophy of Science. 1994
ISBN 0-7923-2702-0
237. L. Haaparanta (ed.), *Mind, Meaning and Mathematics*. Essays on the Philosophical Views of Husserl and Frege. 1994
ISBN 0-7923-2703-9
238. J. Hintikka (ed.), *Aspects of Metaphor*. 1994
ISBN 0-7923-2786-1
239. B. McGuinness and G. Oliveri (eds.), *The Philosophy of Michael Dummett*. With Replies from Michael Dummett. 1994
ISBN 0-7923-2804-3
240. D. Jamieson (ed.), *Language, Mind, and Art*. Essays in Appreciation and Analysis, In Honor of Paul Ziff. 1994
ISBN 0-7923-2810-8
241. G. Preyer, F. Siebelt and A. Ulfig (eds.), *Language, Mind and Epistemology*. On Donald Davidson's Philosophy. 1994
ISBN 0-7923-2811-6
242. P. Ehrlich (ed.), *Real Numbers, Generalizations of the Reals, and Theories of Continua*. 1994
ISBN 0-7923-2689-X
243. G. Debrock and M. Hulswit (eds.), *Living Doubt*. Essays concerning the epistemology of Charles Sanders Peirce. 1994
ISBN 0-7923-2898-1
244. J. Szrednicki, *To Know or Not to Know*. Beyond Realism and Anti-Realism. 1994
ISBN 0-7923-2909-0
245. R. Egidi (ed.), *Wittgenstein: Mind and Language*. 1995
ISBN 0-7923-3171-0
246. A. Hyslop, *Other Minds*. 1995
ISBN 0-7923-3245-8
247. L. Pólos and M. Masuch (eds.), *Applied Logic: How, What and Why*. Logical Approaches to Natural Language. 1995
ISBN 0-7923-3432-9
248. M. Krynicki, M. Mostowski and L.M. Szczerba (eds.), *Quantifiers: Logics, Models and Computation*. Volume One: Surveys. 1995
ISBN 0-7923-3448-5
249. M. Krynicki, M. Mostowski and L.M. Szczerba (eds.), *Quantifiers: Logics, Models and Computation*. Volume Two: Contributions. 1995
ISBN 0-7923-3449-3
250. R.A. Watson, *Representational Ideas from Plato to Patricia Churchland*. 1995
ISBN 0-7923-3453-1
251. J. Hintikka (ed.), *From Dedekind to Gödel*. Essays on the Development of the Foundations of Mathematics. 1995
ISBN 0-7923-3484-1
252. A. Wiśniewski, *The Posing of Questions*. Logical Foundations of Erotetic Inferences. 1995
ISBN 0-7923-3637-2

SYNTHESE LIBRARY

253. J. Peregrin, *Doing Worlds with Words*. Formal Semantics without Formal Metaphysics. 1995
ISBN 0-7923-3742-5
254. I.A. Kieseppä, *Truthlikeness for Multidimensional, Quantitative Cognitive Problems*. 1996
ISBN 0-7923-4005-1
255. P. Hugly and C. Sayward: *Intensionality and Truth*. An Essay on the Philosophy of A.N. Prior. 1996
ISBN 0-7923-4119-8
256. L. Hankinson Nelson and J. Nelson (eds.): *Feminism, Science, and the Philosophy of Science*. 1997
ISBN 0-7923-4162-7
257. P.I. Bystrov and V.N. Sadovsky (eds.): *Philosophical Logic and Logical Philosophy*. Essays in Honour of Vladimir A. Smirnov. 1996
ISBN 0-7923-4270-4
258. Å.E. Andersson and N-E. Sahlin (eds.): *The Complexity of Creativity*. 1996
ISBN 0-7923-4346-8
259. M.L. Dalla Chiara, K. Doets, D. Mundici and J. van Benthem (eds.): *Logic and Scientific Methods*. Volume One of the Tenth International Congress of Logic, Methodology and Philosophy of Science, Florence, August 1995. 1997
ISBN 0-7923-4383-2
260. M.L. Dalla Chiara, K. Doets, D. Mundici and J. van Benthem (eds.): *Structures and Norms in Science*. Volume Two of the Tenth International Congress of Logic, Methodology and Philosophy of Science, Florence, August 1995. 1997
ISBN 0-7923-4384-0
Set ISBN (Vols 259 + 260) 0-7923-4385-9
261. A. Chakrabarti: *Denying Existence*. The Logic, Epistemology and Pragmatics of Negative Existentials and Fictional Discourse. 1997
ISBN 0-7923-4388-3
262. A. Biletzki: *Talking Wolves*. Thomas Hobbes on the Language of Politics and the Politics of Language. 1997
ISBN 0-7923-4425-1
263. D. Nute (ed.): *Defeasible Deontic Logic*. 1997
ISBN 0-7923-4630-0
264. U. Meixner: *Axiomatic Formal Ontology*. 1997
ISBN 0-7923-4747-X
265. I. Brinck: *The Indexical 'I'*. The First Person in Thought and Language. 1997
ISBN 0-7923-4741-2
266. G. Hölmström-Hintikka and R. Tuomela (eds.): *Contemporary Action Theory*. Volume 1: Individual Action. 1997
ISBN 0-7923-4753-6; Set: 0-7923-4754-4
267. G. Hölmström-Hintikka and R. Tuomela (eds.): *Contemporary Action Theory*. Volume 2: Social Action. 1997
ISBN 0-7923-4752-8; Set: 0-7923-4754-4
268. B.-C. Park: *Phenomenological Aspects of Wittgenstein's Philosophy*. 1998
ISBN 0-7923-4813-3
269. J. Paśniczek: *The Logic of Intentional Objects*. A Meinongian Version of Classical Logic. 1998
Hb ISBN 0-7923-4880-X; Pb ISBN 0-7923-5578-4
270. P.W. Humphreys and J.H. Fetzer (eds.): *The New Theory of Reference*. Kripke, Marcus, and Its Origins. 1998
ISBN 0-7923-4898-2
271. K. Szaniawski, A. Chmielewski and J. Woleński (eds.): *On Science, Inference, Information and Decision Making*. Selected Essays in the Philosophy of Science. 1998
ISBN 0-7923-4922-9
272. G.H. von Wright: *In the Shadow of Descartes*. Essays in the Philosophy of Mind. 1998
ISBN 0-7923-4992-X
273. K. Kijania-Placek and J. Woleński (eds.): *The Lvov-Warsaw School and Contemporary Philosophy*. 1998
ISBN 0-7923-5105-3
274. D. Dedrick: *Naming the Rainbow*. Colour Language, Colour Science, and Culture. 1998
ISBN 0-7923-5239-4
275. L. Albertazzi (ed.): *Shapes of Forms*. From Gestalt Psychology and Phenomenology to Ontology and Mathematics. 1999
ISBN 0-7923-5246-7
276. P. Fletcher: *Truth, Proof and Infinity*. A Theory of Constructions and Constructive Reasoning. 1998
ISBN 0-7923-5262-9

SYNTHESE LIBRARY

277. M. Fitting and R.L. Mendelsohn (eds.): *First-Order Modal Logic*. 1998
Hb ISBN 0-7923-5334-X; Pb ISBN 0-7923-5335-8
278. J.N. Mohanty: *Logic, Truth and the Modalities from a Phenomenological Perspective*. 1999
ISBN 0-7923-5550-4
279. T. Placek: *Mathematical Intuitionism and Intersubjectivity*. A Critical Exposition of Arguments for Intuitionism. 1999
ISBN 0-7923-5630-6
280. A. Cantini, E. Casari and P. Minari (eds.): *Logic and Foundations of Mathematics*. 1999
ISBN 0-7923-5659-4 set ISBN 0-7923-5867-8
281. M.L. Dalla Chiara, R. Giuntini and F. Laudisa (eds.): *Language, Quantum, Music*. 1999
ISBN 0-7923-5727-2; set ISBN 0-7923-5867-8
282. R. Egidi (ed.): *In Search of a New Humanism*. The Philosophy of Georg Hendrik von Wright. 1999
ISBN 0-7923-5810-4
283. F. Vollmer: *Agent Causality*. 1999
ISBN 0-7923-5848-1
284. J. Peregrin (ed.): *Truth and Its Nature (if Any)*. 1999
ISBN 0-7923-5865-1
285. M. De Caro (ed.): *Interpretations and Causes*. New Perspectives on Donald Davidson's Philosophy. 1999
ISBN 0-7923-5869-4
286. R. Murawski: *Recursive Functions and Metamathematics*. Problems of Completeness and Decidability, Gödel's Theorems. 1999
ISBN 0-7923-5904-6
287. T.A.F. Kuipers: *From Instrumentalism to Constructive Realism*. On Some Relations between Confirmation, Empirical Progress, and Truth Approximation. 2000
ISBN 0-7923-6086-9
288. G. Holmström-Hintikka (ed.): *Medieval Philosophy and Modern Times*. 2000
ISBN 0-7923-6102-4
289. E. Grosholz and H. Breger (eds.): *The Growth of Mathematical Knowledge*. 2000
ISBN 0-7923-6151-2
290. G. Sommaruga: *History and Philosophy of Constructive Type Theory*. 2000
ISBN 0-7923-6180-6
291. J. Gasser (ed.): *A Boole Anthology*. Recent and Classical Studies in the Logic of George Boole. 2000
ISBN 0-7923-6380-9
292. V.F. Hendricks, S.A. Pedersen and K.F. Jørgensen (eds.): *Proof Theory*. History and Philosophical Significance. 2000
ISBN 0-7923-6544-5
293. W.L. Craig: *The Tensed Theory of Time*. A Critical Examination. 2000
ISBN 0-7923-6634-4
294. W.L. Craig: *The Tenseless Theory of Time*. A Critical Examination. 2000
ISBN 0-7923-6635-2
295. L. Albertazzi (ed.): *The Dawn of Cognitive Science*. Early European Contributors. 2001
ISBN 0-7923-6799-5
296. G. Forrai: *Reference, Truth and Conceptual Schemes*. A Defense of Internal Realism. 2001
ISBN 0-7923-6885-1
297. V.F. Hendricks, S.A. Pedersen and K.F. Jørgensen (eds.): *Probability Theory*. Philosophy, Recent History and Relations to Science. 2001
ISBN 0-7923-6952-1
298. M. Esfeld: *Holism in Philosophy of Mind and Philosophy of Physics*. 2001
ISBN 0-7923-7003-1
299. E.C. Steinhart: *The Logic of Metaphor*. Analogous Parts of Possible Worlds. 2001
ISBN 0-7923-7004-X
300. P. Gärdenfors: *The Dynamics of Thought*. 2005
ISBN 1-4020-3398-2
301. T.A.F. Kuipers: *Structures in Science Heuristic Patterns Based on Cognitive Structures*. An Advanced Textbook in Neo-Classical Philosophy of Science. 2001
ISBN 0-7923-7117-8
302. G. Hon and S.S. Rakover (eds.): *Explanation*. Theoretical Approaches and Applications. 2001
ISBN 1-4020-0017-0
303. G. Holmström-Hintikka, S. Lindström and R. Sliwinski (eds.): *Collected Papers of Stig Kanger with Essays on his Life and Work*. Vol. I. 2001
ISBN 1-4020-0021-9; Pb ISBN 1-4020-0022-7

SYNTHESE LIBRARY

304. G. Holmström-Hintikka, S. Lindström and R. Sliwinski (eds.): *Collected Papers of Stig Kanger with Essays on his Life and Work*. Vol. II. 2001 ISBN 1-4020-0111-8; Pb ISBN 1-4020-0112-6
305. C.A. Anderson and M. Zelény (eds.): *Logic, Meaning and Computation*. Essays in Memory of Alonzo Church. 2001 ISBN 1-4020-0141-X
306. P. Schuster, U. Berger and H. Osswald (eds.): *Reuniting the Antipodes – Constructive and Nonstandard Views of the Continuum*. 2001 ISBN 1-4020-0152-5
307. S.D. Zwart: *Refined Verisimilitude*. 2001 ISBN 1-4020-0268-8
308. A.-S. Maurin: *If Tropes*. 2002 ISBN 1-4020-0656-X
309. H. Eilstein (ed.): *A Collection of Polish Works on Philosophical Problems of Time and Spacetime*. 2002 ISBN 1-4020-0670-5
310. Y. Gauthier: *Internal Logic*. Foundations of Mathematics from Kronecker to Hilbert. 2002 ISBN 1-4020-0689-6
311. E. Ruttkamp: *A Model-Theoretic Realist Interpretation of Science*. 2002 ISBN 1-4020-0729-9
312. V. Rantala: *Explanatory Translation*. Beyond the Kuhnian Model of Conceptual Change. 2002 ISBN 1-4020-0827-9
313. L. Decock: *Trading Ontology for Ideology*. 2002 ISBN 1-4020-0865-1
314. O. Ezra: *The Withdrawal of Rights*. Rights from a Different Perspective. 2002 ISBN 1-4020-0886-4
315. P. Gärdenfors, J. Woleński and K. Kijania-Placek: *In the Scope of Logic, Methodology and Philosophy of Science*. Volume One of the 11th International Congress of Logic, Methodology and Philosophy of Science, Cracow, August 1999. 2002 ISBN 1-4020-0929-1; Pb 1-4020-0931-3
316. P. Gärdenfors, J. Woleński and K. Kijania-Placek: *In the Scope of Logic, Methodology and Philosophy of Science*. Volume Two of the 11th International Congress of Logic, Methodology and Philosophy of Science, Cracow, August 1999. 2002 ISBN 1-4020-0930-5; Pb 1-4020-0931-3
317. M.A. Changizi: *The Brain from 25,000 Feet*. High Level Explorations of Brain Complexity, Perception, Induction and Vagueness. 2003 ISBN 1-4020-1176-8
318. D.O. Dahlstrom (ed.): *Husserl's Logical Investigations*. 2003 ISBN 1-4020-1325-6
319. A. Biletzki: *(Over)Interpreting Wittgenstein*. 2003 ISBN Hb 1-4020-1326-4; Pb 1-4020-1327-2
320. A. Rojszczak, J. Cacho and G. Kurczewski (eds.): *Philosophical Dimensions of Logic and Science*. Selected Contributed Papers from the 11th International Congress of Logic, Methodology, and Philosophy of Science, Kraków, 1999. 2003 ISBN 1-4020-1645-X
321. M. Sintonen, P. Ylikoski and K. Miller (eds.): *Realism in Action*. Essays in the Philosophy of the Social Sciences. 2003 ISBN 1-4020-1667-0
322. V.F. Hendricks, K.F. Jørgensen and S.A. Pedersen (eds.): *Knowledge Contributors*. 2003 ISBN Hb 1-4020-1747-2; Pb 1-4020-1748-0
323. J. Hintikka, T. Czarnecki, K. Kijania-Placek, T. Placek and A. Rojszczak † (eds.): *Philosophy and Logic In Search of the Polish Tradition*. Essays in Honour of Jan Woleński on the Occasion of his 60th Birthday. 2003 ISBN 1-4020-1721-9
324. L.M. Vaina, S.A. Beardsley and S.K. Rushton (eds.): *Optic Flow and Beyond*. 2004 ISBN 1-4020-2091-0
325. D. Kolak (ed.): *I Am You*. The Metaphysical Foundations of Global Ethics. 2004 ISBN 1-4020-2999-3
326. V. Stepin: *Theoretical Knowledge*. 2005 ISBN 1-4020-3045-2
327. P. Mancosu, K.F. Jørgensen and S.A. Pedersen (eds.): *Visualization, Explanation and Reasoning Styles in Mathematics*. 2005 ISBN 1-4020-3334-6
328. A. Rojszczak (author) and J. Wolenski (ed.): *From the Act of Judging to the Sentence*. The Problem of Truth Bearers from Bolzano to Tarski. 2005 ISBN 1-4020-3396-6
329. A. Pietarinen: *Signs of Logic*. Peircean Themes on the Philosophy of Language, Games, and Communication. 2005 ISBN 1-4020-3728-7

SYNTHESE LIBRARY

330. A. Aliseda: *Abductive Reasoning*. Logical Investigations into Discovery and Explanation. 2005
ISBN 1-4020-3906-9
331. B. Feltz, M. Crommelinck and P. Goujon (eds.): *Self-organization and Emergence in Life Sciences*.
2005 ISBN 1-4020-3916-6
332. R. van Rooij: *Attitudes and Changing Contexts*. 2006 ISBN 1-4020-4176-4
333. L. Albertazzi: *Immanent Realism*. An Introduction to Brentano. 2006 ISBN 1-4020-4201-9
334. A. Keupink and S. Shieh (eds.): *The Limits of Logical Empiricism*. Selected Papers of Arthur Pap.
2006 ISBN 1-4020-4298-1
335. M. van Atten: *Brouwer meets Husserl*. On the Phenomenology of Choice Sequences. 2006
ISBN 1-4020-5086-0
336. F. Stjernfelt: *Diagrammatology*. An Investigation on the Borderlines of Phenomenology, Ontology,
and Semiotics. 2006 ISBN 978-1-4020-5651-2
337. H. van Ditmarsch, W. van der Hoek and B. Kooi: *Dynamic Epistemic Logic*. 2006
ISBN 978-1-4020-5838-7
338. N. Nottelmann, *Blameworthy Belief*. A Study in Epistemic Deontology. 2007
ISBN 978-1-4020-5960-5