LUND UNIVERSITY

The Collective Archives of Mind

An Exploration of Reasons from Metaethics to Social Ontology

Mähringer, Gloria

2020

Link to publication

Citation for published version (APA):
Mähringer, G. (2020). *The Collective Archives of Mind: An Exploration of Reasons from Metaethics to Social Ontology*. [Doctoral Thesis (monograph), Department of Philosophy]. Lund University (Media-Tryck).

Total number of authors:
1

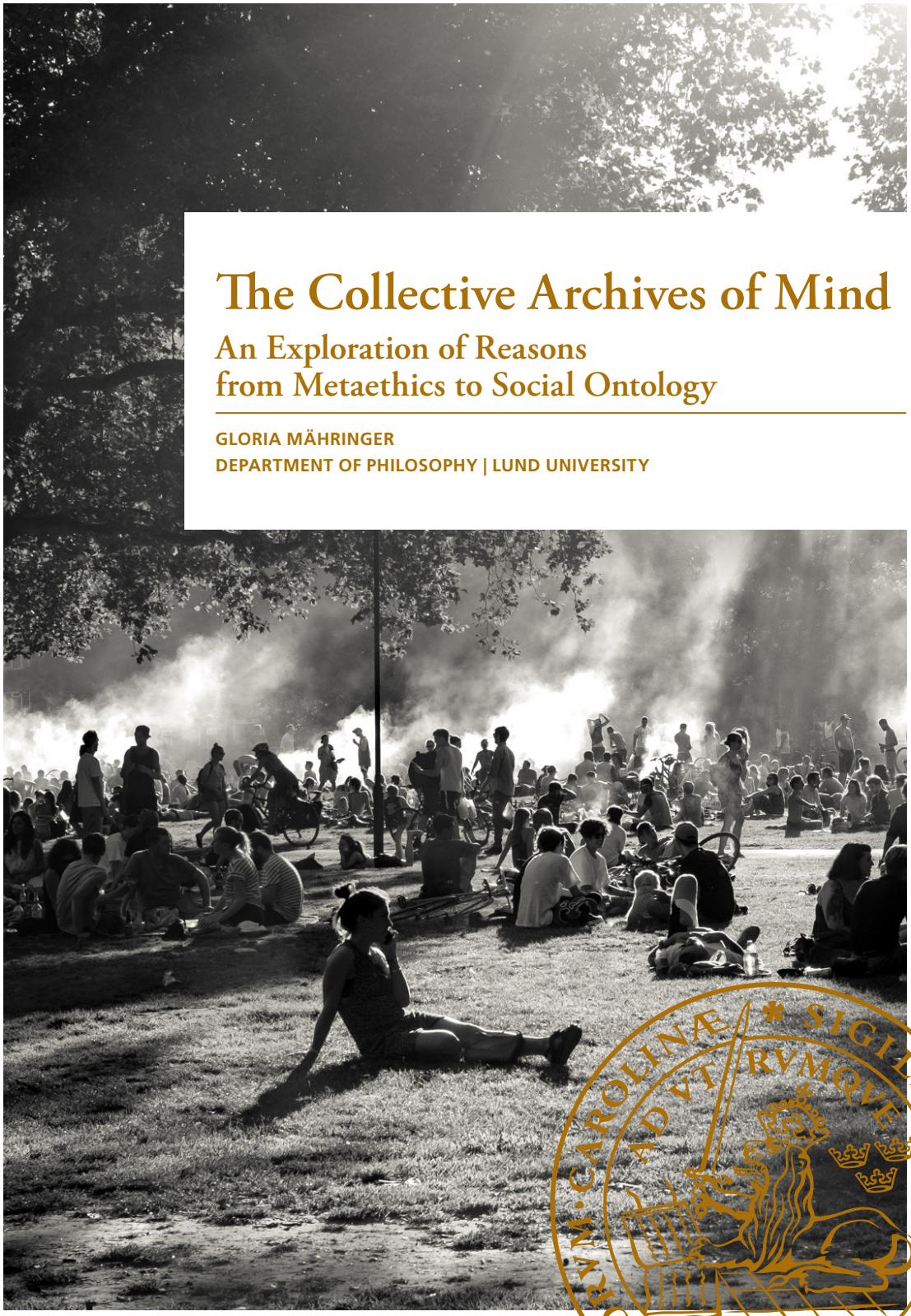
General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00



The Collective Archives of Mind

An Exploration of Reasons from Metaethics to Social Ontology

GLORIA MÄHRINGER
DEPARTMENT OF PHILOSOPHY | LUND UNIVERSITY

The central aim of this book is to explore what it is to be a reason – a consideration of normative weight for a reflective creature. When we understand the reasons supporting our decisions, we become more self-determined in our choices. When we become more self-determined, we usually also become better at explaining and justifying ourselves to others. There seems to be a natural connection between self-determination and communication. Nonetheless, there are moments in which our belief in this natural connection is shattered. We must make concessions we cannot fully support in order to justify ourselves. Or we must face disharmony and instability in order to carry through with our convictions.

This book argues that our capacity to make self-determined decisions depends on communities – communities as the historical source of considerations that can be endorsed as reasons. Self-determined reasoning, so the core thesis of this book, is moving in what we can call the collective archives of mind. However, this does not mean that we cannot overcome or change communal traditions, but in order to do this, we may have to give up the ideal of self-determination – the ideal of understanding our decisions as fully justified by reasons.

Printed by Media-Tryck, Lund 2020 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

Faculty of Humanities and Theology
Department of Philosophy
ISBN 978-91-89213-19-7

The Collective Archives of Mind

The Collective Archives of Mind

An Exploration of Reasons from Metaethics to Social Ontology

Gloria Mähringer

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Humanities and Theology, Lund University, Sweden.
To be defended at the Department of Philosophy, LUX C121 on the 21th of November 2020 at 11:00 am.

Thesis advisors: Björn Petersson, David Alm

Faculty opponent
Ulrike Heuer

Organization LUND UNIVERSITY	**Document name:** DOCTORAL DISSERTATION
	Date of issue: 2020-10-23
Author: Gloria Mähringer	Sponsoring organization

Title and subtitle: The Collective Archives of Mind. An Exploration of Reasons from Metaethics to Social Ontology

Abstract

This monograph discusses the question of what it is to be a reason – mainly in practical ethics – and proposes an original contribution to metaethics.

It critically examines theories of metaethical realism, constructivism and error theory and identifies several misunderstandings or unclarities in contemporary debates. Based on this examination, the book suggests a distinction between a conceptual question, that can be answered by pure first-personal thinking, and a material question, that targets responses to reasons as a natural phenomenon in space and time and that can be answered by help of the sciences. While this book defends a sharp distinction between these approaches, it also argues that the insights gained by the distinct approaches can be fruitfully integrated into a comprehensive picture.

The comprehensive picture promoted in this book, based on both philosophical analysis and resources from psychology and cognitive science, is a picture of (centrally normative) reasons as items in a collective archive of mind – reconfiguring the "domain of reasons" as a collectively established and socially cultivated fund of considerations with normative weight. The developed Collective Archive View combines Reasons Fundamentalism as an answer to the conceptual question, with a social ontological proposal, similar to Haslanger's critical realism, as an answer to the material question.

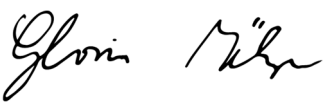
Thereby, it introduces central ideas from the field of social metaphysics into metaethics. At the same time, it offers a way of settling some controversies within metaethics, going on between proponents of Reasons Fundamentalism and Constitutivism, realism and constructivism, as well as proponents of mind-dependence and mind-independence of normative facts. Finally, the proposal enables us to envision and conceptualize genuine normative change within a metaethical theory.

Key words: Reasons Fundamentalism, constitutivism, constructivism, realism, error theory, domain of reasons, self-determination, normative harmony, normative instability, disruption of thought, normative transformation, social ontology, critical realism, affordances of thought, reason relations, collective archive

Classification system and/or index terms (if any)

Supplementary bibliographical information		**Language:** English
ISSN and key title		**ISBN** 978-91-89213-19-7
Recipient's notes	**Number of pages:** 268	Price
	Security classification	

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature  Lund, 2020-09-28

The Collective Archives of Mind

An Exploration of Reasons from Metaethics to Social Ontology

Gloria Mähringer

LUND
UNIVERSITY

Coverphoto by Christoph Strolz,
Design by Luisa Mähringer

Copyright © Gloria Mähringer

The Faculty of Humanities and Theology
Department of Philosophy

ISBN: 978-91-89213-19-7

Printed in Sweden by Media-Tryck, Lund University
Lund 2020

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se
MADE IN SWEDEN

Dedicated to my family
Manuela, Erich, Luisa and Lorenz

We have access not only to our own lives but to almost all the other lives in our cultural circle, access not only to our own memories but to the memories of the whole of our damn culture, for I am you and you are everyone, we come from the same and are going to the same […]

Even if you sit in a tiny room in a tiny town hundreds of kilometres from the center of the world and don't meet a single soul, their hell is your hell, their heaven is your heaven […]

Language is shared, we grow into it, and the forms we use it in are also shared, so irrespective of how idiosyncratic you and your notions are, in literature you can never free yourself from others. It is the other way round, it is literature that draws us closer together, through its language, which none of us owns and which indeed we can hardly have any influence on, and through its form, which no one can break free of alone, and if anyone should do so, it is only meaningful if it is immediately followed by others.

Karl Ove Knausgaard, *My struggle*, Book II
(2013: 543f)

Acknowledgments

Four years ago, the department of philosophy at Lund offered me the opportunity to dedicate four fully funded years exclusively to the development of this philosophical work. This generous opportunity, and the comprehensive personal and institutional support I received during this inspiring and fruitful period, is far from what most people can hope for. I am deeply grateful for my encounter with the Swedish culture and university system, as one of the best in the world, with its well-organized and well-funded doctoral education programme. If I can find a way of supporting, defending or strengthening the Swedish system in the future, my experiences in the past have made me very motivated to do so.

More particularly, I am grateful for having been integrated into a fantastic department and for being part of the practical philosophy research group in Lund. Björn Petersson and David Alm have been exceptionally supportive and competent supervisors to me. I have benefitted from Björn's calmness and open-mindedness that encouraged a quite ambitious project and from David's rigorous scrutiny that hopefully made my claims more precise. Björn and David functioned as a perfectly complementary supervisor team. I thank them for giving me a lot of space and support and enabling me to have a completely stress-free and creative working process.

As the opponent of my final seminar, Johan Brännmark stimulated an incredibly productive discussion of the first full draft of my work and provided me with plenty of "food-for-thought" that helped greatly to improve my thesis. I thank him for his careful in-depth-studying of my work.

The very idea of reconfiguring the "domain of reasons" as a collective archive of mind, was inspired by an intensive critical lecture of John Skorupski's major work, *The Domain of Reasons*, with its highly fascinating account of "self-determination" and the idea of reason responses constituting the process of free thought. The ideas resonated to some extent with the ideas by Harry G. Frankfurt and Christine M. Korsgaard that I worked with during my MA in Vienna, but also involved a break with the subjectivist tradition I previously dealt with. This intriguing tension quickly captured my interest and urged me to try to delineate and reconcile the seemingly opposing claims of

realists and constructivists, of Reasons Fundamentalists and Constitutivists, and of philosophers claiming that normative facts depend on mind and such that argue that they are mind-independent.

I am grateful for an exchange with John Skorupski and for many fruitful discussions in Lund. I am grateful to my teachers in Vienna, Angela Kallhoff, Herlinde Pauer-Studer and Hans Bernhard Schmid, for providing the starting ground for this work and to the practical philosophers in Lund for challenging my previously held assumptions – in this sense especially Toni Rønnow-Rasmussen, Wlodek Rabinowicz, Jakob Werkmäster and Andrés Garcia.

However, the work would not have developed without the numerous less confrontative, but at least as productive discussions with the other participants of the Higher Seminar in Practical Philosophy and the Doctoral Seminar in Lund. I thank Frits Gåvertsson, Mattias Gunnemyr, Eric Brandtstedt, Olle Blomberg, Henrik Andersson, Dan Egonsson, Cathrine Felix, Ingvar Johansson, Matthew Talbert, Caroline Torpe Touborg, Marta Johansson Werkmäster, Signe Savén, Anton Emilsson, Alexander Velichkov, Max Minden Ribeiro, Robert Pál, Marianna Leventi, Thérèse Söderström, Andreas Stephens, Maximilian Roszko, Trond Arild Tjöstheim, Paul Russell, David Shoemaker, András Szigeti, Erik Persson, Fritz-Anton Fritzson, Seyyed Mohsen Eslami, Yan Chen, Berit Braun, Ylva von Gerber, Viktoria Höög and others whom I might have missed.

Regarding the empirical perspective that I bring into the picture, I am indebted to Ingar Brinck, who read the empirically oriented parts of my dissertation, and to Martin Weichold, who introduced me to ecological psychology and enactivist cognitive science.

I further benefitted from regularly attending the Higher Seminar in Theoretical Philosophy at Lund. I thank Erik Olsson, Ingar Brinck, Tobias Hansson Wahlberg, Martin Jönsson, Robin Stenwall, Max Minden Ribeiro, Andreas Stephens, Arvid Båve, Asger Kirkeby-Hinrup, Carl-Johan Palmqvist, Oscar Ralsmark and Patrizio Lo Presti.

The numerous national and international guests at the research seminars contributed a great deal to creating a stimulating environment for doing a phD. Especially helpful for my work were the encounters with Mattia Gallotti and Thomas Schmidt (practical philosophy) and Hans-Johann Glock (theoretical philosophy).

A further benefit was the availability of cognitive scientists at our department that enabled discussions with, among others, Manuel Oliva, Maximilian Roszko, Trond Arild Tjöstheim, Claudia Zeiträg, Gabriel Vogel, Katarzyna Bobrowicz and Zahra Gharaee.

For securing a great institutional environment, I thank the head of department Tomas Persson, the head of third cycle studies Tobias Hansson Wahlberg, and the always supportive administrative staff Anna Östberg, Astrid Byrman, Anna Cagnan Enhörning, Annah Smedberg-Eivers, Ingela Byström, Fredik Eriksson and Jenny Oldbring.

I also thank the department for enabling the participation in national courses at the universities of Stockholm, Uppsala and Umeå. Moreover, I want to thank my group from the Faculty´s introductory course for phD students in autumn 2016, Ida Jansson, Allan Burnett, Nina Jakku and Rasmus Kræmmer Rendsvig for inspiring exchange and for most of us staying in touch all through the phD time.

During my time in Lund, I had the opportunity to gather international experience on various occasions. Most importantly, I would like to mention my visiting scholarship at the University of Barcelona during spring 2019. I thank the Royal Swedish Academy of Letters (Kungliga Vitterhetsakademien) for generous funding and the LOGOS Research Group in Analytic Philosophy, especially Manuel García-Carpintero, for hosting me. In Barcelona, I benefitted greatly from the LOGOS seminars and colloquia, the Graduate Reading Group and other research activities within the group. I am grateful for exchange with Manuel García-Carpintero, Esa Díaz-León, Enrico Terrone, Filippo Contesi, Bartłomiej Czajka, Aarón Àlvarez González, Lisa Vogt, Diana Couto, Andrea Rivadulla Duró, Johannes Findl, Ramón García Moya, Alfonso García Lapeña, Anna Giustina, Anna Ciaunica, Javier Suárez, Grzegorz Gaszczyk, Anton Alexandrov, Matheus Valente, Neri Marsili, Genoveva Martí, Josefa Toribio, Teresa Marques, Víctor Verdejo, Carlota Serrahima, Dan Lopéz de Sa, Alexander Krauss, Sven Rosenkranz, Adam Sennet, Michele Palmira, Samuele Chilovi, Marta Jorba and others that I may have missed.

Moreover, I was allowed to give talks at several international conferences that provided me with extremely valuable input and feedback on my work. Just before starting my phD in 2016, I was allowed to present at the Masterclass with John Broome in Tübingen. I thank the organizer Sabine Döring as well as the speakers and participants, especially John Broome, Herlinde Pauer-Studer, Joanna Klimczyk, Nora Heinzelmann, Simon Gaus, Simon Wimmer, Tao Stein, Bahadir Eker, Peter Königs and Eva-Maria Düringer for valuable exchange.

In 2017, I had a great experience presenting at “Dimensions of Responsibility” in Pavia, the 20th anniversary conference of the journal “Ethical Theory and Moral Practice”. I benefitted from exchange with, among others, Ezio Di Nucci, Marcus Düwell, Lubomira Radoilska, Robin Zheng,

Thomas Schramme, András Szigeti, Alexa Zellentin, Maureen Sie, László Bernáth and Anna Réz.

In the same year, I presented at ECAP 9 in Munich, where I especially benefitted from an encounter with Susanne Mantel, and ENSO V in Lund, where I received valuable feedback on an early proposal about connecting the analysis of reasons with social ontology, most notably from Åsa Burman.

Furthermore, I thank Carla Bagnoli and Wolfgang Huemer for giving me the opportunity to present at their workshop "Norms and Change" in Parma in 2018. I thank the organizers as well as the speakers and participants, in particular Cristina Bicchieri, Elijah Millgram, Francesco Guala, Anne Reichold, Martin Weichold, Mihaela Popa and Damiano Ranzenigo.

Also in 2018, I had the opportunity to present at "The Future of Normativity" in Kent, where I presented the first version of my interpretation of the domain of reasons and could meet many interesting people. I am especially grateful for the very encouraging feedback on my approach by Peter Railton and Hakob Barseghyan, the very active discussion with the audience of my talk, and the exchange with Janis Schaab, Olof Leffler, David Heering, Lizzy Ventham, Michael Smith, Alexandra Couto, Simon Kirchin, Bart Streumer, Nils Franzén and Olle Risberg.

In 2020, I presented a developed version of my introduction of social ontology into metaethics at "Social Ontology 2020". I thank Alessandro Salice and Olivier Massin, both for organizing the conference and for actively participating in the Q&A of my talk. Furthermore, I thank Jonathan Schaffer, Anna Moltchanova, Marcello Ruta, Anna Strasser and Raffaela Giovagnoli for very constructive comments.

For providing travel grants, I thank Hilma Borelius stipendiefond, Knut och Alice Wallenbergs stiftelse, Dagmar Perssons fond, Stiftelsen Roy och Maj Franzéns fond, as well as Erik and Gurli Hultengrens fond för filosofi vid Lunds Universitet.

Apart from that, the generous funding by Erik and Gurli Hultengrens fond allowed me to realize a Graduate Workshop on Fiction and Philosophy in 2018, which I organized together with Frits Gåvertsson, Thérèse Söderström, Ylwa Sjölin Wirling and Anna Persson. I am grateful to my co-organizers, especially Frits and Thérèse, as well as the speakers and the audience, in particular Nora Hämäläinen, Maria Green, Nils Franzén, Tadej Todorović, Diana Neiva, Matías Graffigna, Erin Kavanagh, Axel Rudolphi, Francesca Rodesino and Steven Gouveia.

Moreover, I am grateful for the possibility to serve the Philosophical Society of Lund as a chair since 2017. I thank the people that have been on the board with me, Frits Gåvertsson, Mattias Gunnemyr, Ylwa Sjölin Wirling, Rosanna

Ljungren, Thérèse Söderström, Lina Hallberg, Olle Blomberg, Caroline Torpe Touborg, Melina Tsapos, Robert Pál, Anton Emilsson, Max Minden Ribeiro, Rasmus Magnusson, Marianna Leventi, and accountant Ulf Gallbo, for a wonderful cooperation and the speakers and audiences of our various talks for many philosophically fruitful evenings.

I am indebted to Herlinde Pauer-Studer and Christoph Hanisch for many discussions of theories and ideas that have been central to my work since my time in Vienna. Discussions with Federica Berdini and Xiaoxi Wu helped especially during the early stage of my project. Martin Weichold, who shared my interest in the reconciliation of normative analysis with an empirical perspective on cognition, contributed a great deal to shaping my research interest. I am grateful to Thérèse Söderström for reading and commenting on a full draft of this book.

I further thank Frits Gåvertsson, Thérèse Söderström, Melina Tsapos, Zahra Gharaee, Federica Berdini, Xiaoxi Wu, Katharina Bernhard, Christoph Strolz, Lisa Häberlein and Lisa Vogt for many conversations about philosophy, academia and beyond. Especially the exchange with Frits has greatly shaped my outlook on philosophy and helped me making productive use of the challenging tension between the values that shaped during my childhood and early youth and the culture I encountered in the international academic environment.

The decision to put aside the medical career I was preparing for in order to embark for a full-time doctoral programme in philosophy has not been easy, and has not been done without regret, but offered a unique opportunity to question, revise and reaffirm my reasons and values.

In any case, I would probably not have enjoyed the extensive amount of time for personal encounters and conversations, the many patient and supportive people surrounding me and the time and infrastructure to delve into the wide-ranging and comprehensive philosophical interest that lead up to the present work – not to mention the intense contact with Swedish culture, language, art and literature. In this context, I especially thank Dan Egonsson, Björn Petersson, Ulrika Kristiansson, Jonas Josefsson, Frits Gåvertsson, Anna Karlsson, Jakob Werkmäster and Marta Johansson Werkmäster for their good recommendations and introductions into the Swedish cultural sphere, and Maria Green for regular exchange about international literature, human rights and the public impact of philosophy. For the cover of this book, I thank Christoph Strolz and Luisa Mähringer.

Finally, I would like to thank my family, Manuela, Erich, Luisa and Lorenz, to whom this book is dedicated for being the greatest family I could wish for. I also thank my extended family, in particular my grandparents Hilda and Willi,

Irma and Heinrich, my grandaunt Gabi, my great granduncle Adalbert, my aunt Ingrid and all my other aunts and uncles for providing a stimulating and rich childhood environment. The constructive tension between a traditional rural Catholicism and an open, progressive and socio-critical mindset proved extraordinarily fruitful for intellectual development.

Gloria Mähringer
Wolfersgrün (Frankenwald)
September 2020

Table of Contents

Acknowledgments 11

Table of Contents 17

Introduction 21

Chapter 1
What do we have reason to do?
Thinking about normative questions, questioning the limits of thinking 29

1.1. Introduction 29

1.2. The double nature of reason and its internal disruption 31
- 1.2.1. The double nature of reason 32
- 1.2.2. Intrapersonal disruptions of reason 35
- 1.2.3. Interpersonal disruptions of reason 37

1.3. Disrupted reason in light of metaethics 40
- 1.3.1. Error theory: claiming normative facts that do not exist 41
- 1.3.2. Realism: a theory of irreducible normative facts 43
- 1.3.3. Constructivism: objective normative facts that depend on us 49

1.4. Disentangling methodological confusions 55
- 1.4.1. The ambiguity of constitutivist constructivism 57
- 1.4.2. Misunderstanding metaethics? 59
- 1.4.3. Metaethics narrowly conceived and comprehensive inquiry into reasons 61

1.5. Conclusion 67

Chapter 2
Are reasons a fundamental part of reality?
Bridging the gulf between Reasons Fundamentalism and Constitutivism 69

2.1. Introduction 69

2.2. Reasons Fundamentalism and Constitutivism: conflicting positions? 71

2.2.1. Approaching normativity with different starting intuitions 72
2.2.2. The fundamentalist picture: reasons as facts, rationality as responding to reasons 76
2.2.3. The constitutivist picture: reasons as essentially shareable motivational considerations 79
2.2.4. Deriving reasons? The crucial disagreement 82
2.3. Responsiveness to fundamental normative entities: a presupposition for the constitutivist argument 84
2.3.1. Korsgaard's reflective endorsement method and the argument against theoretical knowledge 86
2.3.2. Subjectivist Reasons Fundamentalism? "Reflective endorsement" as "wholeheartedness" 89
2.3.3. Radical subjectivism and the problem with bootstrapping 93
2.4. Uniting the two approaches in a new picture 97
2.4.1. The limitations of Reasons Fundamentalism 97
2.4.2. A version of Reasons Fundamentalism taking the Constitutivist intuition seriously 100
2.5. Conclusion 104

Chapter 3
Are reasons mind-independent?
The relation between grasping and what we grasp 107
3.1. Introduction 107
3.2. Skorupski's Fregean distinction between mind and thoughts: a first controversy about the nature of thinking 109
3.2.1 Distinguishing the irreal realm of thoughts from the real world 110
3.2.2 Distinguishing the irreal realm of thoughts from subjective imaginations 112
3.2.3 Grasping reasons: Spontaneity, apperception and self-determination 114
3.2.4 A controversy about the relation between mind and what it grasps 116
3.3. The shareability of thoughts: a second controversy about the nature of thinking 119
3.3.1. The Convergence Thesis 121
3.3.2. Reason epistemology in the dialectic between spontaneity and convergence 122

3.3.3. A controversy about the relation between communities and the reasons they share 123
3.4. A Darwinian Dilemma for Radical Mind-Independence 127
3.4.1. A tension between the epistemology and the ontology of reasons 128
3.4.2. Street's Darwinian Dilemma for realism 131
3.4.3. How a theory of mind-independent reason relations avoids the dilemma in Street's formulation 132
3.4.4. A new formulation of a Darwinian Dilemma for radically mind-independent reason relations 135
3.4.5. Mapping the ontology on the epistemology 139
3.5. A proposal: modelling reasons as potential mental states 141
3.5.1. A deflationist reading of the mind-independence claim 142
3.5.2. A thought as a configuration of mind 143
3.5.3. The instantiation of a mental state as the actualization of a potential 145
3.6. Conclusion 148

Chapter 4
The domain of reasons as a collective archive
Leading metaethics into social ontology 151
4.1. Introduction 151
4.2. Responding to a reason: philosophy meets science 154
4.2.1. Spontaneity as a subject: Self-understanding in a unified framework 155
4.2.2. A neuroscience-based model of access to an intelligible world 157
4.2.3. Self-determination in light of reason relations: developmental questions 160
4.2.4. Justification practice: The evolution of reflecting in terms of reasons 165
4.3. An active grasp: reasons as affordances of thought 171
4.3.1. The concept of an affordance in ecological psychology 172
4.3.2. Affordances of thought: the realm of reasons in the human environment 177
4.4. Establishing an archive: reason relations as schemata 185
4.4.1. Schemata as the basic building blocks of cognition 186

4.4.2. Collective fictions and the understanding of self and world 188
4.4.3. Reason relations first-personally – schemata third-personally 192
4.4.4. The solitary mind: moving in an archive? 195
4.5. The Collective Archive View: a social ontology of reasons? 200
4.5.1. Socially constructed facts: causal and constitutive construction 201
4.5.2. Haslanger's Critical Realism applied to the domain of reasons 205
4.6. Conclusion 208

Chapter 5
Losing one's mind
Madness, subversion and normative transformation 211
5.1. Introduction 211
5.2. The Collective Archive View between realism, constructivism and error theory 212
5.2.1. Reasons Fundamentalism beyond simple realism 213
5.2.2. Social constructionism vs. normative constructivism 217
5.2.3. Without reasons, no errors 221
5.2.4. An opening question: embracing or rejecting the traditions we depend on? 226
5.3. The costs and gains of embracing the collective archive of our mind 232
5.3.1. Ownership of thought and normative stability 232
5.3.2. Normative instability with subjective disintegration 235
5.3.3. Normative instability with subjective inflation 238
5.4. Normative transformation and retrospective justification 239
5.4.1. Why value ownership of thought? 240
5.4.2. Risking normative stability for the exploration of better alternatives? 243
5.4.3. Retrospectively justifying transformative explorers 245
5.5. Conclusion 249

Conclusion 251

References 255

Introduction

In Ingeborg Bachmann's short story *Ein Schritt nach Gomorrha*, Mara, a young Slovenian girl studying in conservative post-war Vienna, falls in love with her teacher Charlotte. The story describes a lesbian encounter in a society shaped by traditional gender roles and a segregation of social spheres along the lines of these gender roles. In that encounter, during a business travel of Charlotte's husband, Mara tells her teacher about a state of disorientation.

Mara characterizes herself as not knowing what she thinks. Moreover, she talks about her mind like about a material object which she observes from the outside and about which she judges that it does not work properly, that it must be fixed by a superior role model. In a sense, it seems, Mara has lost the ownership of her thought. Feminist theorists have often attributed such states of disorientation or subjective disintegration to the specific situation that women face when they are the first to enter male-dominated spheres of society.[1] In addition to practical hindrances and psychological obstacles such as being intimidated or discouraged, people in such a pioneering role often face genuine epistemic difficulties as well – difficulties in making proper judgments to begin with, apart from difficulties in defending their judgments and carrying through with what they judge to be the right thing to do.[2] In a state of friction in which fundamental normative presuppositions are challenged, individuals may find themselves with genuine difficulties of knowing what reasons they have. Even if they have some rational arguments for why a change is warranted, they might still have difficulties in endorsing these arguments as reasons. Having a rational argument for something, it seems, is not the same as understanding something as a fully reasonable consideration. Thinking rationally that a position is defensible is not the same as endorsing it in the way

[1] See e.g. Lennox 2007; 1992; Lenk 1981.

[2] See also the literature on "epistemic injustice", in particular Fricker's (2007) concept of "hermeneutic injustice" and the picture of an "epistemic lacuna"; (see also Kidd/Medina/Pohlhaus 2017). I especially thank Mattias Gunnemyr for prompting me to the connection between my work and these debates, and Katharina Bernhard for further exchange about the topic.

in which people who see themselves as doing the right thing on good grounds usually endorse their reasons.

This book suggests that we cannot reduce these phenomena to mere epistemological problems. By contrast, these phenomena reveal something about the very ontology of reasons. When we undergo experiences as the one described by Bachmann's fictional character Mara, in a similar way also reported by Simone de Beauvoir (1959), we are not only blocked from the route to normative knowledge. Instead, we face a fact about the nature of what we could call normative facts. The experiences, which we could describe as disruptions of thought, reveal something about the ontology of normative reasons – about the way normative reasons are in the world, not only about the way we get to know them.

Disruptions of thought are not merely epistemological distortions. By contrast, they may be indicative of what Velleman, in his defence of moral relativism, has most expressively labelled as a "moral black hole" (2013: 71). Moral black holes, according to Velleman, are possible because, as he says, normative force always has a centre of gravity in a communally embedded subject (ibid.: 51). Outside of such a community, Velleman proposes, nothing might have normative force. This means that a moral universe we take for granted might collapse if we leave the established ground of a communal tradition. We might encounter a black hole in which, as we could add to Velleman's description, people make the experiences of Bachmann's fictional character Mara, or even more existentially shattering and morally devastating experiences.

The main thesis developed in this book can systematically account for the normative situation in which these experiences are made. The thesis is that conscious normative reflection in terms of reasons is moving in what this book calls a *collective archive of mind*. A reason, according to a wide range of literature in metaethics and the philosophy of action, is a consideration we see as speaking in favour of a particular action. Moreover, reasons have what many authors call normative "weight". This metaphor refers to our capacity of weighing reasons against each other in deliberation, but also expresses the phenomenology of existential importance that reasons sometimes have for reflective creatures. Successful deliberation depends on reasons having a force on our reflective mind that is non-arbitrary and independent of our decision.

According to a traditional philosophical idea, mind responds to or moves within a realm of reasons that is mind-independent. The most radical formulation of such a claim is probably Frege's (1918; English translation 1956) conception of the realm of thoughts as a "third realm" – an objectively existing realm that is ontologically distinct from both the physical and the

subjective realm. This Fregean idea has exerted a strong influence on 20th century analytic philosophy. More recently, it shows up, for example, in Skorupski's (2010) theory of the "domain of reasons" as a domain of objectively mind-independent irreals.

This book remodels that kind of idea. It proposes the concept of a "collective archive of mind", which is a new configuration of the idea of a mind-independent realm of reasons. While indeed being independent of both individual minds and the physical world, a collective archive, as distinct from a Fregean realm of thoughts, has a *social ontology*. As a kind of social reality, a collective archive has an ontology that is indeed independent of both the physical world and the subjective world, conceptualized as what is merely in one person's head. However, the realm of reasons having a social ontology frames reasons as in a crucial sense community-dependent and allows for the possibility that communities diverge from each other and change over time.

The view developed in this book, which could be labelled the *Collective Archive View* (an overview over its claims is to be found at the end of this introduction), solves various tensions observable within contemporary discussions of reasons and normativity. One tension the Collective Archive View can solve is a tension within metaethics, arising around the question whether we must presuppose the concept of a reason as a fundamental one or whether we can analyse it informatively in non-normative terms. *Reasons Fundamentalism*, made famous by Scanlon's defence of realism (2014), but also embraced by Raz (1999a,b), Nagel (1996), Parfit (2011) and Skorupski (2010), defends the former claim. *Constitutivism*, by contrast, proposes that we can analyse reasons in terms of principles that are constitutive of agency. Different versions of such a view are developed by, for example, Korsgaard (1996; 2009), the already mentioned Velleman (2000), as well as Smith (1994), Katsafanas (2013; 2018) or Setiya (2007).

An alternative way to frame the conflict between Reasons Fundamentalism and Constitutivism is to frame it as a conflict between realism and constructivism (Korsgaard 2008b; see also Hanisch 2016).[3] Constitutivist constructivists accuse fundamentalist realists of answering normative questions "by fiat" rather than by explaining what answers to normative questions have to do with human beings (Korsgaard 1996: 33). Constitutivist constructivism, by contrast, is accused of conflating metaethics with normative

[3] In this specific debate, we can read „realism" and „constructivism" as meaning „Reasons Fundamentalism" and „constitutivism". However, the former terms are usually a bit broader and especially „constructivism" can have meanings very different from the one it has in this context, for example as it is used in Rawls' (1971) political philosophy.

ethics (Hussain/Shah 2006; 2013) and of being completely untenable (Enoch 2006; 2011). However, in such discussions it often seems as if proponents of the two different approaches talk past each other. While a few preliminary attempts to reconcile the different approaches in a single coherent picture have been already made (e.g. Hanisch 2016), this book offers a comprehensive and systematic attempt to develop a Reasons Fundamentalism that takes the constitutivist intuition seriously. The book is thereby integrating an *immediate "first-personal" or agential perspective* on reasons as fundamental items of thought with a *detached "third-personal", critical or analytic perspective* from which we can see that we are responsive to certain items because they constitute us as agents. The two perspectives on ourselves as reflective creatures stand in a tension. We cannot take these two perspectives simultaneously. However, we can take each of these perspectives in turn and integrate them into a comprehensive understanding of what reasons are. This integration reveals a tension in our self-understanding. However, this tension does not necessarily undermine our normative practice and self-understanding, as this book holds against, for example, metaethical error theories. Integrating a proper first-personal conceptual analysis and a more detached approach can nevertheless make us aware of an openness and changeability of our normative practices. It allows us, in other words, to conceive of the possibility of genuine transformation of normative frameworks.

Apart from solving this tension within metaethics, the Collective Archive View can also make a strong case for connecting ontological questions in philosophy to empirical questions investigated in other subject areas, such as psychology, sociology or the cognitive sciences. While the concept of a reason has become very influential in analytic metaethics (see e.g. Star 2018), this development is paralleled by a vital interest in the faculty of reasoning in psychology and the cognitive sciences. However, there is typically not so much interaction between these different fields. Normatively interested philosophers often consider psychological and empirical questions as secondary. Empirically minded philosophers, by contrast, frequently reject the classical philosophical idea of humans as creatures equipped with reason in principle.

This unfortunate divide notwithstanding, more recent research in cognitive and evolutionary psychology offers a hitherto unexplored way of linking metaethical inquiry and empirical investigation. In contrast to the general suspicion against conceptualizing humans as reflective creatures, recent publications by scientists have taken conscious reflection seriously as a narrowly and distinctively defined phenomenon. The cognitive scientists Mercier and Sperber (2017) for example, localize reflection in terms of reasons as a specific functional module within our complex cognitive system. They

thereby grant that there are no globally rational creatures as opposed to globally non-rational creatures and that rationality is basically on a par with various practical faculties. Reasoning, according to their view, is a communicative practice of justification. Reasons, accordingly, are a social currency serving cooperative purposes rather than giving an individual the capacity to act more rationally on an absolute scale. According to this picture, there is no absolute linear scale on which we could rank non-reflective behaviour, implicitly rational actions and fully rationalized choices. Instead, all these phenomena are independent cognitive faculties in their own right.

The proposal by Mercier and Sperber can be supported by various other empirically grounded works. However, as distinct from other works, Mercier and Sperber apply a terminology – of reasons as fundamental facts and of rationality as responding to such facts – that offers a possibility to directly connect to discussions within the contemporary debates about reasons and normativity in metaethics. This book picks up that possibility and proposes a philosophical view of reasons, which combines insight from immediate first-personal reflection, the classical philosopher's method, with evidence about the cognitive and social presuppositions for this type of reflection. The Collective Archive View is a view about both the ontology of reasons and the nature of mind, while "mind" here is the narrow phenomenon of reflective consciousness and does not include everything that can be studied under the headings of consciousness, awareness or subjectivity.

The synthesis of the two approaches invites us to realize that what we cannot help but treat as fundamental has been established over time within a collective practice of justification. We can further realize that what we conceive as our very own thoughts and decisions, determined by reasons that we can understand independently of others, with the power of our own reason, is in fact what is given to us by the shared archives of our mind – by the collective archival tradition. We can appreciate our collective archives as what they are. The argument that knowing that a reason is an item within a collective archive makes responding to that reason simply as a reason an error is untenable, according to this book. Instead of appreciating our archived tradition, however, we can likewise wonder whether the archives to which we currently have access are all there is, or whether these archives offer the best that is possible. In other words, we can wonder about possible extensions or changes of our archived normative landscape.

According to the Collective Archive View, it is not very likely that there is a determinate and universally valid answer to such meta-questions – questions whether we should embrace, reject or transform our normative traditions. What the Collective Archive View establishes nevertheless, is that transformative

endeavours come at a cost. The Collective Archive View can specify the cost of embarking for what we could call new normative grounds. These costs involve, among many other social and psychological costs, giving up the possibility of self-determined reasoning. Self-determined reasoning in this sense is a subjective process that is characterized by *ownership of thought*. Ownership of thought, we can further characterize, has both a *subjective* and an *intersubjective* component. The subjective component involves the experience that one's thoughts are expressive of one's very own self rather than coming from outside or being externally determined. The intersubjective component, as distinct from that, involves the effective capacity of using one's thinking in public justification and communication with others. Both components, this book will argue, are inextricably linked. They are inseparable because the psychological phenomenon of self-determined thinking derives from access to a collective archive, while a collective archive, in turn, is the product of communicative justificatory practice.

If we try to challenge or overcome the foundations of the collective archive that constitutes our thought, we can either do so locally and supported by strong subgroups, cultivating their own archives – or we will have to sacrifice an essential part of our psychological well-being: The capacity to own one's thought and think in a self-determined way. Normative pioneers, in a quite literal sense, take the risk of "losing their mind". As normative pioneers are usually not completely solitary, and subgroups usually do not succeed in establishing completely separate and fully exhaustive life worlds either, transformative individuals will most of the time end up in an intermediate state – an instable state of friction between establishing local new grounds and losing one's mind, a state of which Bachmann's fictional character Mara, representative of many similar or analogous real-world cases, is a witness. Finally, the Collective Archive View is a version of relativism in Velleman's sense, holding that there are no reasons independently of communally embedded subjects. However, the view does not embrace relativism in the sense that our practical concept of a reason, the concept as we actively use it in deliberation, frames reasons as holding relative to communities. In contrast to the latter definition of relativism, the Collective Archive View explicitly holds that we take reason relations to hold simpliciter. However, that does not mean that there really are objective reasons, holding independently of what a subject in a particular time and place finds as collectively established.

The book is composed of five chapters the first of which points out the limitation of classical metaethics in making sense of the phenomenon of disrupted thought. Instances of disrupted thought must be taken seriously, since they are paradigmatic instances in which questions about the status of

normative claims and the possibility of objectivity arise. After introducing the idea of a disruption of thought, the chapter discusses how three different metaethical theories – Mackie's *error theory*, Scanlon's *realism* and Korsgaard's *constructivism* – can account for this phenomenon. While Korsgaard's constitutivist constructivism seems to be the best available candidate, this position is nevertheless unsatisfying in that it conflates different methods. Furthermore, it builds on claims that go beyond mere conceptual analysis and cannot be solved entirely within a discipline relying on pure thinking.

The second chapter elaborates on the argument in favour of a clear distinction between *pure conceptual analysis* – an analysis that philosophers often describe as apriorical – and *investigating the nature of our concepts scientifically* – thereby going beyond apriorical analysis and relying on observations of *material* reality, i.e. empirical reality in space and time. The chapter shows how such a distinction can not only make investigations of normativity more relevant and interesting, but also help to settle some disagreements within metaethical debates. More specifically, the distinction can bridge the gulf between two apparently conflicting metaethical positions, *Reasons Fundamentalism* and *Constitutivism* – alternatively describable as a conflict between *Realism* and *Constructivism*. As opposed to reading those positions as disagreeing on a metaethical issue, this chapter proposes to reconcile both positions by putting forth a Reasons Fundamentalism that takes the Constitutivist intuition seriously. Such an approach embraces Reasons Fundamentalism in metaethics and then, after sharply explicating the demarcation line, goes beyond metaethics narrowly conceived.

The third chapter investigates Skorupski's Irrealist Cognitivism as a potential version of Reasons Fundamentalism taking the Constitutivist intuition seriously. Irrealist Cognitivism holds that reasons are objective in the sense of mind-independent, but, as opposed to Scanlon's realism, the position does not claim that reasons are part of the real world. The chapter scrutinizes this mind-independence claim, based on Frege's conception of thoughts, and makes the case against *Radical Mind-Independence*, and in favour of a deflationist *Modified Mind-Independence*, interpreting reasons as *potential mental states* – potentials that are independent of an individual mind, but dependent on a *collective form of mind*.

Following the modified reading of the realm of reasons as a mind-independent realm, chapter 4 shows why we should conceptualize this mind-independent realm as a collective archive. The chapter examines how empirical theories of psychology and cognition account for the phenomena that Skorupski presupposes in his theory of subjects responding to objective mind-

independent irreals. The chapter draws on theories of *predictive coding*, the *ecological theory of affordances* and *schema theory* within developmental psychology, which all give support to the theory that reasoning is moving within a collective archive of potential mental states. The chapter concludes with an application of concepts from social metaphysics or ontology to the metaphysics or ontology of reasons. Haslanger's (2012) social constructionism as *critical realism*, it will be shown, proves fruitful for understanding the domain of reasons as a variable socially created reality.

The last chapter, chapter 5, picks up the question of transformation and describes the phenomenon of *ownership of thought*, which we risk when we embark for new normative grounds. This chapter discusses the idea of "madness" having a subversive and potentially transformative function. Finally, it suggests a concept of *retrospective justification* in order to make sense of the possibility that the future can reveal reasons that we could not possibly own at the time of acting.

The Collective Archive View in a nutshell

M: (All-things-considered or "central") reasons are items in a collective archive of mind and reasoning is a cognitive movement within a collective archive of mind.

M1: Reasons are irreducibly normative, mind-independent entities.
(a conceptual claim)

- made from an immediate "first-personal" agential perspective (taking normativity "at face value")
- establishes Reasons Fundamentalism
- *There are objective reasons for action.*

M2: Reasons are communicatively created potentials of mind.
(a material claim)

- made from a detached "third-personal" perspective (examining normativity as a natural phenomenon in space and time)
- establishes a social ontology of reasons
- *Nothing is 'simply' a reason.* (critical realism)

Chapter 1
What do we have reason to do?
Thinking about normative questions, questioning the limits of thinking

'I'm no longer sure what I think, or even if I think at all', I noted, completely thrown.

Simone de Beauvoir (1959: 344)

1.1. Introduction

What do we have reason to do? – In this book, our ability to ask and answer this question is put under scrutiny. The book asks what we are doing when we try to answer this question. With this aim, the book is placed in the field of metaethics. Metaethics, as defined by Sayre McCord (2012), is "the attempt to understand the metaphysical, epistemological, semantic, and psychological, presuppositions and commitments of moral thought, talk, and practice." As distinct from normative ethics, metaethics does not answer normative questions but examines what they imply and aim at.

In recent years, metaethics has broadened its focus from the study of morality to the study of normativity more generally conceived (see for example Scanlon 2014: 1). A central normative concept is the concept of a reason. Instead of asking ´What do we have *reason* to do?´, we can also ask ´What *ought* we to do?´ or ´What is *good* to do?´. Some philosophers argue for one of these normative concepts to be fundamental and not reducible to any of the others (see for example Broome 2018). I put this debate aside.[4] What makes me prefer the concept of a "reason" to other concepts is that it indicates the

[4] A possible merit of a distinction between something „being a reason" and something „being good" is mentioned in chapter 5.

relationship between the targets of normative reflection and the very activity of reflecting – the relationship between *reasons* and *reasoning*.

In English, we can also use the term "reason" in order to refer to reasoning as a psychological faculty, when we speak of the *faculty of reason*. In other languages, this difference is more visible. Germans, for example, use "Vernunft" when speaking of reason as a psychological faculty and "Grund" when speaking of reason as in "being a reason for". I find it fruitful to stress the relationship between something being a reason for us and reason as a capacity of our mind. Understanding the faculty of reason as the ability to deal with reasons is not only fruitful for philosophical discussion but also empirically adequate. As chapter 4 will show, this picture is in line with a more recent strand of research in the cognitive sciences, according to which reasoning is the usage of a distinctive module in the human brain, a module for intuitive inferences from "reasons", rather than being a logical or formal operation.

The field of metaethics is complex and comprises a variety of positions. Before I discuss some positions, I will start out with a problem posed by an experience – an experience that, in my interpretation, shows the limits of our individual reasoning capacities. It shatters our natural and deeply rooted belief that proper reasoning allows us to make judgments on more objective grounds – in a way that is universally valid and compelling to every creature equipped with reason. In this experience, questions about the nature of reasons become most pressing and interesting. An analysis of reasons can only be interesting and relevant, so my rationale, if it understands situations in which we start wondering about the nature of reasons – in which we start wondering whether there can be objective reasons and what objectivity exactly means.

After giving a description of this existential key experience in section 1.2., I will, in section 1.3., discuss some influential and elaborated metaethical positions and examine their potential to make sense of the experience. The most promising position, I conclude, leaves us with a puzzling ambiguity. In section 1.4., I suggest that this is due to a methodological confusion, which I hope to disentangle. I propose to distinguish sharply between what I call "metaethics narrowly conceived", a pure analysis of how we use the concept of a reason, how we relate to reasons when asking ourselves what we have reason to do, and what I call a "material approach to normativity", examining normativity as a phenomenon in the lives of biological creatures that evolved socially and historically.

Many works in metaethics, it seems, suffer from one of the following problems: Either they can only give a very limited answer to the question of what reasons are and whether there is objective truth about what we have

reason to do; or they try to give a comprehensive picture of the role of reasons but do so with the resources of armchair philosophy only and thereby become ambiguous or highly speculative. In order to provide a comprehensive picture that is precise and informative, we must go beyond metaethics narrowly conceived. However, we must acknowledge explicitly that when we examine our nature as a "material" reality – a reality in space and time that can be investigated empirically – we are taking an approach to normativity that is fundamentally different from the approach of first-personally reflecting agents, but also from the approach of philosophers employing the self-sufficient method of pure thinking.

The shattering experience I will describe reveals that the pure thinking we can directly engage in has material constituents we cannot directly see or access. Instead, we can only understand them if we approach the faculty of reason and its evolution empirically. The experience of "disrupted thought" or "disrupted reason", as I call it, is the experience of an important psychological faculty not working properly without us being able to find the mistake with ourselves. This experience, I suggest, becomes fully intelligible to us, only if we understand how reason evolved as a psychological faculty. When we understand this experience, so my hypothesis, we can really understand what we are doing when we reflect in terms of reasons.

1.2. The double nature of reason and its internal disruption

The psychological faculty of reason is characterized by two central features that are distinct but seem internally connected. For one thing, reasoning enables us to understand our own actions as justified. For another thing, reasoning enables us to justify ourselves to others. We could speak about reason having a double nature. One part of its nature is inward-directed, the other outward-directed. We can see that both parts are equally central to reason's nature when we experience what happens when these parts fail to converge. As this section will show, it is possible that an act of reasoning does not succeed in realizing both of its essential, internally connected aspects. Such acts of reasoning seem to fail in their very function of being acts of reasoning. Reason, we can say, is internally disrupted. Speaking of internal disruptions of reason is more appropriate in these situations than simply speaking of

disagreements.[5] The difference between a disruption of reason and a disagreement, as I use the term, is that in the former case, a reasoning process fails. In a disruption of reason, the reasoning process – intrapersonally or interpersonally – breaks down. Disagreement, by contrast, is in place when different people fail to reach agreement on a particular point, while preserving reasoning and argumentation in a genuine sense. It is in such situations, I contend, in which the question whether there is objective truth about reasons becomes most relevant.

In subsection 1.2.1., I point out why we should understand the faculty of reason as having a double nature, why full-fledged reasoning requires a successful use of both the inward- and the outward-directed feature. Subsection 1.2.2. gives an example of how reasoning can break down within a single person. Subsection 1.2.3., by contrast, depicts an example from politics in which reasoning breaks down interpersonally.

1.2.1. The double nature of reason

The capacity of reasoning, of reflecting in terms of reasons, plays an important part in human life. Trying to understand the nature of this psychological capacity, we can start by listing some functions that reasons serve for us. We can begin with the guiding role of reasons in everyday life planning as well as in the important decisions of life. Contemporary science, to be sure, often discounts the actual causal influence[6] of explicit reflection on our behaviour (see e.g. Haidt 2012; Kahnemann 2011; Johansson *et al* 2005). Irrespective of this debate, however, we can state that reflection is an existential psychological need for most people – reflection as a conscious activity of making sense of one's actions and decisions, of one's personal life and relation to others. Most people want to engage in activities for which they can see a reason. Moreover, understanding one's actions as done for reasons enables us to understand ourselves and the world we are living in of our own – with the power of our own reason, we could say. This seems to make us more independent of others in our decisions and judgments. We seem to become independent of situational pressures, of social circumstances and unconscious biases. A common saying holds that, once you are aware of all your reasons and know that you have good

[5] I thank David Alm for urging me to specify the distinction between disagreement and disruption.

[6] What „causation" in the relevant sense means is another huge and controversial topic that has to be set aside here (potentially relevant is e.g. List's (2016) distinction between different levels of description). I thank Björn Petersson for pointing out this differentiation.

reasons, you can "stand behind" your actions. You do no longer have to see yourself as drawn by external pressures or other people.

All the above mentioned functions of reasoning are functions that, in the first place, serve our well-being as individuals. In contrast to the functions that reasons have for individuals, we can further list some social functions. By giving reasons for our actions, we can justify our actions, not only to ourselves, but also to others. By communicating reasons for actions, we can discuss which actions are legitimate or acceptable and which are a ground for blame or even punishment. We can furthermore, it seems, reach a non-arbitrary agreement about an issue at hand and make cooperative decisions.

When we take a closer look at the fact that reasons can serve individual as well as social functions[7], it seems that this is no coincidence. It seems that reasons have these functions because of the very same feature. When we really understand what speaks in favour of our actions, we are in a better position to communicate and to justify our actions to others. Usually, justification works in both ways, to oneself and to others. To be sure, humans sometimes find ways of justifying actions, such as a lie or a theft, to themselves, without being capable of justifying this to others. However, attempts to hide it and to create a false picture, suggest that all parties in principle understand the features of the normative situation in the same way, or would understand them if they could see the full picture. Despite these local exceptions, we can probably say that, in general, most people care equally as much for both directions of justification.

It is intuitive to assume that these two aspects are connected to one and the same feature of thought, rather than constituting two separate features, which human thinking happens to possess both. It is the same feature of thought that gives us independence in our private reasoning and that makes us successful in public reasoning. Reason, we could say, displays a double nature which both separates us and connects us to other thinkers.[8]

---

[7] Many scholars distinguish between e*xplanatory*, *justificatory* and *motivating* reasons when asked to list different functions of reasons. The distinction between individual and social functions seems to be a more fundamental one, since all the three different entities must be able to serve their functions on both levels in order to be intelligible as being „reasons" at all. I thank the practical philosophers in Lund, especially Toni Rønnow-Rasmussen, Wlodek Rabinowicz and Jakob Werkmäster for making me think about these distinctions.

[8] Many of the most influential philosophical traditions acknowledge the double-nature of thought. Kant (2009), for example, admits that our capacity to use our own understanding crucially depends on the "freedom to make public use of one's reason". Locke (1980) emphasizes that, "being furnished with like faculties", we are "sharing all in one community of nature".

The probably most straightforward explanation for why private and public reasoning have a tendency to converge is that reason is a faculty that gives us an objective picture of the world. According to this picture, there are not only objective facts about what the world is like but also objective reasons for what we ought to do or, more generally, for which attitudes we ought to have. There is an objective normative reality and reason, according to this picture, is the cognitive capacity of assessing it. Thus, anyone possessing the faculty of reason to a sufficient degree will form the same normative conclusions from given facts. We could simply say that what justifies justifies – no matter whether to ourselves or to others. There just *are* reasons, which *have* a particular force on every reasonable creature.

At first glance, our experience of reason seems to involve an appeal to objectivity in this sense. This intuitive picture, however, can sometimes be challenged. Sometimes, justifications fail, and it is not immediately obvious who is at fault. Sometimes, justifications fail, and the problem cannot be solved by reconsidering one's own reasoning or improving one's argumentation. If we believe that there must always be a solution, at least in theory, to any such communicative problem, we might conclude that this solution, though existing in theory, is currently not available to us. In some situations, however, it might become doubtful whether there is a solution even in theory.

For some people, to be sure, the relation between being able to justify themselves to others and being able to "stand behind" their actions as their own decisions, fully justified to themselves, never becomes questionable. From their point of view, the reasons that are good reasons for themselves and the reasons that they can communicate as good reasons to others are the same. They can, or so it occurs to them, both endorse them and communicate them because they just *are* good reasons. Some reasons just *are* better than others, and, in principle, you are in the same position to see it as everyone else. People who believe this take the existence of an objective ground and the possibility to arrive at it by reasoning for granted.

However, there are other people who are susceptible to the fragility of reason in its double nature. These people can relate to the experience that, though you really "stand behind" your actions and know your reasons, you cannot make a case for yourself in front of others. You cannot make others regard your actions as justified, not even accept your decisions. In these cases, when you want to approach a common ground with others and get along in social life, you find yourself forced to adopt ways of acting and speaking that are not your own best judgements and that do not seem to stand for your own thinking. Members of marginalized or discriminated groups may be among the people who can relate to this experience. In modern societies, this experience becomes more and

more dominant even among people who find themselves attached to a stronger group. In societies that embrace very different cultural and social classes, the experience of clashing ways of reasoning, the experience of an inability to facilitate even a workable consensus, becomes more and more common.

Such experiences shatter our deeply rooted belief in objectivity and in the possibility of arriving at a common ground. Such experiences might even feel existentially threatening. The core of the problem seems to be the essential connection between the two aspects of reason. The connection seems so essential that one aspect has difficulties to persist without the other. We can hardly believe that our actions are justified if all other competent agents deny this. It is important to note that we must speak to a competent agent. If we understand why another person is normatively incompetent, and know others who understand this as well, disruption does not occur. If, however, we are supposed to dismiss somebody else's authority without good reasons, we get into a critical situation. Even if we can maintain our own judgement in that situation, it seems that the situation is not the same. It seems that our reasoning is in acute danger of becoming pathological. Either we must inflate our own judgment and dismiss others for no reason, or we must suppress our own judgment and deny ourselves the authority we deserve. It seems that, either way, our reasoning becomes dysfunctional. It ceases to be full-fledged reasoning. Full-fledged reasoning, it seems, presupposes the possibility of convergence with others. Rightfully believing in the justification of one's own normative beliefs is indicated by a basic success in communicating one's normative beliefs as justified. In a case of full-fledged reasoning, both aspects come together. Both aspects are equally important.

When the basic harmony between inward and outward reasoning breaks down, this can be described as an internal disruption of reason – a disruption of reason in its double nature. Such disruptions can occur in the private reasoning process of an individual as well as in a collective or interpersonal reasoning process on a political scale. In the following, experiences of disrupted reason will be studied in more detail.

1.2.2. Intrapersonal disruptions of reason

Social microcosms, such as family and intimate relationship, play a central role in enabling people to develop as reasoners. The quality of these relationships factors in mental sanity. Failures in these relationships may have a profound detrimental effect on what we conceive as the integrity and sanity ("non-disruptedness") of thought. De Beauvoir offers a good illustration of this when she describes a discussion with Sartre:

> One morning in the Luxembourg Gardens, near the Medici fountain, I outlined for him the pluralist morality which I had fashioned to justify the people I liked but did not wish to resemble; he ripped it to shreds. I was attached to it, because it allowed me to take my heart as the arbiter of good and evil; I struggled with him for three hours. In the end I had to admit I was beaten; besides, I had realized, in the course of our discussion, that many of my opinions were based only on prejudice, bad faith or thoughtlessness, that my reasoning was shaky and my ideas confused. 'I'm no longer sure what I think, or even if I think at all', I noted, completely thrown. (1959: 344)

This note describes a situation in which someone turns out to be incapable of justifying an obviously deep personal conviction to a competent other. The experience is characterized by a loss of confidence in the most basic reasoning capacities, represented in the statement "I'm no longer sure what I think, or even if I think at all".

The term "disruption of reason" can be understood most literally in cases like this. Feminist literature is full of descriptions of such disruptive experiences. More examples can be found in the work of Austrian post-war writer Bachmann who describes intellectually active women in cultural circumstances in which the very legitimacy of thinking women is denied (Bachmann 1971; 1978). A disruption happens either when the personally significant social surrounding does not take up important considerations, which show up for someone, or when the social surrounding imposes judgements and requirements, which the person cannot accept.[9]

We as distanced onlookers, living in another time, might tend towards the judgment that the described experiences are only experiences – that they do not touch normative reality as such. Regarding the cases of disrupted reason that result from sexist oppression, most contemporary readers in the west would probably hold that the women's experiences are due to intimidation and other psychological mechanisms of oppression.[10] They would not hesitate to claim that these women *have* justifications – just that other people do not recognize them or that the women are hindered in spelling them out.

[9] I will differentiate a bit more in the last chapter. Johan Brännmark expressed the worry to me that divergence with others comes in many different forms and degrees so that the picture of a "rupture", indicating a sharp cut, might be inadequate. While certainly the situations in which disruptions occur can vary, and some people may be more vulnerable to them than others, I still think that we can identify distinctive features (with both subjective and intersubjective components) that characterize a person's state *when* it occurs.

[10] See e.g. Abramson (2014) about "gaslighting", or Fricker (2007) about „epistemic injustice".

This, however, is a claim from the distance. In the moment in which the disruption occurs, this distanced standpoint is not available. Sometimes, it seems to be so unavailable that it is questionable whether we can even say that it exists. At least *for* the people undergoing a disruption of reason, the objective standpoint from which everything is clear does not exist, and that seems to be what is relevant if we want to say that there are objective reasons *for* somebody.

People undergoing such experiences can no longer reason. They can no longer recognize any reasons. Even if they insist on their convictions, stubbornly, in the face of fundamental disagreement, this comes at a cost – a cost to the health of reasoning. Being able to maintain one's convictions under such circumstances requires an inflation of one's own judgment and an indifference to the judgment of other people. The judgment of others, however, is an essential corrective – an essential part of the very nature of reasoning. For a person in this situation, the possibility of healthy thought – full-fledged thought in its double nature – is indeed unrealizable. The reasoning process is disrupted in its internal structure.

1.2.3. Interpersonal disruptions of reason

The disruption of reason is most detrimental to individual health when it occurs in the sphere of family and intimate relationships, when what is disrupted is the intrapersonal reasoning process of a subject. The phenomenon of disrupted reason can, however, also occur in an interpersonal process of reasoning, such as in the sphere of politics.

The first examples coming to our mind might be intercultural conflicts. The political scientist Samuel P. Huntington, for example, became famous for outlining what he labelled as "clash of civilizations" (1993). His thesis is that the major conflicts of our age arise from clashes between cultural and religious identities, forming separate cultural spheres further entrenched by historical and economic differences. Some authors, in opposition to Huntington (e.g. Sen 1999; Berman 2003), question the fundamentality of intercultural differences and, accordingly, doubt the impact of such differences in global conflicts. I will not investigate the plausibility of Huntington's "clash of civilizations" at this point.

Nevertheless, I can see a source of conflict that can indeed promote more fundamental clashes than intercultural differences along the lines of west and non-west, or Christian and Islamic. As opposed to the clash of different societies, there are fronts *within* many societies, which indeed warrant the claim that they build up along the lines of fundamental differences in reasoning

– differences in the very type of reasoning to the effect that successful communication in normative and justificatory terms is hardly possible. Such differences in reasoning are pointed out by so-called "moral foundations theory" in social psychology (Haidt/Joseph 2004; Graham *et al* 2009; Haidt 2012). Moral foundations theory holds that humans have at least five different foundations on which they can base normative judgments: care, as opposed to harm; fairness or proportionality; loyalty or ingroup; authority or respect; and sanctity or purity.[11]

Moral foundations theory claims that each of these foundations is basic and independent of the others. Some people rely more on some foundations than on others. One of the probably most interesting findings by moral foundations theorists is that which foundations someone primarily relies on determines one's political inclinations (Graham *et al* 2009; Sapolsky 2017). According to a number of studies performed by Graham, Haidt and Nosek (2009), moral foundations even mark the main difference between political camps such as liberal and conservative, or right and left. Indeed, the conflict potential of these discrepancies is currently increasing in many societies. It seems to split modern societies even more than so-called cultural discrepancies, captured by labels such as west and non-west.

If the discrepancies in contemporary politics are indeed largely due to different moral foundations, this brings an interesting twist into our picture of the political debates in which the two major political camps clash. As Graham, Haidt and Nosek note, there is a widespread consensus according to which "morality is first and foremost about protecting individuals" (2009: 1030; see also Turiel 1983; 2006). The authors trace this consensus in the tradition "from Kant through John Stuart Mill to John Rawls" (2009: 1030; also Turiel 2006). This conception of morality, however, only covers the first, or eventually the first two out of five moral foundations offered by human nature. The bias towards this narrow conception of morality is further reflected in most of the empirical research about "morality", which in fact only investigates empathy and altruistic concern (e.g. de Waal 1996; 2008).

The neglect of other possible foundations, especially the sources of "authority or respect" and "sanctity or purity", so the theory, leads to a misunderstanding of certain political positions. Right-wing conservatives, for example, are sometimes thought to promote incoherent positions or positions against their own interest, while in fact they are not appealing to their interests or their personal preferences at all. Instead, they argue for values which they believe to hold independently of their own preferences and interests.

[11] Haidt (2012: 215) adds "liberty" as a sixth dimension.

Accordingly, the point about defending a tradition might not be a false belief in benefits or an unwarranted fear of negative results but a distinctive value to which they are receptive, and which is not reducible to the principle of benefit and harm.

Consequently, debates between right-wing conservatives or traditionalists on the one hand and left-wing liberals or progressives on the other hand are marked, not that much by argumentative disagreements, but by an incapacity to even basically understand the same kinds of arguments. Both sides of the conflict, it seems, are responsive to entirely different kinds of reasons.[12] If there is no common ground about how to weigh different kinds of reasons, or even about which kind of considerations are reasons at all, communication is more seriously impaired than in a case in which conflicting interests clash.

The parties of the conflict, it seems, are impaired in the very possibility to engage in productive reasoning and argumentation. The relevant processes of reasoning here are collective processes. Reason as the capacity to reflect and communicate in terms of reasons is internally disrupted here as well. The disruption happens interpersonally, rather than within a subject. The cases are analogous in so far as it is in both cases unclear whether the process is still reasoning. As opposed to being systematic thinking in a person or public reasoning in a group, the situations resemble emotional disorientation and emotionally talking past each other respectively. For the individual, the effects might be less detrimental when it is part of a larger political group than when it is an oppressed and marginalized individual. The large-scale dangers connected to disrupted reason in the public sphere and on the floor of politics are, however, more than obvious.

From within the experience of disrupted reason, we may raise existential questions such as "What reasons do we really have? Is there objective truth about what reasons we have?" We want to understand what is happening in these cases, how we can make sense of our situation and how we can gain back a normative compass. The relevance and quality of a theory of reasons, I contend, is determined by its capacity of providing a convincing account of the normative features we find in a situation of disrupted reason.

[12] To be sure, there may be many other explanations for the described political tensions. This is not a sociological analysis of any actual situation. However, we can think of at least some cases in which the problem is ultimately a difference in reason-responsiveness.

1.3. Disrupted reason in light of metaethics

In the previous section, I described situations in which the activity of reason is disrupted. A full-fledged, undisrupted process of reasoning is characterized by responses to reasons to which all reasonable people can respond. The success of reasoning is endangered if one party responds to a fact as a reason that cannot be recognized as a reason by another party. In this case, the former party's reason response gets problematic – it either gets fragile and cannot be maintained, or it turns into a form of conviction with pathological certainty. But, we might ask now, is there an answer to the question what reasons there really are? How should we normatively evaluate a person in a state of disrupted thought?

In this section, I will offer the interpretations following from three different metaethical theories. All three theories share the assumption that we take normative judgments to be truth-apt. This excludes theories, so-called non-cognitivist theories, which deny this and propose to frame normative judgments as, for example, declarations or emotional expressions. I will not argue against such theories, but I will nevertheless limit my focus to theories that are more obviously in line with the profoundly irritating effects of the kind of disagreements that lead up to disruptions of thought. These profound effects, showing up in the described cases, speak in favour of views that assume an appeal to objective truth, truth visible for everyone, in our normative reflection.

However, there are different ways of theorizing about such an "appeal to objectivity", and of connecting a theory about objectivity to the ontology or metaphysics of reasons. Error theorists, for example, who agree that normative judgments are semantically truth-apt, deny that, ontologically, there are any facts that make them true or false. A normative judgment, consequently, is an error – objectivity, according to Mackie, an illusion created by a cultural process of objectification. As opposed to that, contemporary realists, such as Scanlon, argue that there are mind-independent truths about reasons. Constructivists, by contrast, hold that there are truths about reasons but that they depend on us.

In three subsections to follow, I will discuss the different proposals and argue why their answer to the person in the state of disrupted thought is unsatisfying. The most promising ideas, I will suggest, can be found in an interpretation of constructivism proposed by Korsgaard. However, it seems that some of Korsgaard's ideas are open for several interpretations and require methodological clarifications.

1.3.1. Error theory: claiming normative facts that do not exist

Error theory is a metaethical approach that combines our appeal to objective normative truth with the view that there are no objective normative facts. The first philosopher to coin the term "error theory" was John Mackie (1977). Mackie holds that, semantically, our value judgments make a claim to objective truth. Ontologically, by contrast, there is no such thing as moral facts that can make moral judgments true. Consequently, so Mackie's idea in a nutshell, all moral judgements are false.

Before pointing out Mackie's central arguments, some clarifications about scope and terminology are necessary. Mackie speaks about "values" and "moral judgments". Thereby, he picks a field of enquiry which many of those using the term "normativity" only regard as a specific type of normativity, namely "moral normativity" (see e.g. Scanlon 2014: 2).[13]

What makes Mackie's idea relevant to the analysis of normativity broadly conceived as well is that "morality", instead of being restricted to what we owe to other people, can also be defined as what we ought to do unconditionally – independently of our own wants and opinions (Mackie 1977: 27-30, Korsgaard 1996: 33). For deciding whether there is an objective normative reality, which our reflective mind must acknowledge in order to get things right, it is important to examine the possibility of unconditionally valid normative truths. Mackie examines exactly this question.

His argument that there are no objective values or moral facts allows for the possibility to make local normative statements – statements made "in relation to agreed and assumed standards" (1977: 25f). We do that, for example, when we are awarding prices or grading papers. When we want to know whether reasoning can reveal a view on reality that is really objective, and not just in line with particular local reasoning practices, we must ask whether there can be any objective facts which reasoning grasps – facts not only about what there is, but facts about what we ought to do or what attitudes we ought to form. The latter kind of facts would be genuinely normative facts.[14]

---

[13] As Scanlon observes (2014: 1), philosophical debate in the recent decades has shifted from the analysis of morality to the analysis of normativity more broadly conceived. Normativity more broadly conceived comprises all domains in which reflective assessment and reasoning plays a role. This includes ethics but also science and epistemology, which deals with norms that are usually not morally loaded. More recent works in metaethics, such as Streumer's *Unbelievable Errors* (2017), discussed in more detail in the last chapter, have reformulated Error Theory in light of normativity broadly conceived.

[14] Some philosophers believe that facts about what we ought to believe (reasons for belief), what we ought to intent (reasons for action) and what we ought to feel (evaluative reasons) can be analysed in the same way (Scanlon 2014: 1, Skorupski 2010: 2).

Mackie's argument is that such a thing as genuinely normative facts does not exist. As already said, this does not exclude the possibility to formulate ought statements in relation to a standard we assume. As distinct from that, Mackie's point is that an objective standard – one that is not just contingently assumed – cannot exist, since there cannot be objective facts tracked by such a standard. Mackie presents two arguments why it is not plausible to assume that such facts exist, the "argument from relativity" (1977: 36-38) and the "argument from queerness" (1977: 38-42).

The first argument, the argument from relativity, refers to "the well-known variation in moral codes from one society to another and from one period to another, and also the differences in moral beliefs between different groups and classes within a complex community" (1977: 36). Mackie suggests that we can understand such disagreements in a more plausible way when we see them as reflecting people's "participation in different ways of life" (1977:36), rather than as one party getting an objective truth right and the other wrong.

The second argument, the argument from queerness, has a metaphysical and an epistemological component. The metaphysical worry is that such normative facts "would be entities or qualities or relations of a very strange sort, utterly different from anything else in the universe." (1977: 38) The corresponding epistemological worry is that "if we were aware of them, it would have to be by some special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else." (ibd.) As distinct from that, there would be "nothing queer" (ibid.: 42) about normative claims raised in relation to "accepted standards" or existing "psychological constituents". We should therefore understand the grounds for normative statements in relation to such contexts and not as existing objectively.

However, Mackie insists that our ordinary moral talk and practice "involve a claim to objectivity" (ibid.: 48), which, consequently, is false. It is the product of a cultural process of "objectification" (ibid.: 42-46). "Objectification", the generation of an illusion of objectivity, presumably served human needs and purposes of cooperation.

The error theorist proposal offers a quite simple account of the state of disrupted thought. As cultures train people to regard facts as objectively existing, which in fact do not exist, the encounter of people disagreeing about these facts might result in a clash. However, both parties are equally wrong. Core parts of this proposal have a high plausibility when we approach practical normativity with the methods of anthropology and social science. But can these methods really establish the normative conclusion that all normative judgments are necessarily false? I doubt that we can satisfy the person in the state of disrupted thought by showing the cultural or personal genealogy of the

disagreement. We typically don't think our normative judgments are false just because there is a scientific explanation for why we make them that can, and must, do without the assumption of mysterious properties. Indeed, a recent proposal of an error theory (Streumer 2017) embraces the idea that, even if error theory were true, we would be incapable of fully believing it. At the end of this book, in subsection 5.2.2., I will be able to explain in detail why there is a much more plausible way to reconcile the intuitions that drive error theory; that we can investigate the world scientifically without assuming the existence of reason-making properties (one intuition), while at the same time being subjectively incapable of avoiding normative judgments (another intuition). Scanlon argues that the non-existence of ontologically queer facts in the universe is not a problem for taking normative truth seriously. In the following, I will show why his idea is plausible, without giving a fully satisfying account of disrupted reason either.

1.3.2. Realism: a theory of irreducible normative facts

A philosopher who disagrees with Mackie's conclusion is Tim Scanlon (2014). He is one of the leading voices in a more recent wave of ethical realism and defends a "realistic cognitivism about reasons" (2014: 2) – a view according to which there are objective truths about what reasons we have. The ontological question of whether normative facts really exist, so Scanlon's objection to error theory in a nutshell, is not the appropriate question for understanding reasons. Scanlon holds that truths about reasons are "fundamental". He defines this view, which he labels Reasons Fundamentalism, in the following way:

> truths about reasons are not reducible to or identifiable with non-normative truths, such as truths about the natural world of physical objects, causes and effects, nor can they be explained in terms of notions of rationality or rational agency that are not themselves claims about reasons (2014: 2).

Before taking a closer look into how this seemingly difficult idea is defended, we must understand what being "fundamental" in this sense means. The first part of the description – the irreducibility of truths about reasons – says that normativity is a *sui generis* quality of certain facts (2014: 69). It will be explained in a moment how Scanlon thinks that there is nothing queer about such facts, as Mackie supposes. The second part of the description – the inexplicability of truths about reasons – concerns the explanatory

fundamentality of the concept of a reason.[15] The two components of Scanlon's Reasons Fundamentalism are the metaphysical irreducibility of reasons on the one hand and the explanatory fundamentality of the concept of a reason on the other hand.

Mackie would object to both of these components. The first component, that there are irreducibly normative facts, is refuted with the arguments from relativity and from queerness. The second component, the fact that normative truths figure in our explanations as if they were simply a matter of fact, is taken for granted by Mackie. However, Mackie does not see our way of conceiving things as warrant for realism about normative truth. Quite the contrary: He sees the belief in objectivity, which is involved in the way we conceive of things, as supporting error theory, since our conception of things has no plausible ontological basis. Mackie complains that many philosophers are "biased […] towards various kinds of linguistic inquiry" and miss the distinction between "factual" and "conceptual" analysis (1977: 19). He bases his error theory on a discrepancy he sees between what in fact exists in the world and how we conceive of things.

Scanlon formulates sophisticated objections to both of these worries. Let us first consider his objection to the arguments from relativity and from queerness, which Mackie sees as counting strongly against the idea of irreducibly normative facts. To begin with, Scanlon does not think that normative facts must be independent from us in every sense of the term in order to qualify as objectively valid for us. He grants that, though truth about reasons is independent of our particular choices and judgments, it is not independent of "what we are like" (2014: 94). He allows for the possibility that people who differ from each other in terms of their social circumstances, interests and aims can "simply have different reasons" (1998: 71). Nevertheless, what reasons there are for a person in particular circumstances and with particular features is a matter of fact that holds objectively. Scanlon takes this to accommodate the worries raised in arguments like Mackie's argument from relativity. Often, different reactions by different people are equally correct, just because they are different people. When they conflict in a more direct way, "as when one person holds that some consideration is a reason for acting in a certain way in a certain

[15] The view that the concept of a reason is "primitive" or "fundamental" in the sense that it cannot be explained in a non-circular way is shared by many others, such as Raz (1999b: 366f), Nagel (1996: 20), Parfit (2011a: 31) and Skorupski (2010). It is not to be confuted with what is sometimes called "Reasons First View" (Star 2018) – the view that reasons are the fundamental elements within a hierarchical landscape of normative concepts. Whether other normative concepts, such as "ought" or "value", can be reduced to a question of what reasons there are, as some authors claim, will not be examined here.

situation and the other denies this" (1998: 71), it may be due to "the fact that different information or experience is available to our two groups, in which case we need to consider which of our positions is epistemically superior" (ibd.). Scanlon is generally optimistic about this method leading to a common ground. In case it does not, we can regard it as legitimate to hold different opinions. The attitude Scanlon seems to suggest is that belief in objective normative truth, when taken seriously, comes with a self-critical epistemic modesty[16], rather than being connected to what Mackie, quoting Hare, calls "fanaticism" (1977:37), the conviction that anyone can be an infallible authority on normative truth just because something strikes him or her as obviously true.

Scanlon's response to Mackie's argument from queerness takes two steps to explain. As depicted in the previous section, Mackie argues that if genuinely normative facts exist, they would be ontologically queer objects different from all other existing objects. Moreover, they would presuppose a strange perceptual faculty that is different from all our other perceptual faculties (1977: 38). This idea, so Mackie, would be incompatible with a scientific worldview. Against this, Scanlon says that, in one respect, the problem here is not really one about ontology (2014: 17). Asking whether reasons "really exist" is the wrong question. In another sense, he says, it is legitimate and metaphysically innocent to claim that "there are" reasons (2014: 25).

Let us consider the first step, which is to question the general adequacy of the ontological question.[17] According to Scanlon, it is a "misleading question" (1998: 56) to ask what kind of thing in the world a reason is, as if there was a special ontological class of things that are reasons. The things that are reasons, Scanlon says, can be very ordinary things, often natural facts, facts about the empirical world around us, sometimes facts about our psychological states (1998: 57). The colour of a hat can be a reason to buy it. The sharpness of an edge can be a reason not to touch it. "What is special about reasons is not the ontological category of things that can be reasons, but rather the status of being a reason, that is to say, of counting in favour of some judgment-sensitive attitude" (1998: 56). In order to understand what a reason is, Scanlon suggests, we do not have to look for special entities in the universe, but instead ask for what gives certain ordinary facts in the world a force of favouring certain actions for us. Scanlon says that the normative force of a fact can best be described as the relevant fact standing in a certain relation – a relation to a particular agent x in particular circumstances c, for which this particular fact p

[16] See also Bergström (1990: 42). I thank Björn Petersson for this reference.

[17] See also Putnam (2004).

is a reason for a particular action or attitude a. Scanlon (2014: 31) formalizes the phrase "is a reason for" as a four-place relation of the form R (p, x, c, a). This relation, however, is a merely formal description. It is not an explanation in any informative sense. We can simply recognize that certain facts are reasons.

Though these facts are ordinary facts, about which there is nothing odd or queer, their status of being reasons cannot be derived from any natural property of the object or state of affairs figuring as a reason. Their status is simply in virtue of their standing in a certain relation to us in this moment. Why it is this particular fact, and not another one, that is normatively relevant, why it counts in favour of this particular action and not another one, cannot be explained in terms of "what rationality requires" (1998: 30f) or "what desires someone has" (1998: 37f). These two arguments and objections to it will be discussed in more detail in this and the next chapter. Scanlon, in any case, doubts that any of these explanatory pathways can lead us to the truth about reasons. Sometimes, we might have reason to do what is most rational. Sometimes, we might have reason to follow a desire. However, these explanations are not always the correct or best explanations. Moreover, in cases in which considerations like rationality or desire satisfaction are reasons, we must still recognize them as such and cannot derive them from anything non-normative. This is one of the essential claims of Reasons Fundamentalism.

The capacity to recognize reasons correctly is, according to Scanlon, less mysterious than Mackie supposes. Since there are no queer entities out there in the world with which we must somehow "get in touch", we do not have to assume a perceptual faculty that is in any way problematic for a scientific worldview. Recognizing reasons, Scanlon says, is more a matter of "thinking about […] subjects in the right way" (2014: 70) than about getting access to some truth "outside of us". Still, reason relations hold independently of us in the sense that they are not up to our choice or dependent on whether we want them to hold (2014: 94). Likewise, they hold independently of whether we believe that they hold, i.e. whether we have actually recognized them. We can make the discovery that something is a reason and we can become aware of the fact that we were mistaken or ignorant of a reason relation that actually holds. This is meant when saying that there is objective truth about reasons. Scanlon does not think that this poses any metaphysical problem (2014:15).

The first step of Scanlon's response is, as we have seen, to question the adequacy of the ontological question understood as a question for entities that exist "outside of us" and with which we must "get in touch". A part of Mackie's criticism is, however, that there is an error involved in regarding anything as objectively being the case, when it does not exist in this ontological sense.

Mackie would probably classify Scanlon and other proponents of the project of Reasons Fundamentalism as one of those philosophers who mistake conceptual analysis for factual analysis (1977: 19). They are providing an analysis of the concept of a reason and of what we are doing when we use the phrase "is a reason for". This, Mackie argues, tells us as much about the reality of reasons, as it tells us about the reality of a perception to answer the question of "what someone is doing in saying `I perceive…´" (1977: 19).

Scanlon, who is, as mentioned before, not postulating anything incompatible with a scientific world view, holds against this kind of objection that "accepting a scientific view of the natural world does not mean accepting the view that the only meaningful statements with determinate truth values are statements about the natural world" (2014: 18). Referring to Quine (1948), Scanlon argues that our theory of the world – the physical world, important to note – involves also ontological commitments to "mathematical and logical truths as the most abstract parts of our theory of the world, which faces the tribunal of sensory experience as a whole" (2014: 18). It seems natural to say that there are for example numbers or sets and that we can make objective claims involving these concepts. However, insisting that "the physical world contains numbers and sets, in addition to particles and mountains and planets" (2014:18) would be utterly strange.

Scanlon introduces the concept of a domain[18] in order to strengthen the idea that different kinds of truth-apt claims can be made in their own right and according to their own standards:

> I believe that the way of thinking about these matters that makes most sense is a view that does not privilege science but takes as basic a range of domains, including mathematics, science, and moral and practical reasoning. It holds that statements within all of these domains are capable of truth and falsity, and that the truth values of statements about one domain, insofar as they do not conflict with statements of some other domain, are properly settled by the standard of the domain that they are about (2014: 19).

This means that, when the claim that "there is" a reason is understood in the correct way, this existence claim is metaphysically innocent. Many of the claims we make in everyday life are, however, hybrids, which involve ontological commitments in different domains (2014: 21). A reason claim involving witches or spirits, for example, is false. However, this is due to the false assumption that there is such a thing as witchcraft as a causal power in the natural world. Normative judgments that are based on such assumptions,

[18] Also used by Tait (2005), Dworkin (2011) and Skorupski (2010).

such as that the witch who made a baby die must be killed, may be in themselves correct. In other words, they may be correct insofar as the normative relations between the entities that are presupposed are concerned. The judgment as a whole, insofar as it is not purely normative, but involves claims in other domains, is still false. The causally effective natural entities that are presupposed do not exist.

Scanlon argues that certain facts are reasons in virtue of standing in certain relations to us. Scanlon is a realist about these reason relations. When we know all the relevant facts about ourselves and the world around us and think carefully, we recognize them. We, as rational creatures, are responsive to such abstract relations. This is presupposed, but Scanlon takes it as an innocent presupposition that everyone can assent to from their ordinary life experience (1998: 17). This is not to say that anyone is in a position to make an infallible judgment. It is perfectly compatible with seeing ourselves as natural organisms and our rationality – our capacity of responding to reasons – as a natural biological capacity (2014: 22).

The psychological phenomenon of disrupted reason, to conclude, does not touch the picture of objective normative facts. The phenomenon is only evidence of the profound difficulty of normative assessment, of the fallibility of all normative judgments – and probably also of our dependence on other people for using our reasoning capacities in the best way. The picture is, however, clear. In cases of disrupted reason, at least one party must be wrong – either about the truth or about the assumption that there is a relevant truth about the issue at hand.

The realist theory leaves the person in a state of disrupted reason with that stipulated assumption. I think that Scanlon's response to metaphysical worries, as they are expressed in Mackie's argument from queerness, is attractive. It is further in line with the way we relate to reasons. Usually, we have a certain security in recognizing particular facts as reasons. An appeal to a world of scientific facts that does not contain "real" normative properties does usually not impress us very much. However, what does realism have to say when our reasoning capacities lack that natural security? When the lack of security is not because the issue at hand is very difficult, but because you find yourself incapable of communicating something to others that is obvious to you? It is difficult to argue against Scanlon's assumption that there is a truth, however unbeknownst to one or both of the parties. Of course, it is possible that both parties one day arrive at a position where they can reconcile their conflicting views and discuss openly, while recognizing the normative force of the reasons given by the respective other.

However, this is, for one thing, just an assumption. For another thing, it is doubtful whether we can say that the reason relations that people recognize "one day" have been holding all along. As I will make clearer in chapter 3, it does not make much sense to speak of a reason relation as holding when the person for whom it holds must be a very different kind of person in order to respond to it. It might therefore be more appropriate in some situations to understand people reaching agreement after a long time of mutual non-understanding as having become different persons, for whom different reasons become available – reasons of which we cannot claim that they have been holding already for the persons who could in no way see them as reasons. Scanlon's response might be sufficient in cases in which an extension of knowledge or a shift of perspective helps, but seems insufficient in cases in which the very reasoning capacities of particular people are so different that they cannot communicate about certain issues the way they are. In these cases, the realist assumption is, though not proved to be false, an empty assumption.

1.3.3. Constructivism: objective normative facts that depend on us

There is a third option, apart from error theory and realism, to conceptualize objectivity in reasoning. Constructivism, as a metaethical theory, is the view that there are objective truths about reasons, but that these truths, in an important sense, depend on us. According to Street's constructivism (2008a), what reasons we have depends on our desires. Moral truth, she says, is a function of all our desires reflected on in light of each other. According to Korsgaard's constructivism (1996), what reasons we have is determined by the principles of practical reason, structural principles that are internal constituents of our very own consciousness.

In some contexts, it is unclear whether constructivism is to be understood as a metaethical position at all – a position about the ontology or metaphysics of normative facts. Instead, what some defences of constructivism seem to amount to are actually normative proposals – proposals of which kinds of reasons we should accept (given a widespread inclination to accept desire or rationality as reasons in normative discussion, but not, for example, God or any other pre-modern concept). I think nonetheless that it is possible to interpret constructivist formulations in a way that tells us something about the ontology of reasons. However, we must elaborate the distinction between normative and metanormative analysis properly.

For the rest of this subsection, I will focus on Korsgaard's proposal, which can also be specified as constitutivist constructivism. The idea that the totality

of reasons is constructed by rational principles that constitute our reflective consciousness is likely to provide us with some resources to understand the disruption of reason. Street's position, by contrast, seems to be more obviously a normative proposal – a proposal for considering no reasons as normative that we cannot trace back to a desire we have. An idea by Street that is indeed significant for metaethics will later, in section 3.4., be evaluated.

Korsgaard's constructivism, we can say, differs from Scanlon's realism in rejecting the idea that truths about reasons are "fundamental". While the realist, as portrayed by Korsgaard, says "it seems to us that there are reasons because there really are reasons", we have to say instead that "reasons exist because we need them, and we need them because of the structure of reflective consciousness" (Korsgaard 1996: 96). The objection does not target the idea that there are facts that have objective normative force for us, but against the idea that we have to presuppose that "the notion of normativity or authority is an irreducible one" and that it "is a mistake to try to explain it" (1996: 30). Normative force, Korsgaard formulates, derives "from the fact that the agent determines herself to do what is reasonable" (1996: 32). She takes this view to be subtly different from the view that normative force derives "from the intrinsic reasonableness of the action alone" (ibd.). The difference it makes for her is that normative entities are not just there, but that "normative entities" are "products of our own legislative will" and that "values are created by human beings" (1996: 112).

As already mentioned, this seemingly opaque claim does not amount to the claim defended by Mackie that "there are no objective values" (1977: 15). There are objective values, according to Korsgaard, but they are brought into being by the human activity of willing. This again does not amount to the claim that we can arbitrarily will into existence whatever reasons we want. The idea is rather that there are things we can will in a certain serious and correct way, while other things cannot be willed in such a way. A reason, we can say, is something that enables willing. We need reasons because we, as conscious reflective creatures, must act and make choices. "Human beings are *condemned* to choice and action" (2009:1), as Korsgaard writes at the beginning of *Self-Constitution* (2009), where she argues for a view that is inextricably linked to her constructivsm and that is often labelled "constitutivism" (see also Bagnoli 2017; Smith 2015; Hanisch/Baiasu 2016). This view, also shared by Velleman (2000), says that there is only a "self", a "person" or an "agent" in virtue of a reflective consciousness that is "knowing what one is doing" (Velleman 2000: 26) or that "expresses itself in choices and actions" (Korsgaard 2009: 19). When we accept this view on ourselves, we can say that the structure of reflective consciousness is *constitutive* of what we are

(ibid.). When we further agree that we need reasons in order to engage in decisive action, to succeed in acts of willing, we can say that we need reasons in order to exist. The force of reasons is inescapable for us as human beings.

This constitutivist intuition is the starting point for Korsgaard's theory of normativity. We can regard it as one of the major differences between her approach and the one by Scanlon depicted in the previous section. We can see the argument for a constructivism instead of a realism or fundamentalism about reasons as motivated by the constitutivist intuition that reasons have an existential function for us. Korsgaard supposes that understanding what that function is and how exactly reasons succeed in serving that function, is the first step in deciding what reasons we have (2009: 8). Understanding the function of a reason provides us with a standard to decide which of the considerations on which we could potentially act can count as a reason – or a good reason – and which cannot.

Korsgaard further thinks that recurring to such an internal standard – a standard that makes something a reason, Korsgaard calls it a constitutive standard (2009: 28) – is the only way to give a justified answer to the question whether something is a reason. Constitutive standards, she holds, "meet sceptical challenges to their authority with ease" (2009: 29). Korsgaard accuses the realist of answering the "normative question" – the question of why we really ought to do something – "by fiat", by just declaring that a certain fact is normative (1996: 33). Human beings, however, might have a need for understanding why that is so, why they really ought to do what they allegedly ought to do.

In *The Sources of Normativity* (1996), Korsgaard, thereby seeing herself in the tradition of Kant, refers to morality as the "search for the unconditioned", the search for a justification that makes it "impossible, unnecessary or incoherent to ask why again" (1996: 33). Korsgaard sees it as her project as a philosopher to give an answer to the normative question that has this justificatory power. For giving such an answer, Korsgaard claims, you must "place yourself in the position of an agent on whom morality is making a difficult claim" (1996: 16). The normative question as such is a "first-person question" (ibid.). It cannot be answered in a satisfying way by just giving an explanation that is "adequate from a third-person perspective" (1996: 14). Instead, a justificatory normative statement must fulfil three criteria according to Korsgaard. First, it must address someone who is in the position of demanding a justification (1996: 16). Second, it must be transparent to the agent (1996: 17). That is, the agent must understand why it is justified. Saying, for example, that a certain type of behaviour helps to preserve the species is not transparent unless someone sees the species as worthy of being preserved.

Third, the answer to the normative question "must appeal, in a deep way, to our sense of who we are, to our sense of our identity" (1996: 17).

The third criterion, we could say, is the one that also ensures the first two criteria. When something appeals to our sense of who we are, so the idea, it addresses us unconditionally and makes further justificatory questions unnecessary. It is, in a nutshell, the idea on which Korsgaard bases her whole theory of normativity. In *Self-Constitution* (2009), she develops a theory embracing three in themselves distinct topics of enquiry, which she, however, conceives as "intimately related": The "nature of action", the "constitution of personal or practical identity" and the "normativity of the principles of practical reason" (2009: 7).

"Practical reason" is here understood as a mental faculty. Korsgaard also calls it "the active aspect or dimension of the mind" (2008a: 207).[19] The major difference between Korsgaard as a constitutivist constructivist and authors qualifying as Reasons Fundamentalists is that what she calls the principles of practical reason – a standard internal[20] to this mental faculty – are seen as determining what counts as a reason. Understanding the principles of practical reason enables us to derive what reasons we have. According to Korsgaard, every motivational consideration that is supported by the faculty of practical reason, internally structured according to its principles, is a reason. The notion of *a reason* – a particular fact or consideration counting in favour of an action – is not basic. Reasons do not simply exist. The only thing we must presuppose as simply existing is the *faculty of reason* and its principles, which are constitutive of what we are as self-conscious rational beings. Reasons, consequently, are, we could say, constructs of that faculty. They are creations of the minds of rational beings. Accordingly, a constitutivist constructivist as Korsgaard sees an appropriate understanding of the rational mind as a key to deciding whether we ought to count a particular consideration as a reason. Scanlon, by contrast, doubts that a "general account of this kind could succeed" (1998: 17). Therefore, he suggests presupposing the notion of *a reason* as primitive.

Both Scanlon and Korsgaard, however, agree on the fact that normative questions have determinate answers and can be settled objectively. If this objectivist commitment is the crucial feature that warrants the label "realism", Korsgaard accepts a qualified version of the realist label as well. Korsgaard

[19]It corresponds to what other authors call the capacity of responding to reasons, responsiveness to reasons or simply rationality (Scanlon 1998: 17; Raz 1999b; Parfit 2011a: 111).

[20]„Internal" here means „constitutive", i.e. forming that in virtue of which the mental faculty of reason is the mental faculty of reason.

distinguishes between *substantive* and *procedural* realism and accepts the latter (Korsgaard 1996: 35). According to what Korsgaard calls "procedural realism", the rational procedures of arriving at a correct normative statement are real in the sense of fundamental. Substantive normative statements – statements of the sort that x is a reason – presuppose the existence of the procedures. The totality of statements we could arrive at by these procedures is determined, but they exist only in virtue of these procedures. Each time we make a normative statement, we must employ the procedures correctly in order to create or construct a correct statement – a statement that is thereby justified. Korsgaard calls the application of the correct procedures the "test of reflection" (1996: 48). The justificatory certainty, at which we arrive when a consideration survives this test, is called "reflective endorsement" (1996: 50). As opposed to substantive realists, Korsgaard claims that the reflective endorsement test is not only a path leading to normative truth, but it is "normativity itself" (1996: 48, 89). Normativity as a dimension of our human nature consists in exercising our capacity of practical reasoning correctly. Reasons, according to this view, do not really count. They are only the products of a correct exercise of our capacity of reasoning. It is this capacity, or rather its existential function in our lives, that counts in the first place. We will soon see that the "procedural realism", which Korsgaard accepts as a qualified version of realism, is still distinct from metaphysically strong forms of Kantian realism – at least if we interpret Korsgaard in line with her pronounced naturalism and existentialism.

Back to the original problem, Korsgaard's constitutivist constructivism describes very well why reason has the first part of its double nature – the function to justify our actions to ourselves. The phenomenon of disrupted reason, however, is only possible because of an essential connection between the two parts of the double nature. If everything is right, we are able to justify our judgments to ourselves at the same time as we are in a position to justify our judgments to others. The two aspects form a double nature rather than a mere composition of two features because we seem to be in those two positions at the same time in virtue of the same feature. In a nutshell, what justifies an action to ourselves is the same as what justifies an action to others.

Korsgaard indeed claims the *universal shareability* of reasons constructed according to the principles of practical reason. She supposes that considerations having "normative or justificatory adequacy" (1996: 13) for one agent, necessarily have the same justificatory power for all agents, because reasons are "not private, but public in their very essence" (1996: 134f). This idea in fact makes Korsgaard's theory a more promising candidate for making sense of the phenomenon of disrupted thought than realism and error theory, which give too simple explanations of what is happening in instances of

disrupted thought – too simple in that they make stipulations that do not help the agent in the state of disrupted thought. Korsgaard's constitutivist constructivism, by contrast, acknowledges that normativity is basic, existential and inescapable on the one hand, and what she calls "public" or "shareable" on the other hand. Thereby, Korsgaard's position is the only of the three discussed positions that gives the double nature of reason a central architectonic role in the theory.

However, the statement that reasons are "shareable" is ambiguous. One way of understanding it is to say that reasons with justificatory power necessarily have this power for every agent, while "power" here means actual power. According to this interpretation, disruptions of reason should not even occur. Showing others the power of normative considerations that we experience ourselves should be enough to capture those others.

Disruptions of reason are, however, a pervasive experience in life. If reasons with justificatory adequacy are supposed to be shareable but can in fact fail to be shared, Korsgaard must assume the validity of principles for justification irrespective of whether a concrete human consciousness experiences this validity as justifying. This move, however, would destroy one of the most central points of Korsgaard's position and make it indistinguishable from the "substantive realism" she wants to oppose. While the "procedural realism", as which she occasionally labels her position, assumes the fundamentality of the rational procedures of our consciousness for our thinking, "substantive realism" can be understood as the thesis that there are normative facts in the world independently of our particular consciousness. Some realist versions of Kant interpretation (e.g. Wood 2007; see also Arroyo 2011) indeed depict the principles of rationality as independent of our particular human consciousness. Principles, according to this understanding, would be mind-independent facts. Korsgaard, by contrast, wants to see the principles of practical reason as rooted in our mind in the most natural sense. Following the most distinctive Korsgaardian rationale, nothing – neither an action type nor a principle – can be a normative fact that would hold even in a mind- or consciousness-free universe (see also Rauscher 2002). Korsgaard's occasional usage of the term "procedural realism" should not mislead us to attribute a proper Kantian realism to Korsgaard.[21]

[21] A Kantian realist could be realist about the moral principles or about the moral facts to which the principles lead (see Rauscher 2015; Ameriks 2003). I take both versions – if they are proper realist positions – to be different from Korsgaard's "procedural realism". A "proper" realist about moral principles holds that the principles hold in the universe and are independent of the existence of particular creatures with a particular, naturally evolved cognition. Indeed, it seems more in line with what Korsgaard argues in general, to call her a "procedural fundamentalist", rather than a "procedural realist". In subsection 2.4.1., I will

If we take Korsgaard's rejection of "substantial realism" seriously, it becomes unclear in Korsgaard's picture whether people experiencing a disruption of reason are exposed to people with different forms of consciousness or whether they can be shown to make a mistake or suffer from a distortion or a defect of their own reason. The next section will discuss the ambiguity of Korsgaard's constitutivist constructivism in more detail. I show that it is impossible to eliminate the ambiguity without giving up the most distinctive part of the position – the idea that makes the position more informative than Scanlon's realism. To get the picture clear, we must distinguish the method of first-personal thinking about normativity sharply from the scientific methods that give us a material understanding of our reflective faculties. A "material" understanding of reason, as it is used here, is an understanding of reason as an empirical phenomenon in space and time, a phenomenon with a history and a variety of social and biological determinants. Korsgaard fails to make this distinction, which is probably the most important cause for the ambiguity of her position. By pointing out this ambiguity, this chapter paves the way for an analysis of reasons going from metaethics, narrowly conceived as conceptual analysis, into social ontology as material analysis – an analysis of reasons distinguishing the first-personal character of our reasoning on the one hand and its material reality on the other.

1.4. Disentangling methodological confusions

Human reason is vulnerable to internal disruption, because it displays a double nature – consisting of two essentially interconnected aspects. As the previous section pointed out, we can give different theoretical descriptions of the phenomenon of disruption. Firstly, we could say, following Mackie, that both parties make claims to objective truth about something about which there is no objective truth. This conclusion is not true to the fundamentality and inevitability that normative judgments have for us. We typically don't give up our belief in normative truth, just because we realize that a scientific picture of the world does not include queer normative facts. Secondly, we can claim, with Scanlon, that there is objective truth in virtue of normative facts, but that these

criticize the equation of fundamentalism and realism that is frequent in the literature. While "fundamental" can just mean reflectively inescapable, ontologically "real" should best be defined as existing in the universe independently of natural minds. It is obvious that Korsgaard is anti-realist in this sense, even though she accepts the (quite imprecise) label "procedural realism". I thank David Alm and Johan Brännmark for urging me to clarify this.

facts are sometimes extremely difficult to discover. The problem with this approach is not so much that it is false, but that, in some situations, the existence of objective truth seems not more than an empty assumption. Thirdly, we can claim, with Korsgaard, that there is objective truth in virtue of the principles of practical reason, which constitute us as reflective agents. This picture is the only one that acknowledges the existentially fundamental role of reasons in our lives without assuming explanatory fundamentality of reasons. As opposed to Scanlon, who is a metaphysical and explanatory fundamentalist about reasons, Korsgaard takes the idea seriously that there is a correlation between reasons and our psychological capacity of reasoning – a correlation that has implications for what reasons are.

However, Korsgaard's picture is not entirely clear. It is not clear how the situation of disrupted thought is to be evaluated in light of constitutivist constructivism. If Korsgaard wants to maintain a position distinct from realism[22], it seems that people who fail in justificatory communication either do not share the same form of consciousness or the consciousness of at least one party must be dysfunctional, defective or distorted. In subsection 1.4.1., I argue that this remains unclear in Korsgaard's work. Moreover, it is a genuine ambiguity, which cannot be clarified without destroying the most distinctive parts of Korsgaard's constitutivist constructivism. In subsection 1.4.2., I discuss a criticism of Korsgaard's understanding of metaethics brought forth by Hussain and Shah (2006).[23] While I grant a point to Hussain and Shah, I think that what the two authors portray as the nature of metaethical inquiry is a methodological approach that is in an unsatisfying way limited. Korsgaard is right in going beyond this type of metaethics, which I would like to label "metaethics narrowly conceived". However, she must make that step clear and acknowledge the distinctness of the method of metaethics narrowly conceived and another method of approaching normativity as a phenomenon in human life. Subsection 1.4.3. shows the relevance of a more comprehensive approach to understanding reasons. It motivates the step from metaethics narrowly conceived to social ontology, which this book sets out to realize.

[22] As mentioned in the previous section, Korsgaard's occasional usage of the term "procedural realism" should not mislead us to understand Korsgaard as a Kantian realist in the proper sense. Rather, she just seems to employ the term "realism" to stress the inescapability and practical fundamentality of normative truth.

[23] I thank Toni Rønnow-Rasmussen, Wlodek Rabinowicz, Jakob Werkmäster and Andrés Garcia for introducing me to this criticism.

1.4.1. The ambiguity of constitutivist constructivism

As distinct from error theorists, Korsgaard wants to maintain that normative judgments can be true or false. As distinct from non-reductive realists, she claims that this is not the case because of mind-independent facts but because of principles internal to human consciousness itself. Thereby, Korsgaard's position is the only of the three positions to take the double nature of reason, reason's existential role for the self that is essentially connected to a fundamental role in communication, seriously.

That initial strength notwithstanding, phenomena of disrupted reason – phenomena in which justification to oneself and justification to others cannot be reached simultaneously – are difficult to account for within Korsgaard's picture. As pointed out above, Korsgaard claims that justificatory adequacy that holds for one human agent always holds for all human agents. Reasons, according to Korsgaard, are public. This publicity claim must be examined in more detail to find out what happens when tensions between the private and the public get obvious.

Korsgaard rejects what she calls "publicity as objectivity" (1996: 135) and instead builds on what she calls "publicity as shareability" (ibid.). According to the conception of publicity as objectivity, what is a reason can be seen by potentially everyone because it is determined by an "objective feature of the public world" (ibid.), the world that can be seen by everyone. Publicity as shareability, by contrast, is a feature of consciousness itself. In conscious reasoning "privacy must be incidental or ephemeral" (ibid.). Our reflective consciousness, Korsgaard claims, has a deeply "social nature", which both "enables us and forces us to share our reasons" (ibid.). Korsgaard compares the relevant type of consciousness to linguistic consciousness. Building on Wittgenstein's "private language argument", she claims that "the space of linguistic consciousness is essentially public, like a town square" (1996: 139f).

One possibility of maintaining both truth and the construction of truth by the principles of practical reason is to turn realist. Realism in this Kantian picture can either be realism about normative facts, which exist independently of the principles of practical reason but can be determined by help of the principles, or realism about the principles of practical reason themselves. The latter interpretation of Kantian realism means that the principles are normative independently of a particular consciousness that works according to these principles.

Korsgaard, it seems, rules out either of these views. As distinct from mainstream Kantianism, Korsgaard wants to build on reflective consciousness in a "naturalistic" sense (1996: 160) – consciousness as we immediately experience it whenever we reflect. If her point was that the products of

conscious reflection just *are* shareable by all humans, disruption of reason wouldn't be a real phenomenon. Since it is real, we must decide how to interpret it. A possible interpretation is to say that the products of conscious reflection are not in fact shareable, but only conceived as shareable.

According to this view, we are internally necessitated to regard our judgments as shareable. That would explain very well why we suffer from fundamental disagreements and why the experience of having one's important claims rejected, ignored or misunderstood can be debilitating or even existentially threatening. However, it does not establish that there is any truth about who is right or wrong. Accordingly, normative truth would boil down to a projection of individual mind. Korsgaard could be understood as indicating this view when she admits that the relevant "kind of publicity is still inside the reflective standpoint" – a "standpoint created by consciousness" and that "from outside of that standpoint, we can recognize the fact of value, but we cannot recognize value itself" (1996: 161). We can recognize that a conscious being is valuing something, or regarding something as justified, but we can only regard it as justified ourselves, experience value directly, when we are that human being engaged in conscious activity. If, for whatever reason, we fail to share in this kind of conscious activity, there is no normative truth for us. Again, this is a phenomenon open for two different interpretations. If a person is observing the fact that another person is making a value judgment which she herself cannot share, this might be due to either different versions of consciousness or a distortion of consciousness in at least one person.

To solve this problem, it seems inevitable to understand consciousness empirically. We must examine questions such as whether all human beings, or even all conscious beings as such, share the same form of consciousness – exhibiting the same internal structures, working according to the same principles. If we can establish this, we must find ways of explaining the distortions or dysfunctionalities that lead to imperfect expressions of the structures of consciousness and, in communicative situations, to disrupted reason.

All these questions seem to involve the empirical sciences – psychology, sociology, biology and the neurosciences. To be sure, all these sciences comprise different methodologies, and thus different possibilities to investigate, to verify or falsify the proposed claims. What all these methodologies, however, have in common is that they are interested in what we could call the material nature of human cognition – its nature as an observable occurrence in space and time. Korsgaard, however, draws on neither of the mentioned sciences.

The merit of her position – the factor that distinguishes her position from realism – is, however, that she conceives of the principles of practical reason as springing from what we are as actual human beings – as human beings that are both concrete persons and biological organisms. To delineate her own theory, she says that there are, in distinction from her theory, "dogmatic rationalist theories" (2009: 5), according to which human reason, rather than being a psychological capacity in virtue of which we can think consciously, is a capacity to see what is really rational. The term "Human Reason", Korsgaard criticizes this approach, is used as if it were the name of a "module" that looks as if it were inserted into human creatures (2009: 6) – a module that is pre-existing and pre-programmed, we could say, a module that just tracks what is rational in an absolute sense. Korsgaard speaks about "homuncular theories" (ibid.) that raise the question of why humans should have developed such a module and why they should care.

Korsgaard's naturalist and existentialist leanings are what makes her theory potentially valuable for understanding reason in its double nature. However, if we take those leanings seriously, we must ask whether her view can be more than an explanation of the psychological need for presupposing shareability. If different people or groups of people cannot find a common ground, Korsgaard's principles of practical reason can hardly offer an independent justification. Nonetheless, Korsgaard presents herself as being able to establish that normative judgments can be true or false. Either she fails in establishing this or her success is dependent on substantial assumptions about the empirical nature of consciousness. At this point, Korsgaard's position exhibits a confusing ambiguity.

1.4.2. Misunderstanding metaethics?

The previous subsection has shown an ambiguity in Korsgaard's constitutivist constructivism. The ambiguity, it has been suggested, stems from Korsgaard's attempt to show that a) normative judgements can be true or false and that b) this is so because of internal principles of reflective consciousness. As pointed out in the previous subsection, it is difficult to reconcile a) and b). However, the naturalist and existentialist approach motivating b) is the most distinctive strength of Korsgaard's theory – an approach that can potentially give us a philosophical understanding of the double nature of reason, which is so centrally involved in the conflicts that make us interested in the question whether there are "objective" reasons.

In the literature, there are similar, although slightly different criticisms along the lines that Korsgaard's position is ambiguous or that Korsgaard is unclear

about what she is doing in her work.[24] Most prominently, such a criticism has probably been formulated by Hussain and Shah (2006; 2013). Although this book will depart from Korsgaard into a different direction, the criticism by Hussain and Shah can provide a suitable starting point for understanding a confusion concerning the relevant approach to normativity. Hussain and Shah namely believe that Korsgaard confuses normative ethics and metaethics (2006: 266). The two authors defend a clear division of labour between these two disciplines and hold that "not every claim about the nature of morality is a metaethical claim in any interesting sense" (2013:90). They argue that Korsgaard can at best offer a position in normative ethics and does not engage with metaethics at all. Accordingly, she fails to offer an alternative to what she calls substantive (or non-reductive) realism (2006:269). I agree with Hussain and Shah on this point. My own approach is compatible with non-reductive realism as a metaethical position narrowly conceived. However, when doing metaethics broadly conceived, metaethics as a general inquiry into what reasons are, we should not be content with the narrowly defined approach. We can combine this approach fruitfully with a more comprehensive understanding of normativity as a phenomenon in human life. Nevertheless, we should draw a clear-cut distinction.

What, to begin with, is metaethics narrowly conceived? Hussain and Shah point out the noteworthy distinction between "what *makes* an action wrong or a principle normative" and "what *constitutes* the normativity or what the property of being normative itself is" (2006: 270).

According to the relevant use of the term, what makes something normative is what explains why we ought to do it. As an example, Hussain and Shah say: "the fact that brushing my teeth regularly will reduce plague may *make* brushing my teeth good (for me)" (2006: 270). According to this terminology, what makes something normative is a fact that we can point out to somebody in order to explain or give a justification for an action or judgment. Although reducing plague can serve as an explanation in the above-mentioned case, "we do not want to claim, presumably, that the property of goodness itself just is the property of reducing plaque" (2006: 270). We can ask in a general way for "what it means to say that reducing plaque is good" or "what metaphysical commitments such a judgment involves" (2006: 271). The latter, Hussain and Shah claim, is the interest of metaethics, while the question for potential facts that could justify a particular claim is a normative question. In principle, any

[24] I am grateful to David Alm for his patience in discussing how to understand Korsgaard's position and making me aware of the many points at which clarification is needed, and at which my reading may differ from standard readings.

answer to the normative question is compatible with any metaethical account (Hussain/Shah 2013: 85).

We can now turn back to Korsgaard's project. When Korsgaard determines herself to answer the question of why we should care for the claims that morality makes on us (1996: 33), she is apparently interested in a normative question. To be sure, she asks the most general normative question, rather than searching for a justification of a particular action. However, what she asks and answers is a normative question nevertheless. Her answer to this "normative question" as she formulates it, is that "reasons exist because we need them, and we need them because of the structure of reflective consciousness" (1996: 96). As distinct from that, she takes the realist whom she wants to oppose as saying, "it seems to us that there are reasons because there really are reasons" (ibid.).

Using Hussain's and Shah's definition of the formulation "to make something normative", we could paraphrase Korsgaard as saying that what makes x normative is that, given the structure of our reflective consciousness, we need it to be normative. This, it seems, amounts to a theory in normative ethics – and, we might find, not even a very convincing one. By contrast, it fails to provide any insights into what we are doing when we answer normative questions as such, however general or particular in scope these questions might be. This means, the theory also fails to refute realism. Contrary to Korsgaard's description, realists do not claim that the fact that there are reasons rationalizes or justifies our interest in these reasons. They are examining something else, which, according to Hussain and Shah, Korsgaard does not even touch. Hussain and Shah conclude that Korsgaard's work, intended as a metaethical position in its own right, suffers from a fundamental misunderstanding of what metaethics is about (2006: 293). They consequently defend a sharp distinction between attempts to answer normative questions and metaethical analysis, which does not give any answer to what we ought to do and why.

1.4.3. Metaethics narrowly conceived and comprehensive inquiry into reasons

The previous subsection presented an argument for a sharp distinction between normative ethics, attempting to answer normative questions, on the one hand. and metaethics, being interested in the analysis of normative judgements as such, on the other hand. As Hussain and Shah formulate it:

> The point of metaethics is to give an account of what it is to think a normative thought or to show that such an account is impossible, not to tell us which

> normative thoughts to think or to point out which normative thoughts we cannot help but think. (2013: 92)

The two authors show successfully that Korsgaard, as far as she wants to "explain normativity" in the sense of giving us a justification for why we should care for it, offers us a normative claim. Thereby, she takes the possibility of normative claims for granted and does nothing to analyse their structure or implications.

As I will show in this section, there is a way of defending Korsgaard's basic approach. There is a way of defending a position holding that reasons are constructs of reflective consciousness and that outside of the standpoint of reflective consciousness nothing is a reason. However, to vindicate the core of such a position, Korsgaard must acknowledge the distinction between normative ethics and metaethics. She must accept, it seems, that a metaethical analysis cannot, and does not have to, tell people what they ought to do and why. Instead, it has to provide an analytic understanding of what this endeavour is about and how it works. Thereby, it can and should stay entirely neutral with regard to any first-personally pressing "why"- questions.[25]

Let us grant that what we are set out to do is formulating a descriptive account of normativity, which is neutral with regard to normative questions themselves. Let us now try to understand in more detail what it means "to give an account of what it is to think a normative thought", as Hussain and Shah define the task of metaethics (2013: 92). It seems that there is another ambiguity lurking at this point. According to one possible interpretation, we can very well ascribe such an account to Korsgaard. We can indeed say that, in a sense, Korsgaard offers an account of what we are doing when we make a normative judgement. In a nutshell, her account describes normative judgments as internally experienced necessities of reflective creatures – necessities which have their force in virtue of their deep link to our identity as reflective agents. What we are doing when we make a normative judgement, according to this view, is constituting ourselves as agents. We could indeed say that, while this is an analytic account of what normative reasoners are doing, it is neutral with regard to both the question of which particular actions can constitute agents and the question why we should care for being agents at all. Given that the identity-based account of normative necessities can be vindicated, it seems that such an approach would clearly form "an account of

[25] An interesting interpretation of the "normative question" can be found in Risberg (2020, Essay IV, p. 15).

what it is to think a normative thought" without conflating the task of such an account with normative ethics.

When we decide for such a reading of Korsgaard, we will immediately have to deal with another type of criticism against constitutivist constructivism, coming from Enoch (2006; 2011). Enoch, himself defending a "Robust Metanormative Realism" (2006: 195; 2013), seems to criticise Korsgaard exactly from the other direction. While Enoch has no objections to Korsgaard's apparent normative interests, he criticises her, by contrast, for failing to answer "the normative question" in a satisfying way. Disentangling the apparently contradictory approaches can help us to get a clearer distinction between the different things in which an account of normativity can be interested.

In what became famous as the "shmagency"-objection against agency-based accounts of normativity, Enoch argues against constitutivism and in favour of a "Robust Metanormative Realism" (2006: 195). His main argument is that constitutive standards cannot, as Korsgaard claims, "meet sceptical challenges to their authority with ease" (Korsgaard 2009: 29). Constitutivism, according to Enoch, cannot convince the sceptic, who can always say:

> I am perfectly happy being a shmagent – a nonagent who is very similar to agents but who lacks the aim (constitutive of agency but not of shmagency) of self-constitution. […] So what is it for me how you classify my project? (Enoch 2006: 179)

As opposed to Hussain and Shah, who argue that a "normative question" has no room in a metanormative project, Enoch obviously rejects constitutivism on the basis that it cannot give a satisfying answer to what Korsgaard calls the "normative question".

First, we must take a closer look into Enoch's argument in order to understand what he means when he says that normativity is not reducible to what is constitutive of agency. Enoch works with an argument by Rosati (2003), which is meant to refute desire-based accounts of normativity in favour of constitutivism, but which, as Enoch points out, refutes constitutivism as well – and even more obviously so. In the following, it shall become clear why we can agree with Enoch, but only in a particular sense.

Rosati, as Enoch discusses her, claims that desires are "normatively arbitrary", since agents, when deliberating and evaluating, can always step back from their desires (Rosati 2003; Enoch 2006: 178). As opposed to normatively arbitrary desires, Rosati argues, there are some features of our psychology which are constitutively necessary for being a deliberative agent. Accordingly, so the assumption, it is not possible to step back from those psychological features – exemplified by Korsgaard's principles of practical

reason. Consequently, those constitutive features of agency are supposed to be the only candidates that survive the "sceptical challenge".

The shmagent, as Enoch holds against that, is an example of how you can reflectively step back from constitutive standards of agency, nevertheless. The constitutivst can object that even the shmagent needs the constitutive standards of agency – if he wants or not. However, if complying with a standard is practically necessary in such a simple and fundamental sense, this standard is not normative in any relevant sense. If we cannot avoid following a particular standard, there is no point in saying that we ought to. If we are not already following it, or only following it to a degree, we can still ask sceptically why we ought to follow it to the fullest degree. Stepping back from an explanation of normativity, Enoch convincingly argues, is not only possible for explanations in terms of desires – it is equally possible for explanations in terms of constitutive standards of agency. Enoch even notes that "pleasures and pains – reduction to which Rosati (rightly, I think) rejects – seem much less arbitrary and much stronger as candidates for normative significance than any motives and capacities plausibly considered constitutive of agency" (2006: 182).

It seems now sufficiently clear why Enoch considers constitutivism not to provide a viable explanation of normativity. Even if some principles are necessarily applied to some degree in every act of reasoning, we do not have to take these principles *as* reasons. We do not have to accept a reference to these principles as explanation for why we ought to be moral.

We can grant this point to Enoch, it seems, and defend Korsgaard's project by saying – against her own intention though – that she does not have to answer any normative questions. Analysing what we are doing when we make normative judgments (and why they are usually important to us), is just not to be conflated with arguing for why everyone is unconditionally obligated to care for this activity (even if he allegedly does not care).

However, we can ask Enoch how a "Robust Metanormative Realism" is supposed to explain normativity instead. At this point, there is no room for an in-depth analysis of Enoch's own position (defended in Enoch 2013). Although, getting a grasp of his basic line of argument allows us to see where to place his argument for realism. As will be argued in the following, his argument is not placed on the same methodological level, or level of description, as his argument against constitutivism: If Enoch argues that constitutivism fails to convince the sceptic of normative claims by explaining them in terms of constitutive standards for agency, Enoch must – to be consistent – assume that the sceptic will be convinced by explaining normativity in terms of irreducibly normative facts. However, why should a

sceptic, who fails to be convinced by deliberating on the fact "that x is constitutive of agency", accept the fact "that x is irreducibly normative" as a valid justification? It seems obvious that no sceptic can be convinced that way unless he already understands, sees or experiences the fundamental, to himself inexplicable normative properties of x. It seems therefore that the project of convincing sceptics with metanormative analysis must be given up altogether. Metanormative theories do not answer normative questions but explain in virtue of what an answer to a normative question can be accepted. A plausible conclusion, regarding the sceptical challenge, is that, unless somebody just is responsive to certain basic normative properties without being able to ultimately explain this to himself, there is just no way – no way without using normative vocabulary – to explain normativity.

Accordingly, a theory of irreducibly normative facts, however it is spelled out in detail, must therefore be understood as a theory refusing to explain normativity – i.e. refusing to give a substantial analysis of what a normative fact, or a reason, is. A theory of irreducibly normative facts says that analyses of reasons such as "a reason is what serves desire-fulfilment" or "a reason is what is constitutive of agency" always fail because they leave room for the question whether we really ought to do what serves desire-fulfilment or what is constitutive of agency.

This argument, however, is very limited in scope. It can establish that terms such as "p is a reason" are not semantically equivalent to terms such as "p serves desire-fulfilment" or "p is constitutive of agency". If metaethics is, as Hussain and Shah have defined it, an inquiry into what we are doing when we make a normative statement, we can establish that what we are doing when we say "p is a reason" is not saying the same as "p is constitutive of agency". Understanding metaethics in that way, we can grant a very simple point to the metaethicist arguing for a form of non-reductive realism.

However, even if we cannot *linguistically* analyse the statement "p is a reason" in any informative, non-circular way, it seems that *what we are doing* when we say "p is a reason" can be analysed or examined in many other interesting ways. As Korsgaard holds, normativity is a real force in human life and "everything real can be explained" (1996: 13). We can, it seems, approach the question of what an agent making a normative judgment does without assuming that, with the most plausible explanation at hand, we can convince a sceptic that he ought to do it. What we ought to do and what we are doing when we say or believe that we ought to φ are simply distinct questions. Furthermore, there are distinct ways of analysing what we are doing when we make a normative judgment. In one way, it is plausible to say that we are responding to fundamental properties that we cannot fully explain to ourselves. In another

way, however, it seems interesting and obviously possible to step back from being an agent thinking in normative terms and investigate why and under what circumstances humans respond to properties in such a spontaneously inexplicable way. Korsgaard's intuition that this responsiveness has to do with our self-understanding as reflective creatures is at least a plausible candidate for such an explanation. It is a type of explanation, however, that leaves behind conceptual analysis and must therefore make that explicit. It must, moreover, make explicit that it is a type of explanation that draws on an empirical understanding of mind and cognition.

That normativity is irreducible and therefore cannot be explained in non-normative terms seems to be defensible only if we apply a very specific, narrow conception of explanation – a conception of explanation reduced to conceptual analysis. A prominent part of classical metaethics confines itself to this method. Paradigmatic for this approach is Moore's famous "open question argument" (1903). This methodology is clear-cut and distinct, but it might not give us too many interesting and helpful insights into the nature of normativity – insights that are fruitful for understanding complex situations. Mackie surely has a point when he criticises philosophers' bias "towards various kinds of linguistic inquiry" and claims a distinction between "conceptual" and "factual" inquiry (1977: 19). However, attempts to go beyond conceptual analysis, beyond metaethics narrowly conceived, benefit from answering the narrow metaethical question first, acknowledging the scope of the answer, and then link it to insights from investigating normativity with an empirical methodology. To be sure, empirical methodologies in themselves are very distinct and diverse. However, what I take all of them to share is that they investigate phenomena that have come about and that occur in space and time. They investigate "material" phenomena as opposed to conceptual relations we can assess with pure thinking. They investigate thinking, we might also say, from a third-personal perspective, rather than the first-personal perspective of the thinker.

A theory giving us an interesting aanswer to the philosophical question of what we are doing when we make normative judgments must bring the indispensable points of each of these approaches together into a comprehensive picture of what reasons are. But most importantly, a comprehensive theory of reasons must avoid blurring the distinction between different interpretations of the question of "what we are doing" when we make normative judgments. The narrow metaethical question is one version of this question. Another version of the question, the one that Korsgaard probably appeals to, though being sometimes ambiguous, tends towards the ontology of mind and thought as phenomena evolved in space and time.

The two senses of "what normative reasoners are doing" is brought to our attention when we face experiences of disrupted reason – when we make our normative judgements as we usually make them, but are confronted with a failure that we cannot make sense of in terms of our own conceptual apparatus. We use our own capacity of reasoning. We understand a reasonable conclusion as a conclusion shareable by other competent reasoners. Finally, we find ourselves confronted with reasoning that we cannot successfully communicate. Blaming it to the incompetence of others does not always seem an option. Likewise, finding a flaw in our own reasoning is not always possible. Reason – in the double nature, which it has for us – is internally disrupted. It is internally disrupted, however, because something external to us, or at least something beyond the capacities of our own mind, seems to disrupt or disable it. Experiences of disrupted reason invite us to examine normativity in a more comprehensive sense – to examine what our own concepts imply but also what is happening when we apply these concepts and what can happen to disable the success of these concepts. When the workings of our conceptual apparatus is disabled by a real-world situation we face, we realize that the pure thinking we engage in is *at the very same time* a material process – a process in history with social, biological and psychological presuppositions. If we wonder, from within a state of disruption, whether there are any objective reasons, we must first understand what a reason really is. It is not enough to understand how we use this concept.

1.5. Conclusion

This chapter started out with an existential cognitive phenomenon described as "disruption of thought". We are not able to maintain a process of reasoning if individually seen reasons and publicly accepted reasons fail to convergence. I presented disruptions of thought as paradigmatic instances in which questions for what reasons are and whether there can be objective reasons arise. I examined three metaethical theories with respect to how they account for these instances. I suggested that error theory and non-reductive realism (Reasons Fundamentalism) both fail to take the nature of reasoning seriously – the full nature of reasoning as it gets obvious in cases of disruption. Only Korsgaard's (constitutivist) constructivism is sensitive to the fact that human reasoners must be able to share their reasons in order to be capable of reasoning that is expressive (more precisely constitutive) of their selves.

However, read as a metaethical alternative to non-reductive realism, Korsgaard's position is ambiguous. Instead of presenting constitutivist constructivism as a metaethical position in its own right, this chapter suggested a distinction between *metaethics narrowly conceived*, in which non-reductive realism is true, and a *material approach to normativity*, with which we must combine the narrowly metaethical approach if we want to gain a comprehensive, informative picture of what reasons are. Disruptions of reason invite us to go beyond conceptual analysis and ask for the material – psychological or social – constituents and conditions of normativity. The intuition suggested in this chapter and guiding this book is that the material perspective illuminates the insights of metaethics narrowly conceived. The question how agents engaging in reasoning use the concept of a reason must be clearly distinguished from the question of what they are doing psychologically when they engage in reasoning. We must answer these questions separately. Then, we must bring the answers together into a single picture. Having the comprehensive picture and the distinct answers combined in it, we can finally evaluate disagreements within metaethics in a new light – and hopefully see that each of the classical approaches is right in at least one respect. Error theory is right in pointing out that there is a way of looking at the world according to which nothing is a reason. Non-reductive realism is right in pointing out that we cannot explain why someone ought to comply with normative claims if they are not already responsive to normative properties as such. Constitutivist constructivism is right in upholding its interest in why we have the capacity to respond to normative properties without having to "ask why again" (Korsgaard 1996: 33) and what this has to do with ourselves, our reflective consciousness and our shared human nature.

Some metaethical disagreements, so the hypothesis I suggest and hope to have made more plausible by the end of this book, can be solved by separating the *metaethical question narrowly* conceived from a more *material* question and providing a picture that contains both of them. According to the picture offered in this book, *reasons are fundamental items within a collective archive of mind* – irreducible from the first-personal perspective of reasoners and analysable as products of a "collective archival process" from a scientific third-personal perspective. The approach capturing the material nature of reasons best is, as I defend in chapters 3 and 4, a *social ontology of reasons*. In chapter 2, I argue, in more detail than already done, for *Reasons Fundamentalism* as the only plausible answer to the metaethical question narrowly conceived. The strategy of this book, arguing that reasons are fundamental items in a collective archive of mind, is to proceed from metaethics narrowly conceived to social ontology and fuse the two perspectives, while preserving their distinctness.

Chapter 2

Are reasons a fundamental part of reality?

Bridging the gulf between Reasons Fundamentalism and Constitutivism

So I will presuppose the idea of a reason, and also presuppose that my readers are rational in the minimal but fundamental sense I will presently explain.

(Scanlon 1998:17)

Morality is a real force in human life, and everything real can be explained.

(Korsgaard 1996: 13)

2.1. Introduction

This chapter analyses two distinct approaches to the question of what it means to be a reason. At first sight, these two approaches within metanormative theory seem to oppose each other when it comes to describing the relation between reasoning and reasons. One of these approaches, Constitutivism, tries to *explain what a reason is by pointing out the function of reasons* for reflective agents who engage in reasoning. The other approach, Reasons Fundamentalism holds that *what a reason as such is cannot be explained* in a non-circular way. According to Reasons Fundamentalism, the concept of a reason must be presupposed, and the activity of reasoning can, in accordance with this, only be described as responding to reasons. In a nutshell, Reasons Fundamentalism claims that we must presuppose the fundamentality of reasons

in order to account for reasoning.[26] Constitutivism, by contrast, seems to claim the reverse, namely that we must first understand what *reasoning* means for humans in order to give an account of what *a reason* is. This chapter reconciles these two metanormative approaches. More precisely, it argues that Reasons Fundamentalism is right, but that it must take a central intuition of Constitutivism seriously in order to give a comprehensive and informative account of the ontology of reasons – of the question of what reasons are and how they come into the world.

In section 2.2., I introduce the two different positions, their driving motivations and their respective conceptual frameworks. Reasons Fundamentalism, in the relevant sense[27] defended by for example Scanlon (2014), Parfit (2011a,b), Nagel (1996), Raz (1999a,b) and Skorupski (2010), will be exemplified by Scanlon's work in particular. Constitutivism, defended in different versions by for example Korsgaard (2009), Velleman (2000), Smith (1994), Setiya (2007) and Katsafanas (2013; 2018), will be exemplified by Korsgaard's constitutivist constructivism in the first place. This section reveals that Reasons Fundamentalism and Constitutivism offer two widely differing approaches to the phenomenon of normativity, in part guided by different investigative interests, rather than formulating conflicting positions on the same topics of interest. They disagree, however, with regard to the question whether reasons can be "derived" from principles of reflection or reflective agency.

[26] There are two related debates that should be kept apart here: Here, I am interested in the thesis that the possibility of reasoning presupposes that we understand reasons as fundamental. Another thesis would be that rationality must be defined as responding correctly to reasons. Broome (2007), for example, argues that we should call a person rational when she arrives at her conclusions following structural requirements for rationality, even if she gets the reason-giving facts wrong. Parfit (2011a: 113) gives examples of situations in which we have reasons for being irrational. In other words, he thinks it is possible to respond to reasons and be irrational. In this work, I am not interested in how we define what it means to be rational in these situations. Kiesewetter (2017: 161) argues, rightfully I think, that we cannot reject the conception of rationality as responding to reasons, as opposed to rationality governed by structural requirements, merely on the basis of examples involving reason-giving facts of which we are ignorant.

[27] The term "Reasons Fundamentalism", as it is discussed here, is taken from Scanlon (2014) and names a position that is not to be conflated with what is also called Reasons First Approach (Star 2018), a view according to which the concept of a "reason" is the most fundamental concept within the normative domain, i.e. more fundamental than other normative concepts such as "good" or "ought". I do not argue for such a claim in this book, though my project may count as sympathetic to it by choosing the concept of a reason as the most interesting and suitable one for discussing normativity in reflection

Section 2.3. argues that the constitutivist claim that we can derive reasons from principles of reflective agency is false – or at best misleading. Constitutivism, as exemplified by Korsgaard's constitutivist constructivism, must presuppose responsiveness to irreducibly normative entities, as exemplified by Scanlon's proposal of Reasons Fundamentalism. This section illuminates Korsgaard's picture of what she calls the principles of practical reason by comparing it to a picture of agency and self-constitution defended by Frankfurt. By understanding Korsgaard's idea of reflective endorsement in analogy to Frankfurt's wholehearted endorsement, we can reject the idea that the principles of practical reason can play a role in *deriving normativity*, while still holding that they play a role in *explaining the phenomenon of acting for reasons*. However, the relevant explanation is *material*, third-personally describing a phenomenon in space and time, as opposed to *conceptual*, making first-personal activity intelligible to a subject – a distinction suggested in the first chapter.

In section 2.4., I summarize the respective strengths of Reasons Fundamentalism and Constitutivism, which make both approaches indispensable for a comprehensive understanding of reasons.[28] Finally, I outline a way in which the two different starting intuitions, driving Reasons Fundamentalism and Constitutivism respectively, could be reconciled in a single picture. This brings me to reconsidering the metaphysics of the facts we respond to as reasons. I critically discuss the realist version of Reasons Fundamentalism proposed by Scanlon and consider an irrealist version as an alternative.

2.2. Reasons Fundamentalism and Constitutivism: conflicting positions?

There are two different approaches to the question of what it is to be a reason. According to one approach, a reason is a fact that counts in favour of a particular action or attitude. For example, we can say: The fact *that the hat is yellow* is a reason to buy it. According to this way of speaking, *that the hat is yellow* is the proposition stating the fact which figures as a favouring consideration for a particular person in particular circumstances. Scanlon, who

[28] I thank David Alm for thoroughgoing discussions about the implications of constitutivism (see also Alm 2011). Furthermore, I am grateful for the exchange with Herlinde Pauer-Studer, Christoph Hanisch, Carla Bagnoli, Federica Berdini and Xiaoxi Wu.

employs this way of speaking, claims that being a reason is an irreducible property that cannot be derived from another property, nor can it be explained informatively without using vocabulary that is itself normative. Scanlon labels this view Reasons Fundamentalism (2014).

According to a different approach, reasons are not described as facts but as internally rationalized motivational considerations. *I am dancing for the sheer joy of dancing* would be an example of a reason that is used by Korsgaard (2009:12). Korsgaard, as distinct from Scanlon, defends Constitutivism[29] about reasons. According to her view, the property of being a reason is not irreducible. She holds instead that all considerations that figure as reasons in our practical deliberation must fulfil the function of constituting us as agents. This function, so the constitutivst claim, can explain what makes something a reason. As will soon be developed in more detail, constituting us as agents means something like enabling us to understand ourselves as agents and, in virtue of that very understanding, making us reflective agents. This is what we are doing when we are reasoning, according to the picture in place.

After taking a closer look at the different starting intuitions motivating the respective approaches in subsection 2.2.1., I introduce the conceptual frameworks of Reasons Fundamentalism (2.2.2.) and Constitutivism (2.2.3.) in detail. In subsection 2.2.4., I point out the crucial disagreement between proponents of the respective approaches, which otherwise seem to differ in their investigative interests more than in their actual views about reasons.

2.2.1. Approaching normativity with different starting intuitions

There is one phenomenon that we can identify as standing in the beginning of both approaches discussed here: the phenomenon that proper reasoning allows us to make objective normative judgments – or at least judgments that are objectively better than others. There is, it seems, something that makes normative judgments true or false and we are, at least in principle, capable of finding it out. Both Scanlon, who coined the term Reasons Fundamentalism, and Korsgaard, who can be treated as a paradigmatic Constitutivist, believe that this is possible, but disagree about why that is the case (Korsgaard 2008b: 302).

[29] Korsgaard defends constitutivism (2009) and constructivism (1996) separately. As introduced in chapter 1, the two positions can be read together. This chapter means constitutivist constructivsm whenever it refers to Korsgaard's Constitutivism.

To break it down to a simple disagreement, we could say that Scanlon thinks this is the case because some facts are reasons. The relevant facts have their normative stance independently of whether we want it or recognize it. Rationality is the capacity to respond to such facts. According to this picture, the concept of a reason is basic, while the concept of rationality is defined in terms of it. Korsgaard, by contrast, thinks that normative judgments have a determinate truth-value because we can apply our rational capacities of creating reasons in a correct or incorrect way. The concept of rationality is basic, according to this view, while the concept of a reason is defined in terms of it. A reason, according to that picture, is a motivational consideration structured according to the principles of rationality. One approach considers reasons as basic, while the other considers rationality as basic. At first sight, both approaches might seem equally intuitive.

Let us first look at the basic intuition that Reasons Fundamentalists appeal to. There is a point in what Raz says when he explains the impossibility of justifying the claim that something is a reason:

> It is not easy to make sense of the very request for the justification of normativity. We can ask […] does the law constitute a binding reason for action? Do people have good reason to conform to the practices of their country? But what is it to justify a reason as such? (Raz 1999b: 366)

This would be to give reasons for a reason and this, again, would presuppose that we can recognize certain things as reasons, for, as Nagel says, "one cannot criticise something with nothing" (1996: 20). Scanlon's deliberations go into a similar direction:

> Any attempt to explain what it is to be a reason for something seems to me to lead back to the same idea: a consideration that counts in favour of it. ´Counts in favour of how?´ one might ask. ´By providing a reason for it´seems to be the only answer (Scanlon 1998: 17).

Scanlon therefore suggests taking reasons as primitive and presupposing a capacity to respond to them. Parfit, who shares the same thought, says:

> Facts give us reasons, we might say, when they count in favour of our having some attitude, or our acting in some way. But ´counts in favour of´ means roughly ´gives a reason for´. Like some other fundamental concepts, such as those involved in our thoughts about time, consciousness, and possibility, the concept of a reason is indefinable in the sense that it cannot be helpfully explained merely by using words. We must explain such concepts in a different

> way, by getting people to think thoughts that use these concepts (Parfit 2011a: 31).

This is to say that all we can do in order to justify a reason judgment is to talk about it in an illustrative way. We can point out the reasons, show them clearly and hope that the person who should be convinced – be it ourselves or someone else – will recognize that they are reasons when she gets to see things from the right perspective and reflects on the matter carefully enough. When this fails, pointing out how a proper rational apparatus "constructs" reasons, or pointing out which reasons "constitute" a rational agent properly, does not offer much help. This strengthens the fundamentalist intuition. According to this intuition, reasons are facts and rationality is the capacity of responding to such facts. As Raz formulates it: "When studying reasons we study normative aspects of the world. When discussing rationality we discuss our perceptions of, and responses to, reasons" (1999b: 363).

Let us now, in contradistinction, have a look at the intuition that Constitutivists appeal to. When we consider a worry about the fundamentalist view formulated by Velleman, we must admit that there is a point to the constitutivist intuition as well. Velleman says:

> There is a temptation to think that the norm of correctness for actions is that they should be supported by the strongest reasons. But this thought leads into a vicious circle. What counts as a reason for acting depends on what justifies action; which depends on what counts as correctness for action; which cannot depend, in turn, on what counts as a reason. Action must have an independent norm of correctness – a standard not dependent on the concept of reasons – before it can provide the sort of normative context in which reasons exist (Velleman 2000: 15f).

Instead of stating that some considerations simply are reasons, it is legitimate and meaningful to ask for the contexts in which considerations *function* as reasons. This seems to be more than just an interest of professionals in empirical psychology. It seems to play a role in our very understanding of a consideration as a reason. A consideration that figures as a reason is not just a fact favouring a particular state of the world. As opposed to that, it favours an action. It favours an action that we can, in the very course of doing it, understand as done for a reason – as meaningful and justified in a certain respect. Furthermore, it is our very understanding of an action as justified that is driving the action when it is done for a reason. As Korsgaard says, a "practical reason" has the three features of, first, *being normative*, second, *being motivating*, and, third, *being motivating in virtue of its normativity*

(2008a: 208f). While being far from uncontroversial, this threefold characterisation of reasons is at least intuitive in a certain respect.

This intuition motivates philosophers to start the enquiry into normativity not with presupposing a capacity of responding to certain facts as reasons, but with analysing the psychological capacity of taking a consideration as a reason. The rationale goes as follows: The one property that all considerations having the status of a reason – the status of being normative – have in common is the potential of driving an action – an action as an activity that is somehow "internally lucid" or understandable to an agent. Korsgaard calls actions "intelligible objects" that "embody" reasons, in the way in which sentences embody thoughts (2009: 14). According to Velleman, reasons are "considerations in light of which an action would make sense to the agent" (2000: 26). Saying that something "makes sense" is to say that it is "susceptible to explanation and understanding", that it "would provide the subject with an explanatory grasp of the behaviour" (ibid.). If this is the feature that all reasons have in common, further inquiry into this feature seems warranted. It seems we need not be satisfied with stating that some facts are reasons. We can, instead, ask for the kind of representation and processing that is involved in responding to a fact as a reason. The cognitive capacity of responding to reasons is not trivial. Understanding its procedures may well enrich our understanding of what reasons are. We may support this with the intuition that, clearly, in a world without reasoning creatures, nothing would be a reason, so that we can say that reasons are in the world only because there are creatures responsive to them. This intuition motivates the constitutivist approach.

Before we can understand where the disagreements between Reasons Fundamentalism and Constitutivism really lie, we must reconcile their different conceptual and terminological frameworks, which differ so widely that it sometimes becomes unclear whether the different philosophers talking about reasons are really talking about the same thing. While philosophers from both camps paraphrase the word "reason" as "consideration in favour of" (Scanlon 1998: 17; Korsgaard 2008a: 208), they refer to different entities in the first place. Korsgaard describes reasons as understandable intentional purposes that are "embodied" in actions, while Scanlon says that reasons are facts, often ordinary natural facts. This difference is immediately connected to another major difference, which lies in the emphasis the different approaches put on the question of motivation – the question of what happens psychologically when someone acts for a reason or actively recognizes a reason. Constitutivism places the analysis of the psychological phenomenon of acting for a reason in the centre. Reasons Fundamentalism, by contrast, argues that asking how reasons can affect us psychologically is irrelevant when we

ask for what reasons we have.[30] The following two subsections reconstruct the fundamentalist and the constitutivist picture separately, before the last subsection will pinpoint their crucial disagreement.

2.2.2. The fundamentalist picture: reasons as facts, rationality as responding to reasons

While Korsgaard's work is driven by the question of how reasons "get a grip on us" or of how they "address someone" who is in the position to ask the normative question (1996: 16), Scanlon holds that for someone who is seriously asking for reasons, it will be enough to be shown the reasons. To ask for why they should appeal to one's psychological makeup is to take the wrong stance towards reasons. When you really reason about what reasons you yourself have, you are asking straightforwardly, *what reasons do I have?*

Scanlon says that "from this point of view the question of how reasons ´get a grip on one´ properly disappears" (2014: 14). You assess the situation and look out for facts that could be relevant for you and deliberate about what these facts call for. The question of what these facts have to do with you personally does not take room in such deliberations. Reasons in what Scanlon calls the "standard normative sense" (1998: 19, 56) are

> the same kind of things that can be contents of beliefs – propositions one might say. Commonly, but not invariably, these are propositions about the natural world, that is to say about the empirical world outside us or about our psychological states (1998: 57).

In his formalization of the reason relation, Scanlon just refers to a reason as a fact p – a fact that is standing in a certain relation to an agent and a set of conditions in which it counts in favour of a particular action or attitude (2014: 31).

[30] This relates to a claim raised by many authors, according to which we must make a systematic distinction between motivational reasons and normative reasons. Such a distinction seems easy to make when using the conceptual framework of Reasons Fundamentalism but turns out as more difficult in the Constitutivist framework. The Constitutivist framework seems to be committed to a picture according to which a normative consideration and a full-fledged motivational consideration are the same – according to which, in other words, every viable motivational consideration is to a degree normative – or has at least a structural relation to the normative. This view is critically discussed as the "guise of the good"- thesis (see for example Velleman 1992). I am grateful to Hans Bernhard Schmid and Herlinde Pauer-Studer for introducing me to this debate and Toni Rønnow-Rasmussen for many critical discussions.

The idea that reasons, in the standard normative sense, are facts, must be understood in distinction from the idea that reasons in another sense are mental states. Scanlon refers to reasons in this sense as "operative reasons" (1998: 56). He says that "if we concentrate on operative reasons, then it may seem as if the only things that can be reasons are beliefs" (ibid.). A fact cannot function as a reason unless the person acting for that reason has grasped that fact as a reason. She must have formed some sort of belief containing that fact. This is undeniable. Sometimes when we ask for why someone chose a certain course of action or for what made her do such and such, we ask for the state of mind driving her behaviour. We ask for the reason that, as we could formulate it, "operated" on her. However, when we "consider the matter from the point of view of the person who has the reason" (ibid.), we will point to a feature of the situation, rather than to a state of mind. What we cite as reasons are "not beliefs but the sort of things, picked out by ´that´ clauses, that are the contents of beliefs" (ibid.). We might, for example, cite the fact *that an object has a certain colour*. Even in cases in which we refer to our own psychology, we cite the fact *that we feel positive about something*.

To illustrate the distinction between operative reasons and reasons in the standard normative sense, we might refer to Rønnow-Rasmussen's example of asking for Abraham's reason for sacrificing his son (2013): Abraham, believing that he is justified in his action will cite the fact *that it is God's will*. We, by contrast, who might not necessarily believe that God's will can justify killing (or that killing is really God's will), might refer to Abraham's mental state – to his *belief about God's will*. Like Rønnow-Rasmussen, Scanlon emphasizes a difference between giving reasons in a purely explanatory sense and reasons in a justificatory sense (1998: 19). In the former sense, we can say that reasons are mental states. In the latter, the standard normative sense, we have to say that they are facts, propositional entities picked out by ´that´ clauses.[31]

The question that, according to Scanlon, cannot be answered in any substantially informative way, is the question of what makes a reason in the normative sense a reason. According to Scanlon, it is just the fact that this fact stands in a certain relation (2014:31). Truths about whether a fact stands in such a relation or not are fundamental in the sense that there is, ultimately, no

[31] I doubt that the relevant difference here is a difference between "facts" and "mental states". We could say that in either case we cite facts: the fact *that it is God's will*, and the fact *that Abraham believes it is God's will.* However, the basic point, that propositions can be about either mind or the world, can be granted.

way of deriving them from any other facts or explaining them further.[32] The capacity to recognize such truths presupposes a rational disposition. However, the truth itself does not follow from the fact that the rational disposition is employed correctly or that a psychological need connected to that disposition is served properly.

This intuition is captured very well by Parfit's example of a person who has disturbed a snake while walking in a desert (2011a: 34). The person believes that she has a reason to run away in order to save her life. In fact, however, she must stand still, because the snake will only attack moving targets. Thus, in fact, the person has a reason to stand still. However, it appears to her that there is a reason to run away. Given that this appearance is based on a false belief, it is a merely apparent reason (2011a:35), not supported by the decisive reason-giving facts (2011a: 33). When it comes to deciding what the most rational thing for this person is, we should conclude that it is rational to run away, though there is not really a reason to run away. Rationality is based on apparent, not on actual reasons. Accordingly, rationality can require things for which there is no reason. Vice versa, it is possible to be irrational in doing what you have reason to do (2011a: 113). This argument is often cited as one of the strongest arguments in favour of Reasons Fundamentalism.[33] When asking for which reasons you have, you are asking for what Parfit calls the "decisive reason-giving facts" (2011a: 36), not for whether your mind processes impressions and impulses correctly. Accordingly, what you have reason to do is not what your rational mind requires, but what the facts require - facts to which mind can in principle but need not actually respond. Whether failing to respond is the person's own fault and whether the person can be blamed or criticized for it, is an entirely different question. When we ask for reasons, we ask for what reasons there really are, not for how we could make sure not to be blamed for irrationality (see also Scanlon 2014: 14).

[32] This does, of course, not rule out the possibility to explain some normative facts with other normative facts. For example, the fact *that car-driving emits carbon dioxide* might only be a reason against driving in virtue of the *fact that carbon-dioxide related global warming damages living beings*. In this sense, many normative facts are not "fundamental", but derived from other normative facts. This does not touch the thesis I discuss here, namely that every normative fact must be related to or derived from a *fundamental normative fact*, which is not reducible to or derivable from a *non-normative fact*. I thank Johan Brännmark for pushing me to make this clear.

[33] See also Kiesewetter 2017.

2.2.3. The constitutivist picture: reasons as essentially shareable motivational considerations

Constitutivists, such as Korsgaard, object to the view that reasons are facts. Nevertheless, Korsgaard can in principle agree with Parfit's argument that for something to be a reason for somebody, it is not necessary that this person sees it and responds to it correctly. The idea that reason relations hold without the reason being represented in a particular mind is unproblematic for Korsgaard's constitutivist constructivism. Korsgaard can hold that there is a totality of reasons that is independent of an actual representation. Nevertheless, she can insist that this totality is determined by the rational principles according to which normative considerations must be constructed in order to be normative. She can insist that all reasons, which there are, must be understood as considerations constructed in line with rational principles, rather than as simple facts. This idea is defended with what has been introduced as the Constitutivist intuition. This intuition holds that in order to provide somebody with a reason for which she can act, a consideration must be transparent to her and appeal to her first-personally reflective consciousness (Korsgaard 1996: 17). This means that, even if a reason is currently not on someone's mind, we must conceptualize the object being a reason as a consideration for which one *could* act – and for which one could act in a way that makes sense to her and that she could understand as motivated by the reason recognition.

We could say that, in order to be understood as a reason, a consideration must be viewed as an item having a graspable structure – even if it is not currently grasped by mind. According to such a picture, grasping a consideration as a reason and understanding oneself as acting on that reason, is less of a mental state than a mental event – it is a rational activity, an activity in which a rational creature becomes active as a rational creature. Reasons, we could say, must be scripts suited to drive and guide such processes when they are instantiated or represented in mind. Before clarifying this picture with an example by Korsgaard, we must understand the idea of reasons being able to drive mental processes. This idea shows the Constitutivist worry with speaking of reasons as facts: „Reality", according to Korsgaard, „is essentially activity, for all static entities are in general only the result of freezing the observer's mental frame" (2009: 37).

This quote can be interpreted as expressing a spirit that is quite common in modern science. According to a neuroscientific picture, which will be more elaborated in chapter 4, the impression that objects are just given must be understood as the result of constant selecting, filtering, processing and predicting in line with internal mechanisms, which are characteristic for the

creatures we are. Perceiving is an activity, rather than an instance of representing what is out there. The objects we perceive are, strictly speaking, not „static entities", not facts or states of affairs that we just represent. They are, in a certain respect, „made" by our mind. Their being real is, in a certain respect, due to our mental activity of, in a quite literal sense, „realizing" them. Speaking of reasons as facts, according to this rationale, is at best a simplified way of speaking about the motivational or potential motivational processes of agents. Distinguishing between actual motivational processes and potential motivational processes that would be good to have, seems to be the only way of accounting for a distinction between motivational reasons and normative reasons according to this picture.

To be sure, interpreting reasons as constructs of mind does clearly not mean that our mind just constructs objects that would not be of relevance if we did not engage in that constructive process. Of course, external reality is there and affects us no matter how we mentally process it. Reflective agents are clearly interested in this external reality when they ask for which reasons they have. Formulating a philosophical theory of reasons, however, seems to require a more thorough investigation of what such agents are doing. Formulating a philosophical theory of reasons, we must acknowledge that answering the question of what a particular external fact calls for requires more than just a realistic depiction of a natural fact. It involves an actively and subjectively shaped depiction of this fact as normative. Moreover, it involves depicting this fact as counting in favour of a particular action. Thereby, we understand the action as well as how and why it is favoured by the represented fact. When we ask for reasons counting in favour of actions, we inquire into potential courses of action that would be required by these reasons. The question for reasons is not a question for facts simpliciter, but for facts having the capacity to count in favour of actions. In order to be understood as favoured by a reason, an action must first be understandable in light of that reason. We must understand the action as such, the action as a process being internally lucid and making sense to an agent. To capture this intuition, Korsgaard says that „the reason for an action is not something outside or behind or separate from the action at all, for explicating the action, and explicating the reason, are the same thing" (2009: 14). According to this view, the entity we call "reason" is not fully described by the "fact" we cite when explaining. Instead, it is the way in which the representation of that fact rationalizes an action.

When we ask for the reason for an action, we ask for „why" someone did it. „We ask for a purpose that makes sense of the whole action" (2009:14). An action, according to Korsgaard's definition, is an „act done for the sake of an end" (2009: 11). An „act" is a description of a behaviour, a kind of formal

script, like the act of dancing described as a series of bodily movements. An „action", as opposed to that, is the enactment of the formal script driven by a subjective purpose, such as „dancing for the sheer joy of dancing" (2009: 12). The function of an action in this conceptually rich sense is not to produce something, but to make sense to the agent (2009:9). This sense-making is the existential function of acting for reasons. It is, as Korsgaard stipulates, an "end in itself" – done for its own sake (ibid.).

To compare Korsgaard's picture to Scanlon's, we could say that Scanlon employs nothing but a common way of speaking when he states, for example, that the fact *that the hat is yellow* is a reason to buy it. Korsgaard, by contrast, holds that we can only understand the propositional entity *that the hat is yellow* as a reason if we can understand an agent who states *I am buying this hat because it is yellow*. It is only because we can accept *I am buying this hat because it is yellow* as a reason, that we can speak of the fact *that the hat is yellow* as a reason for buying it. Accordingly, the way of speaking employed by Reasons Fundamentalists is parasitic on understanding reasons as internally rationalized motivational considerations. Internally rationalized motivational considerations are therefore the entities that should be called reasons in the first place. Scanlon's concept of "reasons in the standard normative sense" as well as his concept of "operative reasons" (1998:56) depend on the understanding of a consideration as internally rationalized in this sense. Going back to Rønnow-Rasmussen's example about Abraham, we can only speak about his belief concerning God's will as a reason, if we can understand how such a belief can rationalize the relevant action for somebody in Abraham's shoes – even if we do not share the belief or the reasoning ourselves. Even if we speak about reasons holding for someone who does not see them or refuses to act on them, what we are in fact talking about are motivational considerations that *would* drive the agent *if* she got it right – whatever that exactly means.[34]

To sum up, the Constitutivist picture debunks the talk of reasons as facts and replaces it by a description of reasons as internally rationalized motivational considerations. Korsgaard claims that, because of the social, public nature of our reflective consciousness, everything that is internally rationalized must be conceived as shareable with others (1996: 135f). Reasons are therefore essentially shareable motivational considerations.

[34] This picture is open for both views, the view that what we have reason to do is determined by what perfect rationality requires and the view that what we have reason to do is determined by „decisive reason-giving facts".

2.2.4. Deriving reasons? The crucial disagreement

Reasons Fundamentalism provides us with a picture of reasons as facts and rationality as the psychological capacity of responding to such facts. Constitutivism, by contrast, debunks the picture of reasons as facts as a mere way of speaking, which is parasitic on an understanding of reasons as motivational considerations that are internally rationalized. Rationalization, moreover, comes with the degree of shareability with other rational creatures, as Korsgaard claims.

We could now simply conclude that Reasons Fundamentalism and Constitutivism are two different ways of philosophically studying reasons and normativity. We could say that the former approach takes a particular way of speaking seriously and seeks to clarify what this way of speaking entails. The latter approach, by contrast, investigates the cognitive and psychological underpinnings of this way of speaking. In so far as the respective way of speaking is important to us as such, we can say that we might care for what it entails irrespective of what its underpinnings involve. This way, we could choose to see Reasons Fundamentalism and Constitutivism as two independent strands within practical philosophy rather than regarding them as conflicting positions within the same debate.

In this chapter, I will ultimately arrive at the conclusion that Reasons Fundamentalism and Constitutivism can be read as two independent strands of inquiry, which can be reconciled in a single comprehensive picture. In the current state of the debate, however, the two approaches are widely treated as conflicting positions – although often in a way that seems as if proponents of the respective approaches are talking past each other, as for example Smith noted when he described "enormous gulfs" in contemporary metaethics (Smith 1994: 3). Contemporary authors treating Constitutivism as an independent metanormative position, standing against non-reductionist or fundamentalist positions, often centre on the idea of "deriving" reasons or normativity from reflective agency. The vocabulary used by different authors in different places is, however, quite diverse. Accordingly, it is unclear what it means to "derive" a reason. Let us look at some passages before discussing how to interpret the idea. Smith, for example, defines Constitutivism as follows:

> Constitutivism is the view that we can **derive** a substantive account of normative reasons for action – perhaps a Kantian account, perhaps a hedonistic account, perhaps a desire-fulfilment account, this is up for grabs – from abstract premises about the nature of action and agency. (2015: 187)

Using a different term, Enoch describes Constitutivism as the view "that the solution to our metaethical and, more generally, metanormative problems will **emerge** from the philosophy of action." (2006). Hanisch (2016) describes the disagreement between Reasons Fundamentalism and Constitutivism (exemplified by Korsgaard's constructivism) as being about the question

> whether to assign practical principles a foundational or merely a **derivative role** in an account of practical reasons. Recall that for Raz and other "reasons-fundamentalists" principles play a subservient role relative to practical reasons, whereas for many constructivists these reasons are inconceivable unless practical principles guide processes of practical reflection regarding the inclinations that humans encounter. (2016:116)

Moreover, the term "deriving" can be found in the work of both Scanlon and Korsgaard. Korsgaard, when defending her neo-Kantian proposal against substantive realism, states that "obligation **derives** from the agent's own mind" (1996: 31), that "normativity **derives** from our self-conception" (ibid.: 249) and that "normative force **derives** not from the intrinsic reasonableness of the action alone, but from the fact that the agent determines herself to do what is reasonable" (ibid.: 32). Scanlon, who, by contrast, holds that reasons are not derivable from non-normative facts, illustrates his view by what he calls a "reflection on the nature of our own reasons (1998: 42)". When reflecting on our reasons, Scanlon says, we may state things such as *because they are my family* or *because it is worthwhile*. As opposed to that, we may seldomly or almost never cite *that it will satisfy my desire* as a reason for why we are doing something (idid.: 43f). Scanlon takes this to refute the "thesis that reasons for action can or must derive their justificatory force from the agent's desires" (ibid.: 44). In a similar vein, he takes theories deriving normativity from rationality, the mind or the will to be refuted (ibid.: 32).

Scanlon thereby seems to apply a common understanding of the concept of "derivation". When B derives from A, we can get to B by carefully thinking about A. The truth of B, we can further say, is conceptually entailed by A and follows from the truth of A. Accordingly, Scanlon denies that carefully thinking about desires, the mind, the will or the principles of rationality leads us to an understanding of which reasons we have. The only reflection that can lead us to such an understanding is, by contrast, reflection on reasons themselves. Therefore, there is no general non-normative fact from which all reasons can be derived.

This rationale is so plausible that the concept of "derivation" applied by the authors above must either be completely different from the everyday life usage

of the term "derivation" or the Constitutivist idea as such is flawed.[35] The next section will examine Korsgaard's neo-Kantian "reflective endorsement method", according to which no fact or consideration is normative in itself but only gets normative in virtue of being willed by a rational creature. By understanding this view, we can make a distinction, already promoted by Chang, between "bearers of normativity" (the "kind of considerations that can give reasons") and "sources of normativity", ("that in virtue of which" these considerations are reasons) (Chang 2009: 249). We can further distinguish between deriving particular reasons and deriving the normative force of reasons as such. After all, I think, the best might be to replace the misleading logical term "derivation" with a more psychological understanding of "originating in" or "springing from". Thereby, we can proceed towards a more appropriate understanding of the question of what is fundamental or foundational – reasons or rationality.

2.3. Responsiveness to fundamental normative entities: a presupposition for the constitutivist argument

The previous section distinguished Reasons Fundamentalism, claiming that the property of being a reason cannot be derived from any non-normative property, and Constitutivism, claiming that the normative force of a reason derives from what is constitutive of reflective agency. First, I provided a ground for comparing the two positions by translating their widely differing conceptual frameworks into each other. In Scanlon's Reasons Fundamentalism, reasons are facts – at least if we view them in what he calls the "standard normative sense". In Korsgaard's Constitutivism, by contrast, reasons are internally rationalized, and therefore essentially shareable, motivational considerations – no matter in what sense we are talking about reasons. Even if it is true that, in the standard normative sense, we cite facts as reasons, we can only apply this way of speaking because we can understand intentional motivational processes rationalized by the representation of these facts. The entity to be called "reason" must therefore be a motivational consideration of the form *I am doing*

[35] Since this is not primarily a historical or hermeneutical work, I will not engage with the vast amount of literature on "transcendental deduction" in Kantian philosophy that might be helpful for understanding Korsgaard's idea in depth. Instead, I will discuss some more common-sensical understandings to be found in contemporary discussions.

X for the sake of Y – in Kantian terms, this can be called a "maxim", in Anscombian terms, it can simply be called an "intention". Second, I examined the notion of "deriving" reasons or the normative force of reasons, which is commonly seen as the crucial point of disagreement between Reasons Fundamentalists and Constitutivists. Constitutivism, in distinction from Reasons Fundamentalism, holds that what reasons we have derives from what is constitutive of agency.

As shown by reference to Scanlon's plausible illustration of the way we reflect on reasons, this cannot mean that agents can simply reflect on the nature of their reflective consciousness and their reflective agency and will find out what they have most reason to do. Accordingly, we can say that the logical concept of derivation, which usually describes the method of arriving at a truth by considering another truth, in which the former truth is conceptually entailed, is probably not the best concept to apply in this context. This section sheds a light on what Korsgaard means when she says that normativity derives from practical rationality – or, more precisely, from what she calls "reflective endorsement", which follows the principles of practical rationality. I conclude that this picture presupposes responsiveness to substantial fundamental reasons.

In subsection 2.3.1., I describe Korsgaard's "reflective endorsement method" and argue that this method, though being interesting in many respects, fails to achieve what it is originally formulated for – answering a "normative question", a question arising from a first-personal stance towards reasons, a stance comparable to the stance we take when we ask for reasons in Scanlon's "standard normative sense". Subsection 2.3.2. shows, by comparing Korsgaard's conception of reflective endorsement to Frankfurt's conception of wholehearted endorsement, why Korsgaard's idea presupposes a responsiveness to some things as reasons, the force of which we experience as basic or fundamental. This is Reasons Fundamentalism, though it is not realism, as Scanlon frames it. Subsection 2.3.3. examines the differences between Korsgaard's Constitutivism and Frankfurt's radically subjective, anti-cognitivist and anti-rationalist account of reasons. While Frankfurt develops what has sometimes been labelled, especially by critics, a "bootstrapping account" of reasons, Korsgaard avoids the idea of radically subjective bootstrapping by shifting the bootstrapping task to the community with which we share, according to Korsgaard, our form of consciousness.

Thereby, this section paves the way for a defence of Reasons Fundamentalism that, however, is not realist and takes the constitutivist intuition seriously. According to this intuition, mind makes a world normative that, in itself, does not contain any normative properties. Furthermore, this

discussion points towards the idea that the relevant capacities of mind are essentially tied to mind as being shared with others, on which the rest of this book will elaborate.

2.3.1. Korsgaard's reflective endorsement method and the argument against theoretical knowledge

This section argues that Korsgaard must conclude from the presuppositions stated in her own argument against realism that her claim to derive reasons from the principles of practical rationality fails. The principles of practical rationality can neither tell us what reasons we have nor can they help us to assess whether a reason we see is a good reason. Constitutivism, to make a long story short, does not provide an answer to the „normative question", the question of why we really ought to do what is normatively required. Constitutivist principles do not, as Korsgaard stipulates, „meet skeptical challenges to their authority with ease" (2009: 29).

The reason for this lies in a subtle difference introduced by Korsgaard herself – the difference between answers that provide merely „theoretical pieces of knowledge" (2008b: 315) and answers that appeal to „our sense of who we are" (1996: 17) and thereby activate a kind of „practical knowledge" - a kind of effective and embodied understanding of how to act and why, a knowledge that cannot be conveyed by theoretical explanation, but must be presupposed. Korsgaard objects to realism that pointing to the normative stance of facts is to answer the normative question „by fiat" (1996: 33), by just stating a fact, as in reporting the results of a scientific experiment or drawing a map (2008b: 315). Ethics, however, Korsgaards reminds her readers, is supposed to give practical guidance and enable people „to use the map" (ibid.), to tell them when, why and how to perform a certain step (2009: 17). An answer to the normative question can therefore never be a merely theoretical piece of knowledge but must appeal to people first-personally and effectively, so the basis of Korsgaard's argument. We can now ask further for what it means for an answer to appeal to people in this way, to get an effective grip on them.

Scanlon denies that the possible answers we could give in different situations to different people are of any general kind (2014: 44; 1998: 17). There are different kinds of reasons. As a rational creature, you can potentially see all of them as reasons. When you see one of them as a reason, there is no longer a need for a further answer to the normative question: „the ´grip´ that a consideration that is a reason has on a person for whom it is a reason is just being a reason for him or her" (2014: 44). Korsgaard, by contrast, proposes that all reasons that have normative authority for us relate to our conception of

our „practical identity“ - to a description under which we value ourselves, under which we find our lives to be worth living and our actions to be worth undertaking (1996: 101; 2009: 20).

It seems like an obvious and easy objection to this idea that often, if not most of the time, the fact that an action preserves one's identity or feeling of self-worth is not the right kind of reason to justify the action. Korsgaard obviously denies that this is how she wants her claim to be understood. Picking up Kant's conception of „acting from duty“, Korsgaard emphasizes that acting from duty does not mean that acting from duty is the purpose, as in *I visit my friend in order to do my duty* (2009: 11). Rather, the action must be *I visit my friend in order to help him*. The duty, according to that picture, is „the whole package“ (ibid.), the right act for the right end or purpose. How should we now interpret the claim that reasons must be related to our conception of practical identity in order to be normative? The interpretation that is clearly inappropriate is the idea that the fact *that something is related to one's identity* is the reason that justifies everything or that is at the end of every justificatory chain. Enoch (2006; 2013), in his famous “shmagency”- objection to constitutivism, which was discussed in chapter 1, has convincingly illustrated the absurdity of such an answer to sceptics.

To avoid this absurdity, the only plausible interpretation of Korsgaard's claim is to understand the appeal to identity as the capacity to activate practical knowledge, to effectively activate knowledge that is somehow non-propositional, that is implicit and deep-seated in the sense that it cannot occur on our consciousness, but constitutes our consciousness itself. Consequently, we must say that Korsgaard's argument against „theoretical pieces of knowledge“ explicitly forbids to understand the relation to practical identity as an „explanation“ in a theoretically informative sense (2008b: 315). In other words, referring to identity does not explain anything to an agent. Reference to identity, agency or practical reason does not and cannot settle an agent’s “normative question”.

What a reference to practical identity explains is, by contrast, what happens when an agent accepts a particular answer to a normative question as plausible. What happens when someone finds a normative explanation plausible is that this explanation constitutes a piece of knowledge which succeeds in activating practical knowledge – which succeeds in appealing to the person in the right way. The structures of practical knowledge, the structures of active consciousness, are not seen or understood by the person, they only constitute her understanding. The structures of practical knowledge do not occur in justificatory conversation or argumentation. They are just presupposed in the process of recognizing a normative explanation of any kind as justified.

Reasons Fundamentalists presuppose this from the very beginning (e.g. Scanlon 1998: 17).

Ethics, we could say in that spirit, must present pieces of knowledge that appeal to our sense of who we are. If they do not have this appeal to someone immediately, we must fill out the picture with more pieces or present the pieces in such a way that the agent can see their appeal. This is similar to Parfit's idea that fundamental concepts, such as that of a reason, cannot be „helpfully explained merely by using words" (2011a: 31). Instead, Parfit suggests, we „must explain such concepts in a different way, by getting people to think thoughts that use these concepts" (ibid.).

To be sure, accepting a consideration as a reason requires that it is supported by the principles of practical reason – these principles being understood as embodied knowledge constitutive of our conscious thought. This embodied knowledge can be activated spontaneously by various roots in various situations, but there is not one specific route to activating it under all circumstances. No sceptic can be convinced by the „theoretical piece of knowledge" that there are certain things that a reflective creature would endorse and that therefore he ought to endorse them. This idea would fly in the face of all the objections that Korsgaard herself raised against „rational dogmatists" defending „homuncular theories" of the faculty of reason (2009: 6). If you lack the capacity to see a reason all by yourself, explicating formal principles to you does not help, though it might be true that when you see a reason, this process implicitly or naturally follows certain principles. This can count as entailed by Reasons Fundamentalism and it seems that Korsgaard must accept it when she wants to take her own argument against realism (especially 2008b: 315) seriously.

If we say that the appeal to practical identity explains why an agent is responsive to a particular "piece of knowledge", which then is no longer a merely "theoretical piece of knowledge" for her, we must acknowledge that this is an explanation in another sense. It is not giving a justification to an agent who asks what reasons she has. Depending on how much the agent cares for her identity, a reference to identity might lend additional support to particular reason claims. However, an agent, from her normatively interested and engaged first-personal standpoint, cannot derive any substantially action-guiding reasons from understanding her identity.

Avoiding the misleading concept of derivation, we can nevertheless make sense of Korsgaard's idea that reasons, or value generally speaking, in a way "springs from" or "originates in" what she calls "reflective endorsement". Reflective endorsement, we could say, is the "source of normativity", the event that brings normativity into the world. Serious and careful reflection, so

Korsgaard, can sometimes "destroy knowledge" (1996: 86). If a particular piece of knowledge, or belief, however, survives the "test of reflective endorsement" (ibid.: 89), we are justified in holding it. The test of reflective endorsement, important to note, is not only a way of arriving at a justified statement. Reflective endorsement, Korsgaard says, "is normativity itself" (ibid.: 48, 89). This means that, strictly speaking, there is nothing normative in the world than just selves which reflectively endorse a normative belief. It is the capacity of endorsing beliefs as justified on careful reflection – a justification we might experience as unconditional and as stemming from our deep selves – which gives rise to a normatively significant reality. There is nothing of meaning or value in the world apart from such subjective instances of valuing, we might paraphrase.

This seems to be the main message that Korsgaard wants to convey with her claim that normativity derives from our self-conception, our will or the principles of practical reason. If this is correct, she must acknowledge that these principles are not a tool an agent can use as explicit knowledge in a derivation of which reasons she has. The principles of practical reason, by contrast, are a form of implicit or embodied knowledge which allows us to endorse certain actions and not others. Which actions these are therefore *follows*, in a *genealogical* sense of the term, from the nature of practical reason. Nevertheless, it is impossible for us to *logically derive* it from facts about the faculty of reason. The faculty of reason, by contrast, must be provided with substantial content – particular "pieces of knowledge" or items of thought that are capable of activating it. These substantial items, when we encounter them, will be fundamental and inexplicable to us. To be sure, the normative force of some facts can be explained by others. Ultimately, however, we must be responsive to a fact as a reason. Such a fact is encountered as underived – as just given. This drives us to the conclusion that, for solving the first-personal justificatory question, Constitutivism must embrace Reasons Fundamentalism – even if not the realist brand offered by Scanlon. The next two subsections take up with the task of describing how a version of Reasons Fundamentalism must look like in order to capture the idea behind Korgaard's conception of reflective endorsement.

2.3.2. Subjectivist Reasons Fundamentalism? "Reflective endorsement" as "wholeheartedness"

So far, it has been argued that Korsgaard's method of reflective endorsement is not only compatible with Reasons Fundamentalism but presupposes it. Reasons Fundamentalism, by definition, is the idea that, from the stance of a

normatively reflecting creature, it is impossible to explain what a reason is and to derive which reasons we have. This, however, does not necessarily entail realism. There is a further step from Reasons Fundamentalism to realism.

Scanlon, who proposes realist Reasons Fundamentalism, seems to put most of his work into defending Reasons Fundamentalism, rather than metaphysical realism. To be sure, he presents the idea of different domains of fundamental facts, such as the domain of normative facts, the domain of physical objects etc., by help of which he circumvents the objection that there are no normative properties in the physical world. Scanlon might be right that the normative domain is equally fundamental to human existence as the physical world. However, we can still question the claim that this makes the domain of reasons a reality simpliciter – as opposed to a constructed, gradually shaped and eventually changeable reality. To be fair, Scanlon does not explicitly reject the latter picture. However, he does not consider it as relevant either.

As opposed to that, Korsgaard's interest in the relation between ourselves and the fundamental items to which we respond leads us into a more critical and interesting examination of what reasons are – compared to just saying that they are a part of our human reality. This makes the Constitutivist approach valuable even if its questions are usually unnecessary to ask in practical situations, and even if Constitutivism cannot provide answers that can convince a "moral sceptic". As distinct from Scanlon's realist Reasons Fundamentalism, the Reasons Fundamentalism presupposed by Korsgaard's reflective endorsement view is more similar to a view proposed by Frankfurt – a view that usually counts as radically subjectivist, as it stresses the normative arbitrariness of the things that might function as reasons for us.

This subsection proposes that Korsgaard's "reflective endorsement" should be understood as a phenomenon very similar to Frankfurt's "wholeheartedness". Both theories, so the argument, are theories about how human psychology can endorse seemingly arbitrary things as reasons. Both theories develop an idea of how we can take things as reasons without being able to logically derive them from any prior fact or principle. A subtle, but nonetheless crucial difference between Frankfurt and Korsgaard will be discussed in the next subsection. By making this difference, it will be argued, Korsgaard avoids one of the most frequently levelled objections against Frankfurt's radically subjectivist approach to how humans endow arbitrary things with value.

Let us now examine the parallels between Korsgaard's "reflective endorsement" and what Frankfurt (1987) calls "wholeheartedness" or wholehearted endorsement. While there is no need to spell out Frankfurt's

theory in detail[36], we can say that the main problem he has set himself out to solve is the problem of how we can achieve the volitional unity we need in order to be "autonomous" agents (see also Frankfurt 1971). Frankfurt here works with a conception of autonomy that has, at least in the first place, nothing to do with political autonomy or autonomy in relation to others. Autonomy, in this context the same as freedom of the will, refers to the subjective state of conceiving of one's will as one's own. Autonomy, in this sense, is the state in which a person really wants what she wants and the question whether what she wants is the result of external influences, manipulation or social pressure does not come up.

Frankfurt calls this state the wholehearted acceptance or endorsement of a wish as one's own. As Frankfurt describes that state, it is similar to the state in which Korsgaard envisions the agent who has found an appropriate answer to the normative question – a state in which it is "impossible, unnecessary or incoherent to ask why again" (Korsgaard 1996: 33). What Frankfurt, and it seems also Korsgaard, are after, is the phenomenon of perfectly unified or wholehearted willing which is only possible if there is a reason that settles all further questions. According to Frankfurt, no such reasons can be found by rational deliberation or derived from principles of either rationality or morality. As opposed to that, reasons – at least the reasons which really count for us because they are the only ones that enable wholehearted willing – are the product of identification or caring, according to Frankfurt (see also 1982; 1998). Caring about something, in Frankfurt's framing, consists in a strong and effective attachment of our will to an object. Effective attachment means that, whenever the object is threatened, our will is going to be moved by that. Unless external force hinders us, we will act to preserve or protect the object to which we are attached. The objects we effectively care for can, as Frankfurt says, create "volitional necessity" (1987) – a phenomenon of being necessitated not by external pressure but by the internal constituents of one's very own will or self.

Objects having that function for the psychology of a particular person cannot be specified further, according to Frankfurt's theory. They may include persons, abstract ideals, activities or physical objects. They may often include things that are "good" for us in an objective sense and things that are "morally good". However, they might as well be things that are morally or in some other sense "bad" (see Frankfurt 1982). Frankfurt has no theory of what makes things good or bad and is neutral with regard to the question if there is such a thing as objective value. What he stresses instead is that whatever we may rationally

[36] For a detailed discussion see my master thesis Mähringer (2015).

or morally judge to be good cannot give us a reason – unless, of course, rationality or morality are the ideals we most care for. Most frequently, however, the objects we most care for – the objects that are capable of necessitating our will – are persons we love, such as our children, partners or parents. Frankfurt (2004) speaks of love as one of the main sources of reasons.

With the presented setup, Frankfurt defends a view that regards reasons as subjective and contingent entities coming into existence by strong personal attachment. As such, reasons can be completely different for different people and each person could have an entirely different set of reasons if she had been brought up differently or if she had met different people in her life. For these reasons, Frankfurt's position can be considered as one of the most radical formulations of subjectivism about reasons. Since Korsgaard emphasizes universality, which Frankfurt explicitly rejects, it might seem at first sight, that Korsgaard's position is very different in spirit. Taking a closer look, however, we can discern a basic analogy of Korsgaard's and Frankfurt's position.

Indeed, Korsgaard embraces the radical contingency claim defended by Frankfurt when she says that we must „carve out a personal identity" and treat our "contingent identities as the sources of absolute inviolable laws" (2009: 23). Korsgaard holds that „making the contingent necessary is one of the tasks of human life and the ability to do it is arguably a mark of a good human being" (2009: 23). In a paper titled *Morality and the logic of caring* (2006), Korsgaard explicitly reveals her sympathy for Frankfurt's conception of agency (2006: 55). In this paper, she partly agrees and partly disagrees with Frankfurt. She agrees insofar as she emphasizes that our sources of inviolable laws are contingent. As opposed to that, she disagrees with the radical subjectivity of Frankfurt's account of caring – or more precisely with the claim that strictly speaking anything can serve as a source of reasons. „Caring" – the motivationally effective attachment on which Frankfurt builds his theory – has an internal „logic" according to Korsgaard (2006: 57). Korsgaard argues against Frankfurt in two ways.

First, she says that the „logic of caring" entails an interest in diachronic stability or sustainability of the activity of caring (2006: 60-62). Really caring for something means assuming that one's future self will still be able to care. So, caring cannot depend on just any kind of current psychological impulse, but on seeing the object of caring as really worthy of being cared for. Only if we assume that a rational creature as such can see this worthiness, we can assume that our future self, which might no longer be in the contingent grip of the relevant impulse, can see a worth in caring. Secondly, Korsgaard argues – and here she explicitly refers to Scanlon (Korsgaard 2006: 73) – that taking one's personal attachment as a basis for one's caring is „the wrong way" of

caring for something or someone. If „I felt that my child's right“ [to care] „derived from my love for her“ […] „that would be the wrong way of caring about her (Korsgaard 2006: 73; see also Scanlon 1998: 164f). With this example, Korsgaard wants to show that, although our reasons arise from contingent caring relations, we must be able to see a worth in the caring which others can see as well.

While we can establish that Korsgaard buys into Frankfurt's contingency claim, we can see that she combines contingency with a way of transcending the radical subjectivity claimed by Frankfurt. Korsgaard, as opposed to Frankfurt, says that „in valuing ourselves as the bearers of contingent practical identities, knowing, as we do, that these identities are contingent, we are also valuing ourselves as rational beings“ (2009: 24). In a nutshell, Korsgaard says that in order to wholeheartedly will or care for something, we must assume that it is worthy of our care independently of us and that other rational beings can see this worth as well. This, it seems, is a way of describing the same phenomenon as Frankfurt, but disagreeing about one part of the description (see also Mähringer 2015: 92-104). Reflective endorsement in Korsgaard's usual terminology, we can conclude, is in principle the same phenomenon that Frankfurt picks out as wholehearted willing. Korsgaard's departure from Frankfurt consists in the way Korsgaard accounts for the possibility of wholehearted willing. While some authors insist on a distinction between psychological possibility and possibility in the sense of normative permissibility, the cited passages strongly suggest that Korsgaard is describing a feature of motivational psychology in the first place. It is, furthermore, a claim about the nature of consciousness, which she elsewhere describes as "essentially public, like a town square" (1996: 140). This thesis must be specified beyond what Korsgaard says about it. The next subsection shows, moreover, how sharing reasons with other creatures of the same kind can avoid what many philosophers have criticized as the "bootstrapping problem" of accounts which embrace the personal contingency of reasons.

2.3.3. Radical subjectivism and the problem with bootstrapping

The previous subsection has shown that Korsgaard's "reflective endorsement" is similar to Frankfurt's "wholeheartedness" in that the substantial content which is endorsed is entirely contingent and has no intrinsic normative features. By understanding this idea, we can understand how mind can endorse things as reasons without being able to logically derive these reasons from a

prior fact or principle. Everything that makes something a reason is the fact that it is reflectively endorsed or wholeheartedly willed.

Constitutivism presupposes Reasons Fundamentalism, as argued in subsection 2.3.1. However, the version of Reasons Fundamentalism entailed in Constitutivism is not a realist one. It is a theory according to which the items that are fundamental to our practical reasoning are arbitrary objects, which came to matter for us because we developed relations to them. As distinct from Frankfurt, however, Korsgaard claims that what we can reflectively endorse in the fullest sense is not completely arbitrary. As opposed to that, it must be basically "moral", which means, in Korsgaard's understanding, "shareable" by our fellow creatures. This section elaborates on this essential difference between Korsgaard's view and Frankfurt's radically subjectivist position. It shall become clear how Korsgaard, as distinct from Frankfurt, regards an intersubjectively shared realm of fundamental items as a necessary condition for the endorsement of a reason. With this proposal, it will be shown, Korsgaard avoids the bootstrapping problem – a strong objection that has frequently been levelled against Frankfurt-style accounts (by e.g. Herman 2002; Wolf 2002).

So-called "bootstrapping accounts of reasons" (Herman 2002: 257) build on the idea that, as creatures in the middle of a non-normative world, for whom nothing is a reason, we can simply will a normative reason into existence. We do so by "bootstrapping" reasons out of nothing – by mere existential choice and spontaneous commitment. In a certain respect, this idea might appeal to us when we consider the fact that we sometimes just find ourselves valuing or caring for something. Things capture us like another human captures us when we fall in love. While there might be biological or evolutionary reasons for why we react that way, we still cannot give any justificatory reasons for our reaction. We may even find ourselves explicitly denying that the object we love is objectively more valuable than other objects of the same kind.

Despite of this initial plausibility of "bootstrapping", we can find that it is in fact not that easy to respond to a value that no other human being can see or understand as a value. To begin with, it is very unlikely that an individual comes to value something that nobody to whom he is connected values – or at least accepts that the particular individual is right in valuing what he does. The probably most convincing examples supporting this idea can be found among people with marginalized sexual identities and preferences. People with marginalized sexual preferences frequently claim that not only practical possibility and legal rights matter for making different choices possible, but also equal societal respect – being able to see a choice as a choice that is

publicly regarded as legitimate and basically reasonable (see for example Calhoun 2015; Lugones 1987).

We can further say that even if there are occasions in which people are fine with just having quirky tastes or attachments, it seems that these attachments cannot serve all the functions that reason-giving attachments usually serve. While people might sometimes have quirky reasons – effectively and permissibly so – it is usually not the case that these reasons serve very central normative functions – the function to guide life plans, for example, or the function to favour decisions with high stakes and long-term impact. Most people think of reasons serving these central functions as being recognizable by others. Even people who allegedly do not care that much for other people's recognition are usually aware of their reasons as something that can be understood and accepted by a reasonable other. The strong correlation between social isolation and rumination, familiar in psychology, might even support the claim that the need to see oneself as justified gets more pressing in isolation than in times of regular social interaction and communication.

While there is no need to deny the existence of exceptions, it is a pervasive fact about humans that they need the support of others in order to be capable of wholeheartedly endorsing something as a reason – at least if the wholehearted endorsement is supposed to be stable and effective. On the basis of this observation, the idea that anyone can just will or bootstrap reasons into existence by determining herself to do so is not a very plausible picture of human nature. However, there is no need to give up the plausible aspect of a bootstrapping account, the fact that humans actively endow a normatively indifferent environment with value. While individuals cannot will reasons into existence out of their own, they can still do so with the help of others. We could call this a *collectivist version of a bootstrapping account*. Communities with a shared form of consciousness, we could say, are the entities that bootstrap reasons into existence. Communities bootstrap a particular set of reasons into existence, while the individuals growing up and interacting within these communities become responsive to these reasons. Korsgaard's metaphorical formulation of consciousness being "essentially public, like a town square" (1996: 140) indeed fits this picture.[37]

[37] It is still questionable whether we can attribute this view to Korsgaard. If we take her Kantian leanings more seriously than her naturalistic and existentialist leanings, we might tend to attribute a different view to her. We could interpret Korsgaard as saying that the reasons to which a particular individual responds are indeed substantially arbitrary and, in a way, self-invented, while just represented as shareable in the privacy of everyone's mind. In texts in which Korsgaard develops her idea of the publicity of reasons in a more systematic way – especially when she builds on the Kantian resources of the "categorical imperative" and the "kingdom of ends" – this view seems to be what she has in mind (see especially Korsgaard

Putting the question of how to interpret Korsgaard aside, I think that the collectivist bootstrapping account captures the idea that Constitutivism presupposes Reasons Fundamentalism better than a more traditionally Kantian view. In subsection 2.3.1., I have shown by carefully thinking about how we reason that we cannot arrive at any substantial normative conclusions simply by considering principles as abstract as the categorical imperative. We must be responsive to some reasons in order to be able to even start a deliberation. Principles, it seems, can be identified only after the fact. They might sometimes serve an auxiliary function. They are, however, not suited to stand in the beginning of a deliberative process. They might characterize a basic normative response, but usually do neither generate nor justify it.

According to the idea of collectivist bootstrapping, by contrast, reasons are internally represented as shareable because these particular reasons are actually shared. A classically Kantian position, by contrast, would hold that particular reasons can be actually shared because they are constructed according to a universal rational principle. This, however, does not seem to match the nature of reasoning. In chapter 4, I will have a look at the work of cognitive scientists and evolutionary psychologists and confirm both parts of this view – that reasoning is guided by intuitive responses to reasons rather than by logical or structural principles, as well as that reasons are fundamental items that emerged as communicative currency within groups.

This section concludes with a plea for a Constitutivism that embraces Reasons Fundamentalism without restrictions. This amounts to a position saying that there must be reasons, fundamental normative items, underived and inexplicable from the first-personal standpoint of reasoners, and that these reasons must be there independently of a particular agent's decision or will. Their status as "being there", however, does not necessarily require a realist ontology. Reasons Fundamentalism, as it has been defended in this section, is compatible with the idea that there would be no reasons without particular human communities and that, for the individuals responding to reasons, these responses have a self-constitutive function. The last section will summarize

1992). If we read Korsgaard as a Kantian for whom the categorical imperative as a regulative principle, structuring the process of reasoning rather than giving substantial reasons, is the basis of normativity, we must accept this view as hers. We could nevertheless depart from this picture, and still stay in the Kantian terminology, if we argue that the Kantian idea of a "kingdom of ends" as a real-world community in which reasons can actually be shared, is more fundamental than the idea of the categorical imperative, which is only derived from such a community (for interpretations of how to relate the Kantian formulas in a way to capture this idea see for example Pauer-Studer 2016; Calhoun 1994). At this point, there is no space for an in-depth discussion of how Kant or Korsgaard should be interpreted.

these arguments and give the basic outline of an integration of Reasons Fundamentalism and Constitutivism – of a Reasons Fundamentalism combined with a metaphysical view that takes the Constitutivist intuition seriously.

2.4. Uniting the two approaches in a new picture

This chapter tries to bridge the seemingly deep gulf between Reasons Fundamentalism and Constitutivism. So far, I have shown that the differing conceptual frameworks of Reasons Fundamentalism and Constitutivism are compatible. In the very process of reflection, we must take the normative force of reasons as basic, as ultimately inexplicable and underived. We refer to these reasons as facts. However, we can admit that, in order to understand a fact as a reason, we must understand an action rationalized by it.

However, Korsgaard's idea of deriving what reasons we have from principles of practical rationality, and her claim that constitutivist constructivism can answer an agent's first-personal "normative question", is untenable. Still, the questions she and other constitutivists, such as Velleman, ask, are interesting questions. The analysis of reasons offered by constitutivist theories is informative if it is understood in an appropriate way. It represents a significant part of the philosophical interest humans take in reasons and normativity. This is true even if Constitutivism can neither give concrete practical guidance nor convince moral sceptics.[38]

In subsection 2.4.1., I point out why Reasons Fundamentalism, though its basic claim is correct, is limited and does not constitute a comprehensive and fully informative account of what reasons are. Finally, subsection 2.4.2. outlines a version of Reasons Fundamentalism that takes the Constitutivist intuition seriously.

2.4.1. The limitations of Reasons Fundamentalism

The problem with Reasons Fundamentalism is not that it is false, but that it makes a limited claim. I think that Reasons Fundamentalism does not amount

[38]After all, we could find that these are unrealistic expectations towards metanormative theory building – though influential academic debates have been focusing entirely on these questions. In section 1.4., I suggest that the expectation to answer a or the normative question is not only unrealistic, but rests on a conflation of metaethics and normative ethics.

to a theory about the ontological or metaphysical status of reasons. Metaphysicians might disagree. What metaphysicians often mean when they say that an entity is fundamental is that it is not built up or grounded in other entities (see Tahko 2018; also Bennett 2017). If an explanation is fundamental, it refers to a fundamental part of reality – a part that cannot be broken down to factors or constituents making it up.[39]

In this spirit, we can understand Scanlon's step from establishing Reasons Fundamentalism to claiming *realism* about reasons. Scanlon further defends realism by building on the concept of a "domain". Reality allowing for the statement of normative facts is not more problematic, he says, than reality allowing for the statement of physical facts. Each of these facts is stated with respect to its own domain of knowledge and statements from the normative domain simply cannot be evaluated with the criteria of the physical domain.

This idea compels with its obvious simplicity. Rigorous thought about how the world is given to us, how we make claims about the world, makes it inevitable to assume normative facts as equally fundamental parts of reality as physical parts. If we take the term "reality" to comprise the totality of everything that is given to us, that we can refer to, this simple picture is completely sound.

However, I think, that the modern sciences provide us with a possibility to distinguish between "human reality" as what is given to us and "reality" as the world as it is in itself – independently of our perception and cognition. To be sure, each natural or physical fact as soon as we state it gets part of "human reality". Nonetheless, we have the cognitive resources to conceptualize a world that is different from "human reality", even though we have no true representation of this reality. Using these cognitive resources for the analysis of reasons paves the way for an interesting thought: namely, that there is a view in which all there is are creatures whose cognitive systems process certain representations of their environment as reasons – process them in a way that gives rise to the first-personal experience of responding to a reason as a subject. With the methods of the natural sciences, we can investigate what these creatures are doing and what must be the case for such a creature to respond to something as a reason. This promises an account of what that "something" is – an account of what it is to be a reason. In light of such an account, we might

[39]I thank Eric Brandstedt and Johan Brännmark for showing the need to avoid misunderstandings in using the term "fundamental" and Jonathan Schaffer, in the Q&A for my presentation at *Social Ontology 2020*, for pointing out the shares that Reasons Fundamentalists in metaethics take from these general discussions about fundamentality in metaphysics. I hope to show successfully, however, that two perspectives on reasons are possible and equally fundamental in the general metaphysical sense.

want to say that there are not "really" reasons. What it is to be a reason can be analysed further, so, we might conclude, reasons are not fundamental parts of the world. But is the "world" in this sense our human "reality"? After all, we still need reasons to live and to grasp the world we live in, we are not convinced that something is not a reason, just because it is not "really" a reason in the discussed sense.[40]

I would like to preserve both ideas: that there is a "world" in which all there is are creatures responding to facts as reasons; and also that there is a "human reality" of which reasons are fundamental parts. When I distinguished the *material* and the *conceptual* approach to reasons in the first chapter, I have already spoken about a first-personal and a third-personal perspective – the immediate perspective of normatively interested creatures, and the detached perspective of scientists examining what these creatures are doing. This way of speaking, I think, is helpful to reconcile the idea that something can be fundamental from one point of view, but further analysable from another – fundamental from the first-personal view of reasoners, further analysable from the point of view of the scientist who investigates how the first-personal view of reasoners is generated; fundamental as parts of reality as it is given to us, non-fundamental and explainable as parts of the world in and of itself. Finally, we might be able to say that both perspectives are equally fundamental in the sense that both give us an equally valid access to the world.

The most intriguing point about this, I think, is to realize that what is fundamental in our immediate reality and what is analysable when we take another view on the world, is *one and the same thing*. We might be tempted to object that, whatever it is that is described by the scientist, it is *not the same* as the normative reasons we appeal to – since an explanation in terms of reasons cannot be exchanged with a scientific story about our reason-responsiveness. This objection, I think, can be met with Mackie's distinction between *linguistic* or *conceptual* inquiry and *factual* inquiry (1977: 19). How our concepts work, and what there really is, are two different questions. We can *think*. And we can investigate *how thinking works*.[41] We can think. And we can understand that there is a world with thinking creatures whose thinking is exhaustively

[40] See also my critical discussions of error theory in 1.3.1 and 5.2.3.

[41] Of course, we are thinking either way. We must be responsive to reasons in order to engage in scientific investigation. As already mentioned, it is impossible to represent the world as it is in itself. However, we are cognitively capable of bracketing every reason and investigating why we respond to it as a reason. Even if each scientific practice is reason-guided, we can bracket the question whether there is anything we ought to do or believe no-matter-what.

explainable in different scientific terms. Our thinking is our human reality. But the world as it is is different from human reality.

I would expect a metaphysical account of reasons to capture this insight. Accordingly, I agree with Scanlon's view in so far as his realism is just the claim that reasons are fundamental parts of our human reality and that this reality cannot be maintained without conceptualizing reasons as fundamental. However, we might want the term "realism" to imply something stronger – that something is part of the world as it is in itself, rather than just being part of "human" reality. We might want to distinguish between Reasons Fundamentalism, a conceptual claim, and realism, an ontological or, in Mackie's words, "factual" claim. If we decide for this distinction, Scanlon's argument establishes Reasons Fundamentalism, but not automatically realism.

What makes this distinction attractive? With respect to normative questions to which we seek definite answers, there is, of course, not much insight to be gained. Nevertheless, human thought gains significant faculties by having the capacity to see that nothing is "really" a reason in this sense. We might, for example, realize that thought has developed over time. We might lose the certainty that our way of thinking is the right one or the only one that is possible in nature. We might, furthermore, change our relation to other species, other cultures and other historical epochs, because we become aware of the contingency and particularity of our cognitive foundations.

Finally, we might start to consider the possibility that our own cognitive foundations are still changing – that our ways of responding to reasons might evolve further and that we might embark for new horizons. A position that confines its interest to how we reflect on reasons and how we use the concept of a reason misses all these aspects that come with an understanding of what reasons are that transcends the conceptual claim of Reasons Fundamentalism. If the conceptual fundamentalist claim is defended together with a realist ontology, we must make explicit that the realist claim is more than the fundamentalist claim and that it is in need of further arguments – arguments that most metaethicists fail to give. A metaethicist, to be sure, does not commit an error by confining herself in this way. The resulting theory, however, misses out on interesting dimensions.

2.4.2. A version of Reasons Fundamentalism taking the Constitutivist intuition seriously

If Constitutivism is defended as an alternative to Reasons Fundamentalism, claiming that reasons can be derived from the nature of agency, Constitutivism is false. Reasons Fundamentalism, I have suggested, is true. However, it is only

a limited conceptual claim and does not settle the ontological question of what reasons are. At this point, what I called the Constitutivist intuition – the intuition that there are no reasons in the world independently of reflective creatures who need them – can be a fruitful starting point for complementing the conceptual thesis of Reasons Fundamentalism.

The task to be completed in the rest of this book is to develop a version of Reasons Fundamentalism that takes the Constitutivist intuition seriously. In order to pave the way for this endeavour, I would like to introduce some metaphysical preliminaries developed by John Skorupski, who defends a fundamentalist proposal like Scanlon, but puts much greater emphasis on „the relation between self, thought, and world" (Skorupski 2010: 1; see also Skorupski 2012). Skorupski presents the inquiry into reasons and normativity as, primarily, a „road to understanding that relation" (ibid.).[42] He introduces his position as a version of what he calls the „Critical Stance", distinguished from what he labels „Global Realism" (2010: 7). The crucial distinction between the Critical Stance and Global Realism is the way it conceives of the propositions that serve the function of reasons in our reasoning.

I will now outline the main achievements of Skorupski's metaethical theory, which I will inspect more critically in the next chapter and remodel thereafter. To begin with, Skorupski emphasizes the distinction between „factual" and „normative" propositions (2010: 4). Both are statements that can be true or false, but only the former ones, the *factual propositions*, are true or false in virtue of a reality that corresponds to the depiction, that is depicted by help of a receptive apparatus that can be deceived or get things wrong (2010: 8). Factual propositions state facts in what Skorupski calls the „ontologically substantial notion" of facts (2010: 7).

As opposed to that, the contents of *normative propositions*, Skorupski argues, are known in a different way. This knowledge formation, likewise, can go wrong, but it does not go wrong because of a mistake in receptivity. Normative propositions are not depictions of an ontologically substantial reality by help of our receptive apparatus. They state facts in what Skorupski calls the „purely nominal notion of factuality" (2010: 7). This conception of factuality is very difficult to understand and must be picked up again in the next chapter. Our way of knowing nominal facts is, as Skorupski claims, not by receptivity, but by way of „spontaneity" (2010: 12) – by responses that

[42] Johan Brännmark objected that this is true for Scanlon as well. I agree that Scanlon *accepts* most of the things that Skorupski says about that relation. However, it is very obvious that he spends much less diligence *elaborating* on what it is like to respond to a reason, what it means to have "free thought" and how to exactly understand the self in the world. Skorupski's work is therefore more fruitful for my purpose.

genuinely issue from our very nature as subjects (2010: 406). Spontaneity is educated and cultivated by frequently communicating and checking our own spontaneous responses with these of other people with whom we live in a community (2010: 412f). The process of knowing reasons takes place in the dialectic between „individual spontaneity“ and „free dialogue with other inquirers“ (2010: 30). Skorupski's denial of a specific kind of receptivity required for knowing reasons is very similar in spirit to the defence of the „domain of reasons” as a genuine domain of knowledge, also advocated by Scanlon.

Scanlon and Skorupski cannot be read as opposing each other right away (see also Scanlon 2014: 24 FN 9). However, Skorupski's sophisticated picture leads him to defending an irrealist ontology of reason relations, as opposed to Scanlon’s realist proposal. With Irrealist Cognitivism, Skorupski presents a position according to which normative statements can be true or false, but not in virtue of reality to which it corresponds, but in virtue of an irreal realm, consisting of irreal items to which our mind can respond spontaneously. Responses to these items, Skorupski spells out, are „constitutive of the idea of thought“ (2010: 1). Responses to reasons constitute the *process* of “free thought” – an idea that will later serve as a model for my conception of the *activity* of reasoning as *moving* in a collective archive of fundamental items, a collective archive of mind.

Fundamental items enable thinking and acting for reasons conceived as „activities of the self as it is in itself“ (2010: 416). By responding to reasons, we realize our nature as subjects in the most genuine sense. The emphasis of the relation between *responding to reasons* and *being active as a subject* is one reason for studying Skorupski's account in more detail. Another reason is Skorupski's emphasis of the essential role of community in getting to know reasons. The next chapter will ask in an especially critical way for what makes this role so essential and whether the essentiality of community can also have implications for the ontology of the irreal realm in which our thinking moves. The irreal realm shared within a community can be understood as the context in virtue of which we can be subjects at all. Responsiveness to reasons is inextricably linked to our existence as subjects. However, to us as subjects, truths about reasons are fundamental. Skorupski says about the concept of a reason, thereby being more explicit than Scanlon (2014: 2), that

> it is a concept fundamental to all thought. It is pervasive – actions, beliefs, and sentiments all fall within its range; primitive – all other normative concepts are reducible to it; and constitutive of the idea of thought itself (Skorupski 2010: 1).

Skorupski's account is a promising candidate for a version of Reasons Fundamentalism that captures the intuition that responding to reasons must be understood as presupposing a context – a specific type of context, which subjects, we could say, inhabit as subjects and existentially share with other subjects. This is more than what a specification of circumstances within the reason relation gives us. It is an ontological reframing of the reason relation as such. It turns reasons into irreal items of thought. Skorupski says that reason relations are mind-independent, but that the concept of „ontic cognition-independence" is inapplicable to them (2010: 405). Fictions are also irreals. However, fictions, as opposed to reason relations, are subjective irreals – dependent on a mind (ibid: 426). Skorupski describes reason relations, as distinct from fictions, as „irreal and objective – not mind-dependent" (ibid: 420). Some authors interpret Skorupski's irrealism as similar in spirit to a domain-specific realism or a realism in a „non-ontological" sense specified by Parfit (e.g. Olson 2018). Still, I think that the distinction between the real and the irreal makes the sophisticated relation between mind and world much more explicit than the simple solution of demarcating different domains of reality.

The next chapter investigates the theory of mind-independent objective irreals in more detail and suggests that we should understand these entities as, though not depending on the mind of someone in particular, depending on mind as such. A distinction between „mind as such" as distinct from mind in the sense of „my mind" or „your mind" can also be found in Frege[43], to whom Skorupski refers for his conception of propositions stating nominal facts (Skorupski 2010: 7). In the next chapter, I will suggest that the best way to understand „mind as such" within the frame of a scientific world view is to

[43] Geach translates Frege as saying: "Neither logic nor mathematics has the task of investigating minds and the contents of consciousness whose bearer is a single person. Perhaps their task could be represented rather as the investigation of the mind, of *the* mind not of mind." (Frege 1956: 308). In German (Frege 1918), the last bit is "*des* Geistes, nicht eines Geistes". It is very clear in the German formulation that what Frege has in mind is a distinction between "a mind" (the singular and particular mind of a person) and "the mind" (mind as such). In contrast to Geach's translation, my experience is that English speakers often refer to "the mind" in this sense without an article ("mind") and mean a particular mind when they use an article. This distinction between mind as a count noun and mind as a non-countable entity is important. As the next chapter will show, we can understand mind as a non-countable entity in different ways: As the name for a cognitive faculty that varies between species and as one particular entity in the universe. The latter interpretation is the orthodox Fregean one, which I will reject on grounds of scientific implausibility. While this orthodox interpretation is typically ascribed to Frege, he offers, as far as I can see, strikingly low support for the thesis that "mind" is more than "human mind", though his approach allows for the claim that other forms of mind are irrelevant as all scientific knowledge about them is gained by "mind". See also Brandom (1986).

understand it as a proto-typical form of cognition – a form of cognition involved in explicit, conscious self-reflection, a form of cognition that is typical for creatures having an understanding of the world and themselves in it. Skorupski's intuition is that „the world is ´our´ world rather than ´my´ world“ (2010: 31), normative truth is „grounded in cognitive communities of beings with the same natural dispositions“ (2010: 417).

This suggests, in line with plenty of empirical evidence, that cognitive sharing is essential for this type of cognition. But does that mean that "mind as such" is relative to a community? Does it mean that reasons depend on communities in an ontologically strong sense? The next chapter is dedicated to the idea of mind-independence and the Fregean ideas, on which Skorupski's irrealism builds.

2.5. Conclusion

In this chapter, I reconciled two positions that apparently oppose each other: Reasons Fundamentalism and Constitutivism. I offered an integration of the two approaches as painting two different, but compatible pictures on what it is to be a reason. On the one hand, there is the idea that acting for a reason is to constitute oneself as an agent, to understand oneself as relating meaningfully to an intelligible world. Normative reasons, according to Constitutivism, are determined by what enables us to understand ourselves in that way. On the other hand, there is the idea that, as agents who reflect about what reasons they have, we get no help, at least not in general, from considering what would constitute us as agents. According to Reasons Fundamentalism, we are simply responsive to reasons. We understand them if we reflect in the right way and we cannot explain them in any general way.

This chapter proposed that Constitutivism, when taking the first-personal stance of reflective agents, must embrace Reasons Fundamentalism. Reasons Fundamentalism, accordingly, is the view that is correct when it comes to the question of how we think about reasons as agents taking reasons "at face value", asking what we really have reason to do. There is nothing entailed in the very concept of a reason that enables us to derive what reasons we have, if we are not already responsive to some reasons, which in this very process are conceptualized as fundamental.[44] With this rationale, the chapter confirms the

[44]To be precise, not strictly all reasons must be fundamental, as some reasons can obviously be derived from other reasons. However, all this does not work without presupposing some

understanding developed in chapter 1, according to which we must separate a *conceptual* question or a *metaethical question narrowly conceived* from a *material* question of what reasons are. With this chapter, I illustrated how such a distinction can be used to illuminate a complex landscape of metaethical positions, which is hard to disentangle and contains a lot of hard to pinpoint disagreements. Some of those disagreements, I hope to have shown, are only apparent and dissolve if we understand the metaethical question narrowly conceived as distinct from the material question.

When we look at how we use the concept of a reason, how we talk in terms of reasons and how we reflect about what reasons we have, we must adopt Reasons Fundamentalism. While arguments for Reasons Fundamentalism are sometimes claimed to establish "non-reductive realism" or "non-naturalism", these arguments fail to establish anything more than a narrow conceptual claim: The claim that we cannot derive reasons from any non-normative facts, if we are not already responsive to some facts as reasons.

This view, however, is compatible with a bootstrapping account in the style of Frankfurt. As suggested in subsection 2.3.3., Korsgaard's theory of self-constitution can be interpreted as a plausible modification of Frankfurt's radically subjectivist bootstrapping account. Still, the most plausible account remains a "bootstrapping account" – a collectivist, rather than a radically subjectivist one. In any case, this kind of account must count as a version of Reasons Fundamentalism, when it comes to the metaethical question narrowly conceived. When we, by contrast, look at what a reason is from a psychological perspective, asking for the material reality of reasons, the one that is empirically observable in space and time, we must deal with the ideas that accompany the Constitutivist intuition – the idea that acting for a reason is a process of self-constitution and that self-constitution presupposes that we can share the reasons we act for with others.

I finally suggested leaving Scanlon's realist Reasons Fundamentalism behind for a Reasons Fundamentalism that takes the Constitutivist intuition seriously. As a promising theory in that respect, I introduced Skorupski's Irrealist Cognitivism, which understands reasons as *objective mind-independent irreals* – neither existing in the real world, nor being up to an individual mind.

Skorupski does not define his theory as a bootstrapping account. It is clear that he rejects the radically subjectivist version. It is less clear whether he also succeeds in rejecting collectivist bootstrapping. Skorupski's view is worth

---

reasons as fundamental and understanding their normativity as fundamental in the sense of irreducible and ultimately inexplicable.

examining in more detail, because Skorupski wants to establish something that comes very close to what Scanlon claims to establish, without neglecting the Constitutivist questions of how reasoning relates to the self, and how the self relates to the world.

Chapter 3
Are reasons mind-independent?
The relation between grasping and what we grasp

> *Neither logic nor mathematics has the task of investigating minds and the contents of consciousness whose bearer is a single person. Perhaps their task could be represented rather as the investigation of the mind, of the mind not of mind.*
>
> (Frege 1956: 308)

3.1. Introduction

The previous chapter argued in favour of Reasons Fundamentalism – a view according to which we must assume entities that are irreducibly normative in order to be capable of reasoning. Reasons Fundamentalists, such as Scanlon and Skorupski, describe these entities, the reasons we respond to, as *mind-independent*.

I have argued that the claim of Reasons Fundamentalism is limited in scope. In distinction from the common defences of Reasons Fundamentalism in the literature, I argue that we can only understand the thesis as saying that, *from the perspective of reasoners*, we must find ourselves able to *take* certain reason relations as fundamental and holding independently of us. Any stronger claim about the *reality* of reason relations cannot be defended with the means of conceptual analysis only. As shown in the previous chapter, Reasons Fundamentalism is compatible with several ontological positions, even with a so-called bootstrapping account of reasons. According to a bootstrapping account, mind just invents something that functions as a reason out of nothing. Still, the normativity of what functions as a reason is experienced as

inexplicable. In this sense, the force of reasons is perceived as irreducible by the reasoner. But what to do with the mind-independence claim?

The observation that we as reasoners perceive reasons as fundamental parts of our human reality allows us to raise a mind-independence claim in what we can call a *deflationist sense* (see subsection 3.5.1.). But beyond this weak claim that is compatible with a variety of ontological positions, we can ask further, ask in a more comprehensive sense, what it is that the reasoner perceives as irreducibly normative. This allows us to discuss the mind-independence claim raised by some Reasons Fundamentalists in a *metaphysically stronger sense*.

Skorupski's formulation of Reasons Fundamentalism, as introduced in the last section of the previous chapter, is especially promising in this context, since it combines an interest in the ontology or metaphysics of reasons with an equal interest in the essential psychological function of responding to reasons. Skorupski, as distinct from Scanlon, defends irrealism about reason relations. Nevertheless, he insists on a mind-independence claim. With his conception of what he calls *objective irreals*, Skorupski makes an interesting proposal that is rooted in Frege's conception of *thoughts*. According to Frege, a thought is a mind-independent abstract entity that is not generated or produced but, instead, discovered and grasped by mind. Thoughts are neither subjective imaginations nor objects in the physical world. The realm of thoughts, according to Frege, must be recognized as a "third realm".

Skorupski adopts an idea in that spirit when he speaks about the "domain of reasons" as an ontological domain that is neither defined in terms of the real (causally active) world, nor in terms of mind. Skorupski classifies reason relations as not existing in the real world, but fundamental to the process of thought and unchangeable for the mind. Mind responds to reasons, or even better, *grasps* reasons.[45] By grasping reasons, we become capable of what Skorupski calls *self-determination*, a subjective activity that represents ourselves as free subjects in a spatiotemporally and causally unified world. This activity is a genuinely mental one. However, this does not mean that the world is in our mind. Nor does it mean that our mind grasps objects in the world directly. Mind grasps reason relations – facts that stand in relations to us that give them the property of being reasons.

[45] Skorupski uses the term "responding to reasons", which is the most common one in the metaethical literature on reasons. This term is – at least here – equivalent to the term "grasping", as it is for example used by Frege in his idea of grasping thoughts (1918). I prefer the term "grasping" mostly for its expressive strength in describing what responding to a reason is like for us. The suitability of the term will reveal in this and especially the next chapter, in which I explore the analogies between grasping thoughts/reasons and grasping as a sensorimotor process oriented in space.

The aim of this chapter is to discuss the relation between *grasping* (a mental process) and *what we grasp* (abstract items that are independent of our mind). At the end, I suggest modelling reasons as potential mental states. Reason judgments need not be instantiated for there to be a reason, according to this suggestion. However, each reason that there is can be modelled as a potential reason judgement.

The next two sections discuss Skorupski's Fregean theory of mind-independent abstract entities along the lines of its two main constituents, their distinctness from *subjective entities* and *physical entities* on the one hand (3.2.) and their *intersubjective shareability* on the other (3.3.). I distinguish two controversies about how to understand the mind-independence claim in a metaphysically strong sense. I then point out why the most radical mind-independence claim would run into a Darwinian Dilemma (3.4.). Finally, I offer a neo-fregean inspired ontological modification of the idea of a "third realm" – or a domain of "objective irreals" according to Skorupski (3.5.).[46]

3.2. Skorupski's Fregean distinction between mind and thoughts: a first controversy about the nature of thinking

By thinking reflectively about the world, we make the world cognitively accessible to us. To use a metaphor, we could say we intellectually permeate or, in other words, we inhabit the world intellectually. In the process of thinking, we actively make sense of our actions and of the events affecting us. The propositional contents of thinking – the parts of our conscious activity that are *about* something, that have substantial contents, which can be made explicit and spelled out verbally – can be called *thoughts*. What a thought as such is, is a question that is difficult to answer. According to a simple psychologist or neuroscientist view, it might seem obvious that thoughts are produced by minds. Philosophical scrutiny, however, renders it doubtful whether the concept of a thought, being a genuinely semantic concept, applies to something that can be produced or detected in an empirically observable process.

[46] I thank the participants of the LOGOS seminar in Barcelona for comments on the contents of this chapter during my stay there in spring 2019. For comments on the interpretation of Frege, I am especially indebted to Genoveva Marti and Josefa Toribio. Furthermore, I thank Jakob Werkmäster for discussions about mind-independence and Skorupski's book.

A philosophical tradition, famously established by Frege, distinguishes sharply between *processes of mind* and *thoughts as unproduced abstract entities* that are grasped in these processes. Reasons, in Skorupski's work, are described as thoughts in that Fregean sense. Reasons are described as being the abstract entities that govern self-determined thinking. Skorupski thereby presents an idea of a *mind-independent realm of abstract entities* to which mind can respond. Ontologically, these entities are described as objective irreals, to be clarified in more detail soon.

This idea is open to two different interpretations: According to one interpretation, the realm of objective irreals is independent of any particular form of human mind, eternal and uncreated in the strict sense. According to an alternative interpretation, there is an objectivist phenomenology in the psychological response to a reason, but the realm of reasons is a virtual creation of mind, a kind of effective mental objectivation, following the structural features of a particular form of mind. It seems clear that Skorupski favours the first interpretation in this controversy, but many aspects of his thinking point towards a qualified version of the second one. The following three subsections shall have a look at how Skorupski conceptualizes the ontological domain of the *irreal* in relation to the *real* world (3.2.1.), how he distinguishes *objective irreals* from *subjective irreals* (3.2.2.) and how he conceives of our knowledge and *response to objective irreals* (3.2.3.). Finally, subsection 3.2.4. points out that a Fregean distinction between *mental processes* and *thoughts* still leaves room for a crucial controversy about the mind-independence of abstract entities – a controversy that neither logics nor metaethics narrowly conceived, as pure conceptual analysis, can solve with its limited means.

3.2.1 Distinguishing the irreal realm of thoughts from the real world

The main difference between the real and the irreal, according to Skorupski, is that the real is *causally effective*, while the irreal *has no causal stance in the world*. The irreal is not part of the world at all. Skorupski rejects what he calls the "semantic condition" for existence (2010: 421), or elsewhere the "Carnap/ Quine thesis", that "reference to an entity entails its existence" (2017: 594). The totality of everything we can refer to, is called the *actual* in Skorupski's work, while not everything that is actual also exists.

Skorupski holds both that normative judgments can be true or false and that there is nothing in the world that makes them true or false. He thereby occupies an interesting middle ground between realism and error theory. Mackie bases his error theory on the worry that philosophers are biased "towards various

kinds of linguistic inquiry" (Mackie 1977: 19), while, on the "factual" level, there is nothing to which normative statements correspond. Skorupski takes this worry seriously, but nonetheless does not follow Mackie in claiming that people commit an error in making a normative judgment. Instead, he holds that claims about reasons are not "factual" at all. There are no facts to which normative propositions correspond or are even supposed to correspond (2010: 431). Normative propositions are not true or false in virtue of correspondence to facts that are independent of the propositions stating them. Their truth conditions depend on an irreal domain and are met in a different way, as we will soon see.

The world, according to Skorupski, is composed of facts that exist independently of propositions stating them – facts that exist independently of us in the natural scientific sense and that can affect us causally. Skorupski refers to this as the *substantial notion of fact*. According to the "ontologically substantial notion" of fact (2010: 403f), facts are "perfectly distinct from propositions" (ibid.: 404). For substantial facts, the fact that p and the belief that p are distinct (ibid.: 405). Substantial facts are "ontically cognition-independent" (ibid.). When we come to know about a substantial fact, we do so by receptive faculties (ibid.: 404). As examples for substantial facts, Skorupski mentions a flash of light or a plane passing by. Substantial facts have a causal productivity of their own and can affect our sensory apparatus. Skorupski defines the world as "the totality of substantial facts" (ibid.: 405). To be real, in Skorupski's terminology, is to have causal standing (ibid.: 427).

This does not mean, however, that we cannot talk about anything else without committing an error. Skorupski's irrealism emphasizes that when we think and talk, we do not just represent what is the case in the world. We communicate meaningfully by reference to what Skorupski calls facts in the "purely nominal notion of factuality" (ibid.: 403):

> The nominal notion identifies facts with true propositions, or treats them as abstract entities – truths – that stand in one-one relation to true propositions. We understand ´proposition´ as the sense or meaning of what is asserted on an occasion. So the nominal notion of fact is pretty much the Fregean notion: ´a fact is a thought that is true´. (ibid.: 403)

In this sense, the sentences, ´Hesperus is a planet´ and ´Phosphorus is a planet´ express different propositions and state two different nominal facts. A *nominal* fact obtains when a stated proposition is true. The number of nominal facts that we can state need not correspond to the number of substantially real entities that there are. Skorupski holds that "the notion of *ontic* cognition-independence is inapplicable to nominal facts" (ibid.: 405).

Here is Frege's example to make the ontological distinction between the world and our thoughts about the world intelligible: "That the sun has risen is not an object which emits rays that reach my eyes, it is not a visible thing like the sun itself." (1956: 292) To use Skorupski's terminology, the sun, or rather its physical interaction with us, is a substantial fact. A substantial fact, however, is nothing we can think or understand directly. What we think is the thought or the fact *that the sun has risen*, which is distinct from the world. It would be misleading, however, to imagine the world as being populated by two facts, the sun and our thought about it. Rather, the distinction must be emphasized in order to raise awareness of the fact that our cognitive access to the world is not identical to the world. We refer to the world via senses. A *sense* in Frege's terminology determines the *reference*, the object in the world we are talking about, while, however, the reference does not determine the sense, the way we are thinking about it, or the way we make it intellectually accessible to us (Frege 1892; see also Marti 1998; 2007).

Skorupski's refusal to count nominal facts as existing, a terminological nuance that Frege himself does not make, is an advantage in capturing a crucial idea. This idea says that for intellectually permeating or inhabiting a world, as opposed to merely reacting to the environment as a stimulus-response apparatus,[47] we must assume more than just the physical world and the sensual apparatus of an individual organism. We must assume a third category – *senses* or *thoughts* graspable by mind. In Skorupki's terminology, the graspable abstract entities constituting the subjective process of thinking, are called *reasons*.

3.2.2 Distinguishing the irreal realm of thoughts from subjective imaginations

The idea that we can refer to things that do not exist is most unproblematic when we think of what Skorupski calls subjective irreals. *Subjective irreals*, as opposed to objective irreals, are, for example, fictional characters or "putative reals", objects we just assume for the sake of planning, such as a door in a house that is not actually built (2010: 424). It is important to note, in contradistinction to for example Parfit's idea of existence in a non-ontological sense (2011b: 481), that the abstract objects of such talk are not said to exist in some sense, but that they do not exist at all. Still, communication about them

[47] In chapter 4, the simple idea of a stimulus-response model will be taken up again and dismissed as inappropriate for describing any creature that actively survives in an environment.

is unproblematic. As with fictional characters and putative reals, this is unproblematic because the abstract objects are, as Skorupski calls it, "anchored" in either works of fiction or plans that depend on the minds of people who created them. They are "subjective irreals", having a "mind-dependent anchor" (2010: 425). They are anchored in something real, but they are not real themselves.

In addition to that, however, Skorupski defends the view that there are irreal entities that can be called mind- or cognition-independent, that are not anchored in any creation of mind. Skorupski calls them *objective irreals* (ibid.: 420). The domain of objective irreals embraces, among other abstract objects, such as sets or numbers, *reason relations*.

Our cognition can refer to these objects but must take them as independently given facts. "Their irreality", Skoruspki says, "is fundamental and unique to the normative domain" (ibid.: 420). The normative domain, in turn, is fundamental to cognition as such and therefore cannot be its product or construction. There is a crucial difference between being "the product or construct of an act of imagination" and being "the object of cognition" (ibid.: 429). While "fictional objects are irreal because they are intentional objects within imaginative constructs", reason relations "are not intentional objects within an imaginative construct" (ibid.). Skorupski says the following about reason relations: "They are irreal because they are objects not of imagination but of *pure cognition*. A priori truths about them are cognition's norms." (ibid.) It is not just that our cognition *can* refer to objective irreals. According to that picture, it is rather that there needs to be reference to objective irreals for there being cognition as we know it.

An essential characteristic of cognition, which, according to Frege, shows that we must assume the realm of senses or thoughts to be independent of subjective imagination, is the possibility of communication – most interestingly scientific communication. Frege, like Skorupski in his comparison to fictions, takes subjective imagination to be a merely private entity in an individual's head, obeying the individual's fantasy, rather than having its own laws. When we understand the world, by contrast, we grasp something that does not depend on our mind. Thoughts are discovered, rather than created. Our mind must take them as they are. Frege compares the grasping of a thought directly to the grasping of a physical object, such as a pencil (1983: 149).[48] Like reaching out for a pencil, trying to grasp a thought

[48] Chapter 4 will expand on this idea and argue that there is a fruitful way of understanding human interaction with the environment, according to which we can take this analogy even more literally.

is an attempt to discover the shape and reality of something that is independent of whether we reach out for it or not.

3.2.3 Grasping reasons: Spontaneity, apperception and self-determination

Grasping reasons – Fregean thoughts or senses – is discovering something that does not depend on our mind. In a way, it is comparable to grasping substantial objects, although thoughts are not substantial objects, but objective irreals. Objective irreals, as has been established so far, are described as "objects of pure cognition" or as "cognition's norms" (2010: 429). Since objective irreals are not in the world and do not have an influence on our senses, epistemic access to those objects must be more immediate than all kinds of receptive knowledge about the external world. Knowledge of substantial facts, as already mentioned, involves a causal interaction between "productivity" in an object and "receptivity" in a subject (ibid.: 405). Knowledge of a nominal fact, as opposed to that, can only involve two elements according to Skorupski: "spontaneity" and "convergence". The latter is a reflective equilibrium established in communication with others. The role of communication will be a separate topic in the next section. The former, spontaneity, is a specific property of a response – a response that comes from the very nature of the subject.[49] It is crucial for understanding the subjective act of recognizing objective irreals. Spontaneity is necessary for making judgments, while receptivity is not always necessary. There are, according to Kant, some judgments that are purely spontaneous. Skorupski picks up a basic idea from Kant and distinguishes himself in the following way:

> Spontaneity is a property of responses and dispositions to respond. The basic idea is that a spontaneous response or disposition is one that comes in the right way from, is genuinely that of, the actor. Kant gives this basic idea a causal interpretation: spontaneity, self-origination, is ´original´ - uncaused – causation of its activity by the active self. He then takes it that such causation cannot be understood at the empirical level. So spontaneity becomes something whose possibility can be vindicated only by transcendental idealism. We should not follow him in this: spontaneity, as I interpret it, is not as such a causal notion at

[49]The term "spontaneity" is taken from Kant and contrasts with "receptivity" (see Kant 1999/1787, sec. B I).

> all, though there can be subversive causal explanations of a person's response which preclude its spontaneity. (ibid.: 406)

Skorupski works with a conception of a response that cannot be understood causally at all. He does not want us to understand it as what some authors call "agent-causation" (ibid.). The actor does not cause or bring about a response. Instead, a response that is truly spontaneous issues directly "from the actor's nature" – it is expressive of the subject in virtue of its very nature as a subject, we could say.

The difference between *causing something as a subject* and *responding as a subject* might not be easy to draw. Responding as a subject, according to that distinction, is a phenomenon not questionable or refutable in light of empirical evidence. We can draw the distinction by reference to an idea that Korsgaard presents as a response to what she calls the paradox of self-constitution (2009: 20). There seems to be a similar line of argument in Korsgaard, though Skorupski does not explicitly align himself with her in that respect. Korsgaard rejects the idea that we have or are a self that might eventually decide to cause or bring about an action. Rather, she claims, we constitute ourselves in the very process of acting, we are selves only in virtue of acting. Consequently, a paradox seems to arise with the question of how we can constitute ourselves in our actions if we are not already there prior to our actions.

Korsgaard argues that this paradox is only apparent, because "self", rather than being a particular entity, is just the name of an emergent phenomenon – the phenomenon of unified consciousness that dynamically appears in the process of a certain reflective activity. A self, accordingly, is just a form our first-personal reflective consciousness typically takes.[50] It is only available in that specific first-personal way. Claims about first-personal activity, by their very nature, cannot be evaluated directly in light of scientific or other empirical investigations. However, first-personal activity does not establish another reality in the sense of a parallel world. First-personal activity as a psychological phenomenon is not something that refuses biological or neurological observation in general. It is rather that conclusions about first-personally given phenomena, about the "self" in Korsgaard's or the "subject" in Skorupski's terminology, cannot be drawn directly from empirical inquiry. Most likely, science can in principle explain what needs to be the case for an organism to have first-personal experience. However, there is no such thing as

[50] I thank Björn Petersson for an exchange about this view. Helpful resources were also Brinck (1997) and Shoemaker (1996).

a self or a subject involved in the causal chain. The particular reality of being a self is only what is immediately given to a first-personally reflective subject.

In Skorupski's picture, being a subject is just the first-personal phenomenon of responding spontaneously to reasons – a process that comes with a distinctive "apperceptual" quality. Apperception is a mental phenomenon that makes us self-determining agents. Self-determination in this sense requires a "unity of apperception" (2010: 462). Apperception, in distinction from perception, does not represent anything external to consciousness. It is a distinctive state of consciousness. Apperceiving a state, according to this idea, is just being in that state (2010: 459). Apperception is a state of consciousness that comes with consciousness of a "synchronically and diachronically unified field" (ibid.: 464). A self-determiner conceives of itself and others as located in a perceptual field as well as a memory field and thus as located in a "spatiotemporally and causally unified framework" (ibid.: 465).

This way of conceiving of oneself requires the capacity to represent a world consisting of objects, other subjects and oneself as one among them, but it does not consist in these representations. Any particular representation can in principle occur in the quality of *apperception*. The apperceptual features characterize the shape which consciousness takes whenever it grasps something as a reason. The point is not that a subject grasping a reason grasps anything in addition to the reason-giving fact. It just grasps a fact as a subject, as being oriented in a certain unified field.

Apperceptual features being in place do not necessarily indicate that someone grasps a reason correctly. However, we need the general capacity to respond spontaneously to objective irreals, in order to be capable of apperceptual states and thereby become what Skorupski *calls self-determining agents* – agents seeing themselves as causally effective in space, time and, as will be extended in further detail in the section thereafter, intersubjective reality. The domain of objective irreals and the psychological capacity for self-determination are inextricably linked. The realm of objects we can apperceive, instead of perceive, is the third realm, the realm we need as a link between the physical world and the sensual apparatus of an individual organism, if we want to account for minds subjectively inhabiting a world, rather than merely being connected to external input as a stimulus-response apparatus.

3.2.4 A controversy about the relation between mind and what it grasps

In his Irrealist Cognitivism, Skorupski links the ontologically irreal domain of reasons directly to our capacity for self-determination. Self-determination,

according to that picture, is a psychological phenomenon. It is a phenomenal feature, which the self displays in the process of a particular kind of response. This particular kind of response enables its very first-personal activity as a self.

There are two possible ways of interpreting the interrelation between the self and the domain of reasons: An interpretation Skorupski attributes to Kant and rejects is the idea that reason relations are cognition-dependent (2017: 597f). According to that view, reasons are internal constituents of our subjective thinking. In this sense, they are parts of our subjectivity and, according to this view, this is also why we have immediate access to them via spontaneity.

With a conception of objective irreals as cognition-dependent, we could conceive of subjective reasoning as the mental activity of producing ordered ideas. This subjective productivity enables the activity of the self to continue in a functional way. Mind does not directly depict an external reality. Instead, it actively develops a way to make sense of external input that could also be made sense of in a different way. The world, in other words, does not determine exactly how the mind must make sense of it. Mind develops its own repertoire of objective irreals – a process we could call objectification (see also Mackie 1977). It might be fitting to describe this repertoire as virtual.[51] The notion of *virtuality*, as it is commonly used, emphasizes that the structure of the normative domain is a contingent construction that could always be otherwise and that is in no direct way necessitated or determined by how the physical world is. However, as a widespread definition of virtuality entails, mind effectively complies to it as to an objective mind-independent reality.[52]

The objectification thesis about mental activity conceives objective irreals as products of mind. This way of speaking might still be criticized for using the concept of a thought or proposition in a misleading way. It can be

[51] I thank the audience of my presentation at *The Future of Normativity* (Kent 2018) for help in specifying the idea of the realm of reasons as a "virtual space".

[52] It might be that the complex mental activity of producing thoughts, in this view virtual items, which is essentially connected to self-reflexivity, had an evolutionary advantage for human animals. It might, however, as well be that this kind of mental productivity is only a by-product of evolution, not necessary or, even if that seems unlikely, potentially hindering for human survival. This question cannot be answered by philosophy. The point that is relevant for philosophical understanding nevertheless is that, if the domain of objective irreals is a virtual space created by objectivation, reasons can, in a sensible way, be called products of mind. As fundamental norms of cognition, reasons, of course, cannot be intentional products of imagination, such as works of fiction developed by a creative individual, as in Skorupski's notion of subjective irreals. However, as the idea of objectification suggests, mental productivity can take place on other levels of mind as well. Later on, another sense of fiction will be considered – a type of fiction that is non-intentional and more fundamental to thought than an individual's explicit and purposeful imagination.

misleading to use that semantic concept as if it stood for an object that comes into existence in space and time and can be detected in the physical world. We can grant that this picture is inadequate. It seems appropriate to call semantic entities such as thoughts or the propositional contents of thoughts uncreated in a very trivial sense. Being uncreated in that trivial sense means being "outside" of space and time in the sense that these categories are inapplicable. It does not make sense to ask when or where a semantic concept was produced. We can only ask when or where the capacity to refer to it or use it was developed.

This rationale makes the distinction between a mind as a psychological system and a thought as a semantic entity intelligible. It supports the philosophical tradition established by Frege, to which also Skorupski adheres, which draws a clear dividing line between *psychology*, describing the empirical process of thinking, and *logics*, describing the "laws of thinking", or the concepts fundamental to thinking, as they are. These laws cannot be discovered in the workings of a psychological apparatus. They are not made or produced by a psychological apparatus. These laws, according to Frege, are a matter of fact that can be discovered by pure thinking, not a process that can be observed empirically. In line with Frege, we can best describe discovery by pure thinking, or spontaneity according to Skorupski, as grasping something that is beyond and independent of one's mind and to which the structure of one's mind complies or is necessitated to comply. In Frege's seminal text *Der Gedanke* (1918; English 1956), it seems to be mainly this idea of grasping, as opposed to generating in imagination, which gives rise to the famous theory of a "third realm" – the realm of thought which is neither part of the external world nor a subjective imagination. In the trivial sense mentioned above, we can plausibly describe the realm of thought as uncreated.

The controversy about how to understand the realm of objective irreals in relation to mind is, however, untouched by this Fregean rationale. The objectification thesis, viewing reasons as mental products, is compatible with concepts being uncreated entities in the trivial sense.

We can, nevertheless, deem it adequate to speak about reasons as produced if we take seriously what Skorupski defines as *actuality*. The actual is the totality of everything we can refer to (see 2010: 425f). Reasons only have normative force for us *in so far as they are actual*.

If the objectification thesis is true, the only actual reasons are those being actual in relation to our particular mind. Reasons we could refer to if our mind was different are not actual and therefore of no relevance. It does not even make sense to ask for a totality of non-actual reasons. The number of possible, but not actual reasons might not even form a finite totality. The totality of actual reasons, however, is finite and distinct for every existing mind and even

for every particular state of mental development. The realm of actual reasons can therefore be described as *evolving with creatures in space and time* and, in this sense, as *produced.* This does not stand in opposition to the basic (and trivial) Fregean rationale of distinguishing between *productive minds* and *unproduced thoughts*.

An alternative to the objectification thesis would hold that there is a realm of objective irreals and that cognition is the psychological capacity to understand the structure of this realm. Forms of neurological organization might be variable, but any form of neurological organization that constitutes cognition – defined as self-determined thinking – is responsive to the realm of objective irreals. Accordingly, there is but one possible form of self-determined mind and all minds having the capacity for self-determination move in the same irreal realm and respond or comply to the same irreal structures. According to this picture, reasons are mind-independent in a strict sense that involves more than what is required for the phenomenology of grasping. It involves assuming a realm of objective irreals being determined by the structure of the universe and being mandatory for all possible forms of cognition.

The controversy we face, even if we take the Fregean distinction for granted, is a controversy between *thoughts as objectifications of mind* on the one hand and *mind as complying to strictly mind-independent objective irreals* on the other hand. It seems that Skorupski tends to the latter side of the controversy. As the next subsection will reveal, he can at least convincingly reject the view that objective irreals are *solipsistic* objectifications of an individual mind. However, this gives rise to a new controversy – a second controversy about the nature of thinking.

3.3. The shareability of thoughts: a second controversy about the nature of thinking

The previous section discussed a controversy about how to understand the relation between mind and the abstract propositional entities that are grasped in thinking. One side of the controversy holds that the propositional entities, which Skorupski describes as objective mind-independent irreals, are *objectifications of mind*, potentially different for different minds. In distinction, the other side of the controversy holds that objective irreals are *independent of any particular mind*.

There are strong considerations in favour of the latter view. It seems obvious that, in many domains of life, people training their reasoning capacities come to the same conclusions, while deviating views sooner or later turn out as based on insufficient training, lack of diligence or strong subjective factors blurring or hindering insight. Apart from that, it might seem obvious that mind is not a solipsistic machine producing is own objectifications. Mind is much too irritable by other minds coming to other conclusions. A reflective creature genuinely cares for the insights of sufficiently developed others.

Skorupski captures this basic fact about the nature of reasoning by formulating what he calls the "convergence thesis", portrayed in the next subsection. He locates reason epistemology in a dialectic between *individual spontaneity* and *convergence with a community*. As I will make clear in the subsection thereafter, convergence is a pillar of reason epistemology in such a genuine way that it rules out an individualistic objectification thesis.

However, the shift to communities as genuinely involved in reason epistemology gives rise to another controversy, structurally analogous to the first one: One side of the controversy holds that objective irreals are *objectifications of communities of mind*. The other side holds that objective irreals are *independent of any particular community of mind*. According to the latter picture, we share a community of mind because we have the same mental capacities. Everybody having these mental capacities recognizes the same logical structures. According to the former picture, by contrast, we recognize the same structures because we developed our mental capacities together. Mind, according to that view, is a communal product – community is not only a necessary assistance for developing mental capacities.[53]

Skorupski is again leaning to the stronger mind-independence claim. However, his arguments in the second version of the controversy are less strong than his arguments in the first one. Likewise, Frege, who speaks about the realm of thoughts as a "common treasure" of humanity[54], fails to provide arguments for anything stronger than the view that thoughts are shareable and

[53] A very powerful formulation of this controversy is provided by Baier, who dates the discussion back to Locke. Locke says that we are a mind community "furnished with like faculties" (Baier 1997b: 1). While Locke could be interpreted as saying that we need a community of beings furnished with like mental faculties in order to develop these faculties, Baier understands the thesis in a stronger way. She thinks that "a closer community is involved" (ibid.) in the development of mind.

[54] "denn man wird wohl nicht leugnen können, daß die Menschheit einen gemeinsamen Schatz von Gedanken hat, den sie von einem Geschlechte auf das andere überträgt" (Frege 1892: 30).

independent of an individual's mind. This view, however, is compatible with both sides in the second controversy.

3.3.1. The Convergence Thesis

A core part of normative epistemology that was described in section 3.2. is spontaneity – a capacity of responding to normative considerations in virtue of one's very nature as a subject. If a response is truly spontaneous or not, is something we can, at least in principle, find out by first-personal self-examination. Spontaneity "is typically marked by a certain experienced or felt harmony" (2010: 407). This experience is essentially first-personal and can be characterized as "an experience, impression or spontaneously persuasive representation of a given response […] as reason-supported: normatively apt, proper." (ibid.)

However, this experience does not necessarily mean that we recognize a reason correctly. Skorupski says that "a purely first-personal equilibrium is not enough for warrant. It needs to be tested against, and can be defeated by, the normative responses of others." (ibid.: 412) Skorupski formulates what he calls the convergence thesis:

> if I judge that p, I am rationally committed to holding that either inquirers who scrutinized any relevant evidence and argument available to them would agree that p or I could fault their pure judgments about reasons or their evidence (ibid.).

The convergence thesis says that whenever an agent who shares our own "epistemic state" – made up by the evidence available from our current perspective – disagrees with our judgment, we either need reasons for blaming it on "incompetence, inattention, or insufficient care" or will find ourselves incapable of claiming warrant (ibid.: 412).

In cases of disagreement with others "who relevantly share my epistemic state", Skoruspki explains, "I might start to wonder whether my normative disposition is indeed spontaneous" or "I might conclude that it is genuinely spontaneous but also wrong" (ibid.). In the latter case, the agent might find that she must educate her spontaneity in discussion and interaction with others. Alternatively, she can conclude that the divergence in this particular case just shows that there are "no genuine reasons in this domain" (ibid.: 413). In this case, divergence constitutes no genuine problem.

There are different ways of solving a problem posed by diverging judgments. Each case shows that human reasoners are subject to a strong

necessitation of establishing convergence with others. It is more than just external pressure. It also seems to be more than just a psychological need for being supported by fellows. Skorupski describes convergence as a core part of normative epistemology – working as a *dialectical counterpart* of spontaneity.

3.3.2. Reason epistemology in the dialectic between spontaneity and convergence

Spontaneity and convergence are described as the "twin pillars" of normative epistemology (Skorupski 2010: 413). According to Skorupski's theory, the road to knowledge about reasons involves a dialectic between spontaneity and communication. Both ends of the dialectic, according to that picture, are equally important. Spontaneity as an individual disposition can only develop a potential of getting normative truths right when it is educated, trained and continuously tested within a community. Moreover, describing the two epistemological elements as twin pillars seems to indicate more than just a practical importance of communication. Emphasizing the two elements as the antithetical ends of a continuously employed dialectical process assigns convergence a place in the very core of reason epistemology. This makes the communicative element so genuine that spontaneity cannot even be understood as a possible road to normative warrant without the embeddedness in an active communicative context. Spontaneity can only be understood as a disposition for normative responses if it is understood as a disposition that is genuinely susceptive and vulnerable to the perspective of others.

A spontaneous response is characterized by a first-personally experienced normative harmony. In a truly spontaneous response, this harmony is unquestioned. The subject, in its very response as a subject, is "unified" or "wholehearted". However, this unquestioned normative harmony is essentially fragile. It can always break down (ibid.: 408). The experience of normative dissonance makes us incapable of achieving true spontaneity and urges us to question our dispositions. It urges us to "sort out the problem, either by achieving harmony or finding a convincing way of explaining the disharmony away" (ibid.). Even if we currently feel perfect harmony, certain pieces of evidence might give us a reason to question this harmony, as Skorupski emphasizes (ibid.: 409). A perspective beyond the immediate subjective experiences always matters. Thrown back on their immediate subjective responses, humans have no chance of gaining reliable normative knowledge.

Accordingly, the radical individualistic version of the objectification thesis, discussed in section 3.2., is very implausible. The idea that an individual mind generates a virtual space of reasons in order to entertain its activity of sense-

making and self-understanding seems absurd considering the role of others in understanding and sense-making. The role of others in these activities is essential and constructive, not only practically limiting. Unless we assume the remote and unlikely possibility that the very idea of other people is an objectification of our mind, we can reject the objectification thesis in its radical individualistic disguise. It can count as established that a mind responding to objective irreals responds to entities that are clearly independent of this particular individual mind. If we want to understand the realm of objective irreals as a created virtual space, as mentioned before, it is at least a space we share. It is a virtuality accessible by more than one mind, rather than our own mental objectification.

This insight can be used as an argument in the first controversy, but still leaves the second controversy unsettled. In the final subsection, I will give a formulation of this controversy and discuss if and how Skorupski's Fregean approach can relate to it.

3.3.3. A controversy about the relation between communities and the reasons they share

To sum up this section, it has been convincingly established that thoughts, the abstract propositional entities grasped in thinking, must be independent of the individual mind. It is implausible to envision the realm of objective irreals as a virtual space that the individual mind creates for itself by objectification, as claimed by one side in the first of the discussed controversies. As opposed to that, thinking, or reasoning, is genuinely susceptible and vulnerable to the reasoning of others. Convergence is not only a practical or psychological need, but a component essential to the core of reason epistemology. Convergence is one of the twin pillars of reason epistemology, forming the dialectical counterpart of spontaneity, without which individual spontaneity could not even be understood as a route to normative knowledge. *Epistemologically*, we can gain knowledge of reason relations by combining spontaneity and convergence. *Ontologically*, reason relations are, by Skorupski, described as objective mind-independent irreals.

Understood as a substantial ontological claim, there are two conflicting ways of interpreting the mind-independence claim about objective irreals: According to the first interpretation, which we could label the *Radical Mind-Independence view*, reason relations hold independently of any community of mind. According to such a view, individual mind needs the exchange with a community with sufficiently developed reasoning practices in order to develop its own reasoning capacities properly. However, the particular way these

reasoning capacities work, the particular outcome they produce, is not shaped by particularities of a concrete closer community insofar as reason is developed properly. There are mind-independent normative structures, and everyone furnished with the right kind of faculties has the potential to become responsive to them, given the right training.

There is, however, a different view to describe the ontology of objective irreals. There is a view we could label the *Moderate Mind-Independence view*. This view allows us to speak of *individual mind-independence*, while maintaining what we could call *collective mind-dependence* for objective irreals. The idea of collective mind-dependence needs further specification, which will be provided in the next chapter. The point of this interpretation is to grant, on the one hand, that reason relations must be described as independent of an individual mind and deny, on the other hand, that there are normative structures independent of communities in which individual mind is shaped.

A researcher paving the way for the moderate view is Annette C. Baier, who argues that not only a common biological nature, but "a closer community is involved" (Baier 1997b: 1; see also Baier 1997a) in the development of mind. Baier distinguishes between those who take "reason to be an inborn capacity, complete in each person" and those who take "it to be the socially cultivated outcome of a certain inborn intelligence and capacity for language" (1997b: 12f). She takes the latter side in this controversy. Baier, thereby also building on the works of Wollstonecraft (2007) and drawing from developmental psychology, defends a "social view of reason". She emphasizes that such a view "does not doom one to undervaluing independent thinking, nor to overvaluing deference to the thought-community in which one grew up." (1997b: 14). Baier wants to bring to our attention that independent thinking, as well as individual intending, is a mental process that we learn to do by doing it with others. When we consider this fact about the genesis of our mental capacities, we realize, according to Baier, that the individual mental activity is derivative on a shared or common activity. According to a genuinely social view of reason, we could say that the shared activity is prior to the individual one. This priority must be understood both *temporarily* during development and *logically* in determining the internal structure and content displayed in the mental activity.

Returning to Skorupski's picture, the explanation for convergence, according to a social view of reason, is not that every reasoner has the same biological equipment for discovering mind-independent logical structures. It is, instead, that all reasoners who have a capacity for self-determination are part of a collective that creates its own common structures. Self-determination

succeeds when the individual makes proper use of the commons. The self-determined mind, according to a social view of reason, is essentially a common product. Community is an essential part of its ontology, shaping the structures of mind, not only a necessary means for unfolding its innate structures.

If we adopt this view, we can distinguish between the *individual mind* and the *mental commons*, or the "commons of mind" in Baier's terminology. Accordingly, a possible interpretation of the ontology of objective irreals could be that they are part of the mental commons, but independent of individual mind. The realm of objective irreals, according to that view, can be conceived as a shared virtual space – a virtual commons that is underlying the activity of individual minds. Individual minds have a capacity of responding to items of the shared virtual space. The externality of the shared virtual space makes it possible that these responses can succeed or fail and that they can even fail without the individual immediately recognizing its failure. A shared virtual item can be recognized correctly or incorrectly. However, the externality in this picture is of a kind that distinguishes the shared items of thought from objects of the external physical world. Thoughts, or the propositional items of thought processes, are not external to mind in this latter sense but underlying the very own activity of a self-determined mind. The distinctive phenomenology of self-determined thought involves the virtual commons in an essential way. The virtual commons are, we could say, *actualized* in an individual's self-determined thought. The individual experiences this actualization as grasping, as complying to the structural demands of something given as a matter of fact beyond its individual will.

We can see the mental commons as an irreal or virtual space having some of the most central properties that Frege takes his "third realm" to have – an ontological category between external world and subjective imagination. However, interpreting the realm of thoughts in light of Baier's social view of reason adds a distinctive note to the Fregean picture. It makes the individual mental process understandable as grasping, rather than generating, while maintaining that objective irreals are *products of objectification*, as introduced in section 3.2. As also introduced in section 3.2., the irreal domain of reasons can be described as virtual in order to emphasize that it is contingent, could always be otherwise and is not determined by external necessities, although it presents itself as objectively forcing and unchangeable to the individual that finds itself within its coordinates.

Does Skorupski offer any arguments for or against any of the sides in this second controversy?[55] There are certain aspects that make it seem like he adheres to the more radical interpretation of the mind-independence claim. Still, he considers some ideas at the end of *The Domain of Reasons* (2010), which point towards the possibility that there is more than one correct way of reasoning about the world. That notwithstanding, Skorupski considers this possibility as barely relevant. Accordingly, he offers no compelling argument against what I formulated as the Moderate Mind-Independence view.

Let us briefly consider what Skorupski has to offer, before the next section will reject Radical Mind-Independence. Skorupski emphasizes, especially in his argument against constructivism (2017), that neither spontaneity nor convergence, nor the presence of both features *make* a reason claim true. The dialectical interplay between individual spontaneity and communal convergence can only be understood as our best way towards warranted reason beliefs. A conclusion we must not draw from Skorupski's outline of reason epistemology is that we can derive any truths about reasons directly from the presence of either spontaneity or convergence. Even the presence of both is only justifying our reason beliefs, but not *making* the reason statement itself true. This means that the results of the dialectical process are not by definition the normative truths. Applying a particular epistemological procedure does not produce or construct normative facts, as a version of constructivism that Skorupski opposes (2017: 600) would hold. According to Skorupski, normative claims are not true because they are the results yielded by a specific procedure. If they are true, they are "simply true" (2017:600). This makes Skorupski a Reasons Fundamentalist in Scanlon's sense. It further suggests that Skorupski understands reason relations to hold independently of both individual mind and any sort of mental commons or collective mind.

Nevertheless, Skorupski seriously examines the possibility of diverging communities having their respective community-relative truths (2010: 502ff). Like mentioned before, he slightly favours the view that communities are only epistemic limitations to normative knowledge, not the ontological origins of reason relations. An observation driving Skorupski to this view is the fact that reason relations have the function to "generate a world theory" (ibid.: 503). Different communities might achieve this in better or worse ways, so that world theories are unlikely to be the products of communities, according to Skorupski. Moreover, when communities fuse or interact, they quickly start to remodel their world theories. When they do so, they make revisions not at will,

[55] I am very grateful to John Skorupski for extensive personal feedback in the beginning of this work and for urging me to develop this position.

but after serious consideration of evidence. Finally, however, we can state that Skorupski, for his part, remains open for the possibility that there might be two equally successful cognitive traditions. In this case, we would "have to accept that there was more than one optimal world theory" (ibid.: 503). We can either "accept that there is some unknowable truth about the world as it really is" or we could conclude that "the question which world theory is really true is empty" (ibid.: 504). Skorupski, after all, gives no definite answer. To conclude, he just holds that his theory gives us "as much cognition-independence as common sense requires" and emphasizes once more that "we do not in any sense ´construct reality´" (ibid.: 504).[56]

In this book, by contrast, I want to show that questioning or asking beyond this "common sense" can be philosophically rewarding and that taking a qualified stance in the second discussed controversy matters for understanding reasons and normativity. Some senses of "constructing reality" might not be that odd to assume, in light of contemporary sciences, and an awareness of such constructive processes contributes to our philosophical picture. In the remainder of this chapter, I will reject the Radical Mind-Independence view (3.4.) and propose to frame reasons as potential mental states (3.5.). This is a helpful framing of reasons within the moderate view.

3.4. A Darwinian Dilemma for Radical Mind-Independence

Reason relations, according to Skorupski's ontological position, are objective irreals – neither parts of the external physical world nor products of subjective

[56] I think that Skorupski's rejection of the picture of construction is prejudiced in at least two ways. For one thing, there might be forms of mental productivity that are constructive in nature, without being consciously intentional processes started at will. For another thing, Skorupski seems to link construction or creation exclusively to the concept of fiction as a creative act of individual imagination. Indeed, Kurt Sylvan in his criticism from the standpoint of Kantian Constructivism accuses Skorupski of basing his rejection of constructivism on the mistake to treat it "as of a piece with fictionalism" (Sylvan 2016: 620). Linking construction and fiction might be one problem. There is, however, another problem that might be even more crucial. It seems narrow-minded to suppose that a product of individual imagination is the only type of fiction that there is. According to a broader definition of fiction, collectively developed products or myths and narratives evolved within communities might also count as fictions. Moreover, such a notion of fiction, which we could call collective fiction, might in fact be a fitting description of what the items forming the shared virtual space of reasons are. I will talk about collective fictions in the next chapter.

imagination or fiction. According to the corresponding epistemological position, the route to knowledge of this kind of irreal items is a dialectical process between spontaneous individual judgment and finding convergence in communication with other spontaneous creatures. However, Skorupski emphasizes, normative truths are mind-independent and do not follow from the presence of empirical phenomena such as spontaneity or convergence. The structures of the domain of objective irreals hold independently of the socially embedded psychological processes in which mind is trying to grasp them.

According to a view we can label *Radical Mind-Independence*, reason relations hold *independently of any particular mind or any particular mental community* gaining access to them. According to an alternative view we can label *Moderate Mind-Independence*, the structure of the realm of objective irreals is *independent of individual mind* (individually mind-independent), while being a *product of a mental community* (collectively mind-dependent).

This section points out why Radical Mind-Independence requires a cut between the epistemology and the ontology of reasons that cannot be defended successfully. I will argue that a normative theory must map the ontology and the epistemology on each other in order to make sense of its epistemological part at all. The systematic cut between the ontological picture and the corresponding epistemological story produces a sort of Darwinian Dilemma. This makes a moderate view, acknowledging collective mind-dependence, preferable.

3.4.1. A tension between the epistemology and the ontology of reasons

A crucial implication of the thesis that objective irreals are mind-independent is that a correct application of the epistemological procedure is no guarantee for seeing the normative relation correctly. Skorupski's theory of normative knowledge can count as fallibilistic. Fallibilism is a consequence of a strictly mind-independent conception of objective irreals.

The fact that a response is truly spontaneous and converging with the responses of others makes it likely that it is a correct reason recognition. It therefore *justifies* a reason belief. However, it does not *make* a reason statement true. What makes a reason statement true is the fact that stands in a reason relation – a fact that is objectively irreal. Consequently, there is a conceptual distinction between reason relations that hold objectively and reason judgments that can reasonably be expected of a socially embedded subject following the route of spontaneity and convergence.

In accordance with that distinction, Skorupski introduces two distinct notions of autonomy. Autonomy is here understood as a property of a judgment that is not only self-determined – done from genuine understanding having apperceptual features – but also self-determined on a correct basis. Now, saying that the reason judgment on which a self-determined act is based is correct can mean two different things: On the one hand, there is "Hegelian Autonomy", measured entirely against the common sense of a community. This notion of autonomy serves as a basis for moral evaluation of persons. On the other hand, there is "Kantian Autonomy", a state of having full warrant about reason relations. This notion of autonomy is only constituting a high personal ideal that we cannot expect morally. Accordingly, Hegelian autonomy is what we achieve when we follow the dialectical route of spontaneity and convergence in the best way that is reasonably available to us. Kantian Autonomy, as distinct from that, is a concept referring to a state that we must conceive as in principle possible, while there are no criteria that tell us for sure when we have reached that state.

The sophisticated picture that allows for the conception of Kantian Autonomy entails a gap between reason relations and our psychological faculties of recognizing them. Some sort of ontological gap between an object and our psychological faculties of recognizing this object is natural to assume in the case of external physical objects – objects that have a causal influence on our receptive apparatus. An ontological gap between an object of recognition and the procedure of recognition is, however, more difficult to defend in case of irreal objects of cognition. It is a crucial feature of such objects that they do not exert a causal power on us that could be undercut. The failure of recognizing them correctly must be explained differently.

Some philosophers argue that this is only possible if the ontological structure of reasons and the epistemological procedure of recognizing them are structurally interlocked. Epistemology and ontology must be, we could say, mapped on each other. What a reason is, according to a view that maps the ontology on the epistemology, must be defined in terms of what it is that certain epistemological procedures yield. Otherwise, it would become unintelligible what reason are and how our procedures should aim at them at all.

As Sylvan criticizes, there is a tension between Skorupski's spontaneity-based epistemology and his cognition-independent metaphysics that makes it questionable how spontaneity can be a basis for reason recognition at all (2016: 619ff). If truths about reasons are not truths that are "derivable from constitutive principles of self-determination" (ibid.: 620), so Sylvan, we could not understand spontaneity as leading to warrant – not even potentially. Instead, it would resemble to what Sylvan, borrowing from McDowell (1994),

calls ´frictionless spinning in the void´. Sylvan therefore concludes that, if Skorupski wants to maintain his spontaneity-based epistemology, he must give up cognition-independence and turn to Kantian Constructivism as envisioned by Korsgaard (2008b) or Markovits (2014).

Unfortunately, the discussion between Sylvan and Skorupski stays focused entirely on the notion of "deriving" reasons, as Skorupski's reply to Sylvan (Skorupski 2017) shows. In his reply to Sylvan, Skorupski admits that every normative truth corresponds to a possibility of being recognized in a certain psychological process: "Truths π are a reason for x *if and only if* x can tell, recognize, know that they are" (2017:600). However, this does not amount to constructivism according to Skorupski, since, as he says, the right-hand side of this biconditional does not *make* the left-hand side true (ibid.). In chapter 2, I rejected the claim that we can derive what reasons we have from what is constitutive of our capacities to respond to them – thereby supporting Skorupski on that matter.

Constitutivist constructivism fails both as a directive normative theory and as a metaethical analysis in the narrow, conceptual sense. However, I stipulated that asking for what reasons are independently of the ambition to derive what reasons we have, and independently of the mere conceptual question of how we understand reasons, can help us to understand the force of reasons in a more comprehensive and critical way. Taking a more comprehensive approach to the phenomenon of normativity, we can still make room for the claim that the epistemological route to reasons and the ontology of reasons must be structurally interlocked.

We can gain insight into the nature of reasons when we understand that facts that are reasons are, by definition, facts that are open to be discovered by certain procedures – even though we do not derive the truth of a reason statement from the fact that the epistemological procedures have been followed correctly. Even though we do not get any help in reasoning, we gain a philosophically interesting perspective by understanding what these entities are that are open to be discovered by those procedure – what they are, what they have in common and how they are determined – materially rather than conceptually, as distinguished in the first chapter.

If Radical Mind-Independence about objective irreals were true, there would be a systematic cut between the epistemology and the ontology of reasons. Sharon Street formulated a "Darwinian Dilemma" for theories involving such a cut. In subsection 3.4.2., I will present the dilemma as it was introduced by Street. In subsection 3.4.3., I will show how Skorupski, and even Scanlon, can avoid the dilemma in Street's formulation and offer a new, refined formulation

– designed so as to tackle Skorupski's Irrealist Cognitivism, as in principle every view embracing the idea of mind-independent reason relations.[57]

3.4.2. Street's Darwinian Dilemma for realism

In her article *A Darwinian Dilemma for Realist Theories of Value* (2006), Street challenges the thesis that normative realism is compatible with natural science. She argues that realist theories of value run into a so-called Darwinian Dilemma. This dilemma, Street holds, is "not distinctly Darwinian, but much larger" (2006: 155), making normative realism incompatible with "any good scientific explanation" (ibid.). The only way of sidestepping the dilemma, according to Street, is a form of (constructivist) antirealism (ibid.: 152).

Street's depiction of the Darwinian Dilemma goes as follows: There is an undeniable evolutionary influence on our evaluative attitudes. If evaluative truths hold independently of how our attitudes are shaped, we need a theory about the relation between the evaluative facts and the processes in which our attitudes evolved. There are two possibilities, none of which is acceptable so that we are presented with a dilemma. We end up with the first horn of the dilemma if we hold that there is no relation between evaluative truths and the development of our attitudes (2006: 121). This view entails that if a particular attitude gets the truth right, this match is a product of mere chance. Accordingly, most of our attitudes would be "off track" (ibid.: 122). This horn of the dilemma presents us with a hardly acceptable form of scepticism. The second horn of the dilemma is reached if we assert a relation between evaluative truths and the development of our attitudes (ibid.: 125). We could argue that ancestors who were able to grasp normative truths were favoured by evolutionary processes. Accordingly, there is a kind of "tracking relation" between the evolution of evaluative attitudes and evaluative truths. Evolution of evaluative attitudes tracks evaluative truth. Street argues that this assumption is "unacceptable on scientific grounds" (ibid.: 109).

According to the antirealist picture defended by Street, "each of us begins with a vast and complicated set of attitudes" (ibid.: 153), which are continuously subjected to selective pressures. Street argues that the relation between selective pressures and evaluative truths is best described by an "adaptive link account", instead of a "tracking account". Street describes the relation as follows:

[57] I am indebted to John Skorupski for sharing his objections to Street with me (mainly with respect to the idea of "tracking"), so that I could adapt the argument of the Darwinian Dilemma accordingly.

> Evaluative truth is a function of how all the evaluative judgements that selective pressures (along with all kinds of other causes) have imparted to us stand up to scrutiny in terms of each other; it is a function of what would emerge from those evaluative judgements in reflective equilibrium. (ibid.: 154)

Evaluative truths being "a function" touches the dependence relation or, as Street formulates, the "direction of dependence" (ibid.) between the truths and our attitudes. According to the realist, the truths are "prior", while according to the antirealist, the evolutionary causes shaping our attitudes are "prior" (ibid.). Priority of evolutionary causes in relation to truth means that these causes "gave us our starting fund of evaluative attitudes, and evaluative truth is understood to be a function of those attitudes." (ibid.)

The idea of normative truths being a function, rather than a fund that exists and must be tracked, can be identified as *irrealist.* Moreover, emphasizing that they are a function of existing attitudes that evolved under evolutionary pressures is a way of describing normative truths as products of evolutionary processes – at least according to Street's picture. In contradistinction, Skorupski, insofar as he defends Radical Mind-Independence about objective irreals, seems to combine *irrealism* with the idea of a *"mind to truth" direction of dependence*. That means Skorupski shares irrealism with Street but opposes her with regard to the direction of dependence. The following subsection will point out how Skorupski's irrealism can avoid the implausible assumption of a tracking relation but runs into a Darwinian Dilemma nevertheless if he refuses to describe objective irreals as produced or created in a crucial sense.

3.4.3. How a theory of mind-independent reason relations avoids the dilemma in Street's formulation

Street suggests depicting evaluative truth as a function of evolutionarily evolved attitudes. The "direction of dependence", according to such a picture, goes from our psychology to evaluative truth, not vice versa. Skorupski, by contrast, argues for mind-independence of normative truth. This section points out how his type of account, building on the concept of a reason relation and thereby making normative facts indexical to subjects, succeeds in avoiding Street's formulation of the Darwinian Dilemma.[58] The formula of the reason

[58] The argument would also work for Scanlon's formulation of the reason relation. For a more general application of the idea see my manuscript *Reason Monolithism: A New Darwinian Dilemma* (Mähringer 2020). I thank the participants of the phD seminar in practical philosophy at Lund, especially Anton Emilsson, Frits Gåvertsson, Marta Johansson Werkmäster, Marianna Leventi, Robert Pál and Alexander Velichkov for helpful feedback.

relation, employed by Skorupski, Scanlon and others, allows us to understand the truth about normative facts as mind-independent without assuming the scientifically dubious "tracking relation" between our attitudes and evaluative truth that was criticized by Street.

Indeed, the idea of a "tracking relation" seems the most widely criticised element of Street's theory (see e.g. Deem 2016; Copp 2008; FitzPatrick 2014; Enoch 2010). In the following, I will only focus on the objection that is relevant for the confrontation between Street and theories about mind-independent reason relations. The objection targets the very idea that a theory of mind-independent normative truth requires such a thing as a "tracking relation" between mind and normative facts. FitzPatrick, for example, holds that such an idea indicates "either bad science or a poor view of the moral facts, and so can be discarded without much fuss" (2014: 239). As FitzPatrick holds, not all our faculties should be understood as the product of evolution, as "we clearly engage in […] intelligent and independent thought when we do higher mathematics, or science, or metaphysics, or philosophy generally (including philosophical reflection on these very issues!), among many other things" (ibid.: 242). He claims that we developed these intellectual capacities "despite the evolutionary influence in the background" (ibid.: 242f), not *because* of it or *as a product* of its operation on us. In a similar vein, Copp (2008) insists that, no matter how strong the evolutionary force was, there are other forces driving our development as well – such as "culture" or "deliberation and reflection" (Copp 2008: 190).

Most importantly, this line of criticism rests on a distinction between being the *product of evolutionary forces*, being the *product of culture* and being the *product of reflection*. Against the background of a classical genome-based theory of evolution in the tradition of Darwin, we can probably allow for such a distinction. It seems that both culture and the activity of reflection can change judgments, while the genetic equipment that was shaped by evolutionary forces remains the same. So, we can admit that there are influences on our judgments different from the evolutionary influences of genetic mutation and selection. We can distinguish between *genetically hard-wired dispositions*, such as for example a disposition for empathy or a disposition for helping behaviour, on the one hand, and *reflective faculties* that help us to evaluate these dispositions, considering consequences and implications, on the other hand.

The concept of a "reason relation" as employed by Skorupski, Scanlon and others, offers an elegant solution to the tracking problem that is in line with this rationale. A tracking problem, we can agree, occurs only for a naïve understanding of moral realism, which assumes a substantial set of attitudes or actions that are, we might say, good in themselves or required as such. As

opposed to that, describing normativity in terms of a reason relation is a way of avoiding substantial assumptions of this kind. According to the concept of a reason relation, a fact p, or a set of facts π, being normative means nothing but that it is standing in a relation R. According to Skorupski's formulation, this relation holds between a fact, a point in time, a degree of strength, a subject and an action or attitude: R (π, t, d, x, φ). Skorupski claims that this relation holds independently of whether we recognize it. However, if we have all available evidence and employ our cognitive faculties correctly, we can recognize that it holds. The interesting point is that this idea postulates no mind-independently correct attitudes, because each possible attitude is only normatively required in relation to a particular subject at a particular point in time. I think that even reasons according to Street's own account could be expressed in this way.[59]

Now, we can conceive the subject variable x as standing for a particular subject as it is shaped by whatever evolutionary, cultural and personal influences we can think of. Moreover, the time variable t accounts for particular circumstances, shaped by various natural factors. Thus, the idea of a reason relation allows for particularism in the most radical way. There is no need for tracking because the reason relations, in which a subject finds itself, are indexical to this very subject at this very point in time. They are, we could say, actual only for this very subject at this very point in time. The subject recognizes this if it employs its cognition correctly. There are no "mind-independent existents" to track. If we make the distinction between *reflective faculties* and *genetically hard-wired dispositions* claimed by Street's critics, we may now find that the idea of mind-independent reason relations does indeed escape a Darwinian Dilemma in Street's formulation – a dilemma between scepticism and tracking.

However, in the subsection to follow, I will reject this distinction and reformulate the Darwinian Dilemma to target the idea of mind-independent reason relations (see also Mähringer 2020). The reformulated Darwinian Dilemma opens up between scepticism and what I call Reason Monolithism, an assumption that is as scientifically dubious as the idea of a "tracking relation".

[59] The remaining difference between Street and Skorupski, as I can see it here, is that Street makes the assumption that all the facts standing in that (irreal) relation to us are conceptually related to, and therefore derivable from, desires we have. It seems clear to me that this is a normative assumption, as it is presented by Street. We do not have to find it plausible, but we can understand why such a suggestion can be attractive in a secular world view in which people tend to reject justifications that do not ultimately trace back to individual interests or desires. However, this normative position leaves the metaethical analysis untouched.

3.4.4. A new formulation of a Darwinian Dilemma for radically mind-independent reason relations

In this subsection, I will argue that the distinction between reflection and contingent dispositions, on which the rejection of Street's Darwinian Dilemma rests, is scientifically questionable. To begin with, we must settle for a workable terminology when talking about the distinction. To capture all versions of the objection, I propose to frame it as a distinction between *cognition* (faculty of reason, reflective capacity) and *objective psychological dispositions*. The latter include all features that a cognizing creature can objectively have, and that can, in principle, be empirically observed and established. It includes all biological and psychological features that have been shaped by genetic selection, culture, personal history and the like. All these objective psychological dispositions, as I take it, can be understood as a product of evolution – even if this exceeds the distinctly Darwinian theory of evolution.[60]

Equipped with this distinction, we can look at the model of a reason relation in more detail: R (π, t, d, x, φ). There is a set of facts π that is a reason (of strength d to φ at time t). The variable x denotes the subject for whom π is a reason. It embraces all the psychological dispositions this subject might have. *Cognition*, in this model, is the responsiveness to π as standing in R. Cognition can grasp π in the relation R. Cognition, for example, can understand that a disposition for liking ice cream relates to the act of buying ice cream – making the fact that there is ice cream available a reason for buying it, given that there are no stronger reasons related to this. Cognition is the faculty to understand how any possible consideration (i.e. "there is ice-cream") *relates* to our dispositions ("I like ice-cream") and our possible actions ("I buy ice-cream") and how a particular related consideration *relates* to other considerations (e.g. "Ice-cream is unhealthy", "ice-cream is expensive" etc.). While objective psychological dispositions may vary widely from subject to subject, cognition understands how these dispositions *relate* to actions and other dispositions – how they *are* related in the normative domain.

However, neglected by many philosophers, there are two different ways of conceptualizing cognition. The picture indicated by the formulations above suggests that cognition is the faculty of understanding how certain variables *are* related. According to an alternative conception, however, a system of cognition is but one model of *relating* different variables, which, however,

[60]We should remember that Street's Darwinian Dilemma is not "distinctly Darwinian, but much larger" (Street 2006: 155).

could also be related according to another model. Cognition, according to the former conception, complies with logical structures independent of mind/ cognition in the strict sense of what I previously defined as *Radical Mind-Independence*. The latter conception, by contrast, depicts cognition as a contingently evolved and multiply expressed faculty of structuring the representation of reality according to norms not directly prescribed by external structures.

We could also pinpoint the difference as being about the question whether the fact that π stands in R is only a perspectival fact or a fact holding from an absolute standpoint – a "standpoint of the universe" that does not have alternatives (see also Veluwenkamp 2017). To be sure, π is an indexical fact. It is only a reason in relation to a particular subject at a particular point in time. However, there is a relevant difference between this indexical fact being *perspectival* or *absolute*. If it is perspectival, only the subject with its particular cognition can see it. If it is absolute, by contrast, the subject for whom it holds *might* be in an epistemically superior position to assess it, because it has more knowledge about its own situation and psychology. However, the reason relation holding absolutely would imply that, at least in principle, every creature with cognition, or more generally every entity having the faculty of reason, can see that this relation holds for that particular subject, given its objective psychological features and situation. The perfect cognizer, having full knowledge about the subject in question, would see this even if she has totally different psychological features herself. She would see this even when the subject *for* whom the relation holds does not see it.

I think that this conception of cognition is scientifically implausible. It involves what we could call *Cognitive Monolithism* or *Reason Monolithism* (monolithism about the faculty of reason). Reason Monolithism is the assumption that cognition has a single perfect expression in which all creatures having cognitive development share to the extent to which they are developed. Cognition is one specific thing. There cannot be two genuinely different forms of cognition. Subjects, in other words, cannot differ with regard to their form of cognition, but only with regard to their objective psychological dispositions. According to Reason Monolithism, there is only one expression of the *cognitive disposition*, while all the variable dispositions are portrayed as *non-cognitive*. The cognitive disposition, by contrast, is conceived as historically and culturally invariable.

It is not easy to prove that Cognitive Monolithism is false. It seems, however, already enough to point out that depicting it as plausible in a scientific world view is extremely difficult. If we want to maintain that all cognitively active creatures have evolved slowly with many historical interim

stages, Reason Monolithism must make two assumptions about the development of cognition. Both are difficult to defend. First, Reason Monolithism must assume that the evolution of cognition is unidirectional and universal in all possible cognitive creatures.[61] According to this picture, a complete and objective mental grasp of the universe is possible. Second, Reason Monolithism must assume that creatures in an interim stage fall short of something that is, in a sense, normative for them. This problem would still be left, even if a complete and objective grasp of the universe itself was in principle possible. Even if a mind possessing a complete and objective grasp of the universe was possible, we need a reason to believe that our mind in its current form is this ultimate upshot of evolution. Alternatively, we must assume that our mind is supposed to reach this ultimate state or that the perfect end state of cognitive development is somehow normative for us as we are now.

Here is a consideration that makes this idea very implausible: Evolution is a historical process. It is most likely impossible to mark the exact historical point at which mental creatures became subjects in Skorupski's sense – creatures having cognition that can respond to reason relations. Likewise, it is most likely impossible to draw a sharp dividing line between subjects and non-subjects in this sense. Accordingly, it is probably more appropriate to conceptualize the development of subjectivity as a continuum[62] with different stages and expressions. All the subjects occurring on this continuum are different versions of subjects. Their very way of being a subject – their very form of cognition – must be conceptualized as being different. Cognitive dispositions, or *forms of cognition*, we could say, *constitute* us as subjects, while some of our objective psychological dispositions might only define which kind of persons we are, without touching our type of subjectivity. Creatures with the same type of subjectivity, according to this rationale, would be responsive to the same reason relations. A creature with a different kind of subjectivity, by contrast, would not think that these reason relations hold, not

[61] Among scientists, it seems that the burden of prove is usually put on those who assume a unidirectional development towards a universally valid or correct end state. A position aiming at this is presented by, for example, Daniel Dennett (2017). His position is, however, ambitious and controversial. Dennett tries to explain in terms of physics, information theory and the theory of evolution why the universe allowed for the evolution of a particular species capable of grasping the very principles of this universe and therefore the whole truth about it. Dennett also includes truths about reasons into his theory (2017: 33-52). There is no room for a proper discussion of Dennett's scientific theory at this point.

[62] „Continuum" here does not mean that the continuity is linear and unidirectional. The evolution of cognition can and did continue in many different directions.

even if this creature possessed every possible information about the other subjects and their objective psychological features.

Finally, I would like to illustrate with an example how this seemingly remote and "deep-philosophical" consideration can be relevant: Think of a person who is only inclined towards egoistic goals and, for the sake of the argument, completely lacks an inclination for altruist behaviour. Yet, it is possible that this person judges that she has a reason to acquire an altruistic disposition – for example because she has been told that this is worthwhile.[63] Ultimately, it seems plausible that our objective psychological features and our tendencies to judge, our subjectivity or cognition, are complexly and closely related. However, they can be conceptually distinguished. Typical humans, for example, do both *have altruistic dispositions* that can be studied and explained empirically and *judge there to be reasons* for altruistic behaviour. Nevertheless, they can *lack altruistic dispositions*, while this still does not settle the question how to judge. The same goes for dispositions to favour close family relations[64] or dispositions to attach to one's place of birth. We could be subjects that judge there to be reasons for developing and following such dispositions and subjects that do not judge so.

I do not claim that this is up to us. All I claim is that, completely in line with Scanlon and Skorupski, that there are reason relations we can discover when we think in the right way and consider all relevant facts. Where I differ from Scanlon is that I can see different people using their reason in the same flawless and serious way and having the same epistemic state, while discovering different truths. The reasons they see present themselves in the same forcing and non-arbitrary way. An example I find convincing is G. A. Cohen's observation that equally intelligent and educated people can be found among Marxists and traditionally religious people, although these world views oppose each other in many of the most fundamental matters (Cohen 2000). Still, I can very well imagine that an adherent of one of the respective positions would feel personal inclinations for the other, while still being unable to think they are right, because it "goes against reason" – against *their* reason, I would specify. I find it plausible to say that they differ in their very *thinking*, not just in some objective psychological features which thinking can be *about*. Different kinds of thinking within the human species are, of course, less

[63] See a similar example about feeling gratitude, discussed by Skorupski (2010: 396).

[64] While the majority of humans probably thinks that this is warranted, some philosophers believe that it would morally enhance us to give up this more exclusive favouring in favour of care for more global goals and the future of humanity (see for example Persson/Savulescu 2012).

profoundly different than different kinds of species-specific forms of cognition. Still, there are varieties of expression both within and between species. If there was one way of responding to the world that counts as "having the faculty of reason", we can wonder: Do all the species that fall short of it have reason to develop it fully? I don't think that this question is as outright absurd as it may sound. However, it is difficult to answer it with a straightforward ´yes´. This is so since all reason relations hold only for particular subjects. So, no subject can have a reason to be a different subject. Which kind of subject to be cannot be read from features we objectively have as the natural persons we are, nor can we plausibly assume that there is a "standpoint of the universe" from which each species is supposed to have a particular form of subjectivity.

To conclude, the assumption of Radical Mind-Independence involves either radical scepticism or a scientifically implausible idea. The scientifically implausible idea involved in a theory of radically mind-independent reason relations is not a tracking relation, but Reason Monolithism, which is implausible in a similar way. To be precise, the scientifically implausible assumption in this case affects both horns of the dilemma. Either we say that there is this one and only perfect faculty of reason, but we can never know whether we have it, or we claim without good grounds that we have it. Street's critics, such as FitzPatrick (2014) or Copp (2008), might be right about their distinction between the forces of genetic evolution and the force of reflection, which makes a tracking relation between Darwinian evolution and normative truth unnecessary. However, they can hardly deny on good scientific grounds that *reflection as a psychological faculty* is separate from *dispositions that shape in a historical and variable process.* The most plausible way of understanding reflection, I suggested, is to understand it as a *variable cognitive faculty* – a faculty of making sense of one's surrounding that is not a monolith but can succeed in various forms.

3.4.5. Mapping the ontology on the epistemology

If we do not accept a monolithic picture of cognition – a picture according to which the evolution of cognition can only go into one direction and approach one perfect state of expression – we must refuse Radical Mind-Independence of reason relations. As argued above, the nature of cognition is most likely not monolithic. Therefore, the reason relations determining someone's situation must be described as relative to, and in correspondence to someone's form of cognition. The *form of cognition* defines someone's being a subject. Someone's *being a subject* is something more fundamental than just the

emotional dispositions and preferences that a subject *has*. It is the way of *relating* these dispositions to each other and to actions and decisions. The way of being a subject, in other words the form of cognition or form of mind, is what determines the relation of the different variables in the relation R. It is, we could say, what *constructs* the reason relation R. As this constructive process could be achieved in different ways, according to a non-monolithic picture of cognition, it is unlikely that it is a response to radically mind-independent irreals.

Accordingly, the ontology of objective irreals must be construed in correspondence to our subjective ways of processing. The ontology must be "mapped" on the epistemology. As I understand it here, this means that we can examine the procedures in which normative insights are achieved in order to get an understanding of what the objects of these insights are. What it does not mean, by contrast, is that we must look for facts relating to what we are like, when asking for what we have reason to do. This is where I differ from Street. Reason recognition, I think, is discovery. Reasons force us, sometimes against our inclinations and personal interests. Accepting only reasons that can ultimately be traced back to a desire, while rejecting other reasons as ill-grounded, is only a very particular normative position.

But which understanding of reasons can we get from considering the way our mind finds out about them? Normative epistemology, as portrayed in this chapter, takes place in a dialectic between spontaneity as a response of the subject and convergence with a community of other subjects. Skorupski calls "establishing warrant […] in practice a collective, archival process" (2010: 124). As discussed throughout this chapter, the crucial question that arises in Skorupski's work is the question whether this statement only describes the epistemological process or says something about reason ontology as well. If we reject Radical Mind-Independence about objective irreals and understand the need to develop the ontology in correspondence to the epistemology, we must conclude that the "collective archival process" does not only *establish* warrant in practice. As opposed to that, we must assume that this process also *creates* reason relations. The realm of objective irreals, we could say, *is* a collective archive – an idea that gave rise to the title of this book. The idea, of course, must be specified and justified in more detail. I will embark for this task in the next chapter.[65]

[65] Skorupski indeed formulates at one point that "what reasons there are depends on a common nature. Norms are grounded in cognitive communities of beings with the same natural dispositions" (2010: 417). Having rejected the idea of Radical Mind-Independence, we can now decide for a clear interpretation of this statement, which occurs in the chapter about reason epistemology and seems to be relativized at other places. We can explicitly read the

If this picture is correct, the realm of objective irreals is a virtual space that is created by communities and that allows for its individual members to develop and exercise the capacity of self-determination. If this picture is correct, Skorupski's concepts of Hegelian autonomy and Kantian autonomy would ultimately collapse. What *autonomy* is, according to this picture, is a *state of self-determination in convergence with one's community*. If there are no objective irreals independently of the established collective archives, a distinct concept of Kantian Autonomy makes no longer sense. As a consequence, there can be no way of having a warranted reason belief independently of other people.

Before defending this view by combining Skorupski's position about the domain of reasons with an empirically grounded picture of the cognitive phenomena he describes, I will outline a few preliminaries for this combination. In line with what has been established so far, section 3.5. will distinguish a deflationist reading of mind-independence from an ontologically stronger thesis, which must be modified. We must acknowledge that, in an ontologically substantial sense, reasons *are* what spontaneity and convergence can *arrive* at.What can spontaneity and convergence arrive at? How we can exactly describe it, is a question for the next chapter. The next section will first suggest a way of modelling the ontology of what our epistemological procedures arrive at.

3.5. A proposal: modelling reasons as potential mental states

This chapter examined the distinction between reasoning and mind-independent entities to which reasoning responds. I showed that an idea of mind-independence is helpful in describing the reality of individual reasoning, but rejected an interpretation called *Radical Mind-Independence*. Instead, I favour a Moderate Mind-Independence view that construes the mind-independent entities an individual mind can discover and grasp in correspondence to its *form of cognition*. Both ordinary experience and a

idea of normative "grounding" in cognitive communities as normative truths holding only in relation to communities and being – in a certain way – communal creations. This reading, as argued so far, is the most plausible way of understanding Skorupski's conception of objective irreals in line with natural science.

convincing epistemological theory suggest that a form of cognition is not solipsistic but shared by several individuals.

This last section comes up with a proposal to define reasons ontologically in structural correspondence to a form of cognition – to a form of mind that can grasp them. The proposed framing "maps" the ontology on the epistemology in that it takes the idea seriously that there are no reasons other than those we can arrive at with spontaneity and convergence. What it is that we arrive at when successfully employing spontaneity and convergence is a distinctive mental state. The difference between the moment of arriving at, of discovering or recognizing a reason and the state of ignorance is this: having actualized a potential of our mind and having it as a mere potential. Ultimately, there is nothing else in the universe than minds with potentials. In this section, I will introduce a neo-Fregean reading of the mind-independence of thoughts, adopting deflationism about mind-independence (3.5.1) and depicting thoughts as configurations of mind (3.5.2.). Subsection 3.5.3. explains the difference between an actual mental state and a potential mental state. It argues why the potential is independent of the actual state of mind, "out there" to be discovered, why still structurally dependent on the mind of which it is a potential.

3.5.1. A deflationist reading of the mind-independence claim

Skorupski adopts the Fregean ontology of thoughts when he develops his idea of the domain of objective irreals. For Frege, thoughts exist in a "third realm", different from both the realm of physical objects and the realm of subjective imaginations. Some interpreters of Frege, such as for example Tyler Burge, were what we could call deflationists about Frege's mind-independence claim. Burge, for example, considered Frege's talk about a third ontological realm as nothing but "a picturesque metaphor" (Burge 1992: 634). This means that there is no deep or substantial metaphysical claim involved in it. It is just a way of capturing how things inevitably present themselves to us. According to such a view, certain truths are just "self-evident" and there is no need to be overly concerned with scepticism. According to deflationism about mind-independence, certain normative truths are fundamental to practices and activities in the sense that you cannot derive them from facts about the practice. In this sense, thought contents are discovered, not created. There is no other sense, according to a deflationist view, in which we could look for anything that is there independently of mind. Burge holds that "reason and judgement – indeed mind - are partly defined in terms of acknowledging the basic laws of truth" (Burge 1992: 649). Accordingly, there is no need within the limits of

reason to question these basic laws. This view is metaphysically innocent and compatible with construing thoughts or reasons as being ontologically dependent on and structurally corresponding to particular forms of mind.

The defence of Reasons Fundamentalism, given in chapter 2, involves the deflationist reading of mind-independence. As argued in subsection 2.4.1., Reasons Fundamentalism is in itself indifferent to whether reasons are dependent on mind in another, ontologically stronger sense. However, we can enrich a metaethical position by asking for the emergence of our reasons in relation to our historically evolved and contingent form of cognition. In the following, I will suggest an understanding of thoughts that accepts mind-independence in the deflationist sense.

3.5.2. A thought as a configuration of mind

A philosopher who could agree with the deflationist reading of mind-independence, Michael Dummett, criticizes Frege's talk of an ontologically independent realm of abstract entities as "a textbook example of philosophical mythology" (Dummett 1991). In addition to the deflationist reading, Dummett introduces another reading of mind-independence, compatible with the deflationist reading.

As opposed to a strong metaphysical reading of Frege's "Third Realm", Dummett suggests a merely "grammatical" interpretation. Against a strong metaphysical reading, he states: "Frege's mistake is to assume that all objects are self-subsistent", while, in fact, "some are intrinsically "of" or dependent on other objects" (Dummett 1991: 249). We can only treat thought contents or reasons as grammatical objects, which, in Dummett's terminology are not self-subsistent objects: „Senses are always „of" something, namely whatever serves as the medium of their expression, in communication or in soliloquy" (ibid.: 262). Dummett here uses a notion of "dependence" that is different from the notion of dependence as logical truthmaking – the notion that allows a thinker to arrive at a truth by deriving it from another truth. Dummett's notion of dependence as being "of" something rather than self-subsistent is nevertheless a pervasive notion of ontological dependence: The idea that a thought or a reason response is a thought or a response "of" a mind deserves thorough understanding. We must take the "being of" seriously if we want to account for the important difference between *grasping something* and *thinking about something*. We can think about a tree, which is a physical object that has an existence independent of mind. In contradistinction, we can grasp the thought "This tree is green", which is an abstract entity. When we grasp a thought, we

do not think about the thought as we think about the tree, *we think the thought itself.*

Paying attention to this difference is crucial for understanding the Fregean idea of grasping as well as Skorupski's idea of objective irreals, to which we are responsive in virtue of spontaneity, not in virtue of being causally affected by them. Thinking a thought, or responding to a reason, must be conceived as an instance of mind being forced into a certain shape. It must not be conceived as mind aiming at or interacting with a shape that has an independent existence and structure. While the forces that bring mind into that shape are clearly beyond its control and awareness, the shapes that mind takes are nonetheless shapes "of" that mind. Metaphorically speaking, we could say they are "made of" mind, they are configurations of mind in the moment they are being grasped. They are, in that very moment, mental states. Dummett brings up an interesting consideration: „senses if they were self-subsistent objects would be the only objects not given to us in any particular way, but whole and entire in the fullness of their being" (Dummett 1991: 256). Senses, Fregean thoughts or reasons, can, at least in principle, be instantiated in our mind "whole and entire in the fullness of their being". This seems to be exactly what Skorupski has in mind when he speaks about the possibility of *warrant* – about the possibility to grasp all the normative relations applying to oneself correctly (Skorupski 2010: 108ff). The aprioricity of reasons guarantees for that. While the objects in the world that are presented to us through our senses are always given in a particular way and from a particular perspective, this does not hold for objective irreals, to which we respond spontaneously. The ideas of spontaneity and aprioricity thus present us with a specific challenge.

The probably most suitable way to meet this challenge is to depict the recognition of a reason as an instance of our mind taking a certain shape rather than as our mind contacting or trying to contact an abstract entity. We can speak, in the same vein as Korsgaard[66], of mind as embodying a reason in the moment of recognizing it. This means that in the moment of recognition *the reason that there is* and *the mental state that occurs* become *identical*. Some people oppose this view (see e.g. Mantel 2018), but their claims, as far as I can see, are only conceptual. They establish that our concept of a normative reason

[66]Korsgaard (2009: 14) speaks of „actions" embodying reasons. Depending on our conception of action, we can of course interpret this as meaning something different from what is presented here. However, if we take "action" as Korsgaard uses it here as mainly meaning an intentional state, we arrive at a similar picture.

and our concept of a mental state are distinct.[67] But they fail to establish that there is anything more in the universe than minds employing concepts – minds being in distinctive states that is. We can say that a reason is a *proposition*. But we can also agree that, ultimately, there is nothing in the world but creatures with *propositional attitudes*. Still, it is possible to say that propositions hold independently of mind, independently of whether we instantiate or grasp them. However, a presupposition for it to make sense to say that they hold is that we *could* grasp them if we had all information. Modelling the propositions that we have not yet discovered as, hitherto unknown, propositional attitudes that our mind has a potential to form, enables us to reconcile these claims.

3.5.3. The instantiation of a mental state as the actualization of a potential

In this subsection, I suggest conceiving of the *reasons that there are as potential mental states*, and the mental state of a reason judgment as an *actualization of a potential*. In the moment in which a reason is recognized, the reason and the occurrent mental state are *identical*. Previously, I suggested that a thought that is instantiated is a configuration of mind. Instantiating a thought, accordingly, is adopting a particular shape of mind – a shape of mind that, however, has an importance for us even when it is not instantiated. Still, the shape that can be instantiated has no subsistence or existence independently of a mind that may take that shape.[68] Let us look at the distinction between an

[67] I am grateful for an exchange with Susanne Mantel (2017 in Munich) and Hans-Johann Glock (2019 in Lund).

[68] To begin with, we must specify the concept of a "mental state" and the idea of a "shape" that mind takes. In the context of this book, "mind" shall be defined as the phenomenon of an explicit, verbally communicable, some might say *occurrent*, stream of reflective consciousness. This narrow notion of consciousness or mind might by far not be the only phenomenon that philosophy of mind can examine and discuss. Nevertheless, there are good scientific reasons, to be pointed out in the next chapter, to believe that this narrowly defined phenomenon is the one relevant for understanding reasoning or reflecting in terms of reasons. The definition of a "mental state" connected to this understanding of mind, can be distinguished from another widespread understanding: Some philosophers have something in mind like *dispositional attitudes*, when they talk about mental states – attitudes such as *beliefs* or *desires* that can occur in consciousness or not (e.g. Persson 2005). However, the idea of a mental state as a static, inert attitude that a person holds is incompatible with the conception of mind as reflective consciousness that is assumed in this book. Reasoning, according to this book, is in the first place a processual instantiation of a series of mental states – a series of instantiations that constitutes the thought process. Each of the mental states involved in this series can be described as a content presented to a subject in a certain way. Examples for contents might be "It is bright in my room." "The sun has risen." or "The

actual mental state and a *potential mental state*: When a reason judgment is instantiated, actually and vividly represented in a particular consciousness, the content turns from a mere potential into an actual mental state representing it – an instance of representing a thought content first-personally. We can draw a distinction between the *potential* and the *actual mental state* - between the general complex of information coding for a *type of mental state*, and the particular *token mental state*.

The distinction between *potential mental states* and *actually instantiated mental states* leads us to the final point of this chapter – the idea that reason responses are actualizations of potential mental states. Accordingly, the Fregean distinction between thoughts as abstract mind-independent entities and psychological instances of thinking boils down to the distinction between complexes of semantic information with the potential to be instantiated by mind and instances of actual instantiations of these complexes. In the proposed neo-Fregean reading, the complexes of semantic information are nothing but mental states not actually instantiated. We do not need a strong metaphysical assumption of "potentials" as existing. We can accept that there are no entities that have subsistence over and above the minds, even though reasons are mind-independent in a deflationist sense – i.e. not derivable for a mind that is not directly responsive to them. While we can think of semantic information as having bearers, for example letters or sounds of language, communication of thoughts or reasons is not essentially the transmission of sounds or signs. Instead, it is essentially the transmission of thoughts or reasons themselves, which is only possible if the participants in a communication are responsive to the same abstract entities. Therefore, it makes sense, for understanding the nature of these entities, to speak of them as forming an ontological category of its own and as existing only in the form of being instantiated in a mind, which, however, is forced into a shape by responding to them, which is not self-generated or private. Therefore, we can speak of potentials independently of actual instantiations.

---

snake is poisonous." "I fear the snake." "I ought to run away." Contents are, we could say, complexes of semantic information. The way in which these contents are given when they occur in a process of reasoning – that is not just in the process of free association, fantasizing or dreaming – is characterized by a particular cognitive phenomenology. When we reason seriously and at the best of our abilities, the cognitive phenomenology has the features of what Skorupski calls apperception – a way of holding a thought content first-personally and rightfully, as a self-determining subject in a causally and spatiotemporally unified framework. To understand the view developed here, it is essential to understand a mental state as an instance of holding or entertaining a thought content in exactly that way. Accordingly, a mental state is *not* an *attitudinal component of the psychological system of an individual*. It is a conscious *occurrence*.

Each mind, to be capable of thinking, to actually be a mind, must have a set of potential mental states it can instantiate. These potentials inherently belong to the mind, are "of" the mind, to use Dummett's formulation. Where these potentials come from and how they are determined is the question that will be picked up in the next chapter. It is, in an important sense, an empirical matter. Metaethics, asking for how we are using the concept of a reason and if there is anything about this concept that allows us to derive what reasons we have in a general form, stays in its place. What the thesis that reasons are potential mental states does not say is that a reasoner can derive that she has a reason from the fact that she can instantiate the respective reason belief. Examining which mental states we can instantiate, and why we can instantiate these rather than others, leaves normative and conceptual questions untouched. Likewise, saying that reasons are potential mental states does not imply that reasons cease to be reasons when we grasp or actualize them.[69] Our subjective reality in grasping is not touched at all. The analysis that reasons are potential mental states is a complement of that subjective reality – saying that beyond that subjective "reality", there is nothing but minds with various potentials, which they are internally driven to realize. This is neither to debunk our subjective experience nor to advocate normative nihilism. It is only to understand our nature and ourselves in the world in a more comprehensive way.

We miss an essential part of what reasons *are* if we do not examine what it is that we do when we understand something as being a reason. The empirical, material approach, with which the conceptual approach is enriched in this book, turns the matter of enquiry upside down. The approach does not ask how we know whether something is a reason, but what we do when we know that something is a reason. The methodology of this approach is to start with the state of recognizing a reason. To grasp something as a reason is to instantiate a mental state. To grasp a reason not merely as a reason, but to grasp it as the reason that is normative for oneself in this moment is to instantiate a very distinctive mental state – the state of representing a particular reason judgment as first-personally warranted. In order to instantiate this distinctive mental

[69] I thank Johan Brännmark for this objection. The problem with it is mainly, I think, that it is based on a misunderstanding of the thesis as a purely linguistic claim. To be sure, it sounds odd to say that we respond to a potential mental state when we respond to a reason. The terms are clearly not linguistically equivalent. My point is, however, that what we are doing when we respond to a reason is realizing a potential mental state. This is a *factual* identity claim. Assuming *linguistic* identity of the terms of description results in questions and objections that do not apply, since I do not make a linguistic identity claim. However, it is rewarding to get such objections, which help to clarify the approach. In this context, I also thank Sven Rosenkranz and Adam Sennett for objections when I presented the idea of reasons as potential mental states at LOGOS in Barcelona.

state, it must – almost trivially true – be a potential of our mind. The worthwhile philosophical endeavour, which benefits greatly from the empirical sciences, is to proceed from this insight into examining the origin and the determining conditions for those potentials of our mind – a starting ground for this is built with the ideas of spontaneity and convergence.

To be sure, the notion of a potential mental state has some vagueness. It is unclear at what point we should speak of a potential. Is it when we can realize it as we are right now? Is it when we can train our minds to reach it over many years? Or is it when humans can reach it, given a sufficient progress of human societies? I think this is a genuine vagueness. It relates to the question when a subject has undergone enough change to count as a different subject. We might simply have to accept this vagueness. But we must also accept that reasons can only be reasons in relation to particular subjects. Our subjectivity, and thus our reasons, can change over time. This is where interesting questions, some of which discussed in the final chapter, can arise.

3.6. Conclusion

This chapter was dedicated to the notion of "mind-independence", as it is involved in Reasons Fundamentalism. Reasons, according to Skorupski's Fregean proposal, are mind-independent entities, to which mind responds spontaneously. Two things can count as established so far: First, in line with chapter 2, there is a *deflationist* reading of this claim, put forth by some neo-Fregean scholars, according to which the mind-independence claim does not involve any strong metaphysical or ontological claim, but only the claim that we cannot derive reasons from any facts about our mind, cognition or practice of reasoning. Second, when it comes to a mind-independence claim in a *metaphysically strong*, non-deflationist sense, we can distinguish between a radical and a moderate version of the claim. According to *Radical Mind-Independence*, reason relations are independent of any particular form of cognition. According to *Moderate Mind-Independence*, reason relations are individually mind-independent, but not collectively mind-independent. Instead, they are dependent on collectively shared *forms of cognition*.

The radical version, I argued, runs into a Darwinian Dilemma as it involves the scientifically questionable assumption of Reason Monolithism. We must either say that there is a perfect end state of cognitive development in evolution, but we can never know whether we have reached it – or that the way our cognitive system works is that perfect end state. Our faculty of reason is

"reason" – the faculty to grasp the universe as it is. As opposed to that view, it is more plausible to understand cognition as a way of making sense of the world and acknowledge the possibility that different cognitive systems can achieve that in different ways. It is undeniable that individual mind must take reasons as they are – that they must be mind-independent in the deflationist sense, but also in a metaphysically stronger sense, given the need for convergence that restricts individual thinking. However, it is far from clear that communities sharing a form of mind must "take reasons as they are".

Finally, I introduced an ontological model of reasons as *potential mental states*. This model is in line with Moderate Mind-Independence and avoids the cut between reasons and mind involved in Radical Mind-Independence. Ultimately, there is nothing in the world but different forms of mind with different representations they can potentially instantiate with the distinctive quality of reason judgements. The reasons which there are, but which we have not yet discovered, are potentials we have not yet actualized. This ontological framing serves as a starting ground for asking what defines and determines these potentials of our mind. *Whatever defines and determines the potentials of our mind*, so the upshot of this chapter, *determines our reasons*. This leads into a material investigation, an investigation of mind as a cognitive phenomenon that evolved in space and time and can be studied empirically, which leaves both the domain of normative questions and the terrain of conceptual analysis behind. It tries to study the normative domain as if from outside – with the eye of somebody detaching from normative judgments and the question about their truth. This endeavour, I will try to show, opens up a striking new perspective on the phenomenon of being a self-determined agent through reasoning – an agent who is intellectually inhabiting a world, a world that she understands, and herself within it.

Chapter 4
The domain of reasons as a collective archive
Leading metaethics into social ontology

The stories others tell about you and the stories you tell about yourself: which ones come closer to the truth?. But actually, that is not the question on my mind. The true question is: In such stories – is there, as a matter of fact, a difference between true and false? Is the soul a place of facts? Or are the alleged facts only the deceptive shadows of our stories?

(Pascal Mercier 2008: 142)

4.1. Introduction

This chapter undertakes a reconfiguration of the domain of reasons as a collective archive of mind. Thereby, it proposes a social ontology of reasons that enriches and complements metaethical Reasons Fundamentalism. If it is true, as this chapter will suggest, that what we grasp when we understand ourselves as grasping a reason is an item in a collective archive, a product of collective tradition, this reveals a striking insight to us: Namely that our reflective understanding in terms of reasons, the capacity we usually associate with mental power, autonomous decision-making and independence from others, is essentially dependent on others with whom we communicate. We could perhaps say that our intellectual independence is essentially a social mode of existence, or a social status.

The assumption on which this chapter builds is the identity of the normative reasons that there are for us and the reason judgements our mind can potentially instantiate as correct reason judgments. This assumption was defended in the previous chapter. While, as agents, we cannot derive the truth of a reason statement from the correctness of the epistemological procedures, we must still

acknowledge that all reasons, which there are for us to discover, are potentials of our epistemic procedures. The procedures of normative epistemology, as introduced in the previous chapter, are a dialectical process between spontaneity as a subject and convergence with other subjects. What does it mean to be a potential of these epistemic procedures? Can we say more about what determines the mental states that we can potentially instantiate as a result of these procedures?

The aim of this chapter is to spell out what the psychological nature of responding to reasons means for the ontology of reasons – for the question of what reasons are. Following the account of *The Domain of Reasons* (2010) by John Skorupski, reasons are conceptualized as fundamental items of thought – fundamental items to which we respond spontaneously in reasoning, while these responses constitute our process of thought. In thinking, we *grasp* reasons that are mind-independent in the sense that we discover them and are forced by them. In the first chapter, I distinguished between analysing reasons in the conceptual and in the material sense. What, now, are the entities, which we conceptualize as fundamental when grasping them, in the material sense? While asking this question has no direct bearing on our first-personal agential understanding of reasons, I suggested in the previous chapter that it can still be asked, thereby contradicting orthodox Fregeans. It can be asked because all there ultimately is in the world are creatures grasping facts *as* reasons. So, we can ask for a description of what we are doing when responding to reasons, thereby viewing mind and reasons from a detached, scientific, empirically interested stance, rather than examining them conceptually or first-personally, with an immediate interest in answering normative questions.

The first step is to pay close attention to the psychological features of grasping reasons as Skorupski describes them in his philosophical account, and then look for empirical approaches to the described phenomena. In section 4.2., I consider several scientifically based approaches to the phenomenon of responding to reasons as a subject. Despite of some controversies in developmental psychology concerning some characteristics of reason responses, there is wide consensus about the social character of the most essential characteristic*: self-determination in light of reason relations*. The next sections build on psychological theories for developing an informative and illustrative picture that corresponds quite neatly to the philosophical picture of "grasping" reasons. The picture of "grasping" evokes two characterisations of reasons, which are here taken at face value. First, the picture evokes an analogy between processes of reflective thought and grasping as a sensorimotor process. Second, the picture of grasping assumes that reasons are, in a sense, mind-independent – out there to be discovered, a

bit like physical objects for which our hands reach out and which we must take as we find them. The reflective mind is determined by reasons, rather than vice versa.

Section 4.3. analyses our activity in the realm of reasons by help of the conceptual tools provided by so-called ecological psychology. Ecology understands organisms as being interlocked with an organism-specific environment, which enables them to respond as effective unities. An organism-specific environment, according to this theoretical framework, is made up by so-called *affordances.* This section suggests a reading of reasons as a specific type of affordance – a type of affordance that is analogous, but also distinguishable from the many other types of affordances to which human organisms can respond. Some of the most recent developments within the neurosciences allow us to reconcile two seemingly contradictory paradigms: According to one paradigm, the distinction between "the physical" and "the mental" or between "the rational" and "the non-rational" is a part of intellectual history, which modern evolutionary biology renders untenable. According to another paradigm, reflecting in terms of reasons is a specific mental faculty that can be sharply distinguished from non-reflective or so-called implicitly reflective processes – a faculty that can only evolve in environments communicating verbally, that is presumably unnecessary for most functions in life, but that may still be of high personal value for people.

Section 4.4. combines the view of reasons as affordances with a genealogical perspective that allows us to describe the realm of reasons as established in a collective archival process and thus warrants the label of a "collective archive of mind". The section describes the structure of a reason in metaethics – a fact standing in a certain relation – in analogy to the structure of a *schema* in psychology – a small body of information that relates persons, actions and situations and is culturally acquired.

Finally, section 4.5. turns back to the ontological question and asks whether reason relations can count as "socially constructed" if the proposed account is true. Following literature in social ontology, I distinguish between the *causal* and the *constitutive* sense of construction. While the reasons we are responsive to might in part be causally socially constructed, I argue that they are in any case constitutively socially constructed. I then turn to Haslanger's ontological proposal of *critical realism* as a fruitful position for understanding the ontology of reasons. The point of defending social constructionism, read as critical realism, is not to deny that something is real, or fundamental to our nature, but to emphasize that it is a collectively created reality, which could be different –

and to examine, furthermore, why this reality is so difficult to change and what such change would involve.[70]

4.2. Responding to a reason: philosophy meets science

According to the philosophical picture developed so far, conscious reflection is a process that is constituted by reason responses. As described by Skorupski, such responses come "directly from the nature of the subject" (2010:405). In a Kantian spirit, he calls this type of response "spontaneous" as opposed to "receptive". Among philosophers of mind with scientific leanings, the idea of subjects responding spontaneously, as opposed to biological organisms responding receptively, may provoke resistance. Others, by contrast, might strengthen that this is a "transcendental" approach which can neither be falsified nor verified by empirical data. I can see a basis for agreement with both sides. On the one hand, I maintain that we cannot really think beyond the limits of our thinking and are therefore entitled to take the indispensable foundations of our thinking as basic. On the other hand, however, I emphasize that there is nothing in the world but biological organisms with the subjective capacity to think. We can thus try to give a scientific account of what we are doing when we employ this cognitive capacity. While no scientific explanation can successfully debunk or eliminate what is most basic for us and all our explanations, looking for such an explanation can still be fruitful and throw an interesting light on our nature.

The process of responding spontaneously as a subject, as described by Skorupski, is centrally characterized by first-personal access to a world – a world as a unified, intelligible system. Human subjects, according to Skorupski, locate themselves within such a system and understand their actions as determined by themselves on the basis of reasons. Self-determination in this sense and responsiveness to reasons are essentially connected. After introducing Skorupski's characterisation of spontaneous responses in subsection 4.2.1., I consider scientific approaches to the idea of unified world-access and the idea of self-determination in light of reasons. The function and origin of the latter idea, discussed in subsections 4.2.3. and 4.2.4., can be investigated with strictly empirical observation. Methods of investigation

[70] I especially thank Ingar Brinck for reading and commenting on this chapter, as well as Olle Blomberg and Eric Brandstedt for encouraging early works that lead up this.

include developmental observations in humans and other species as well as assessments of the neurological basis of reflecting in terms of reasons, for example with fMRI or indirectly with eye-tracking technology.[71] The idea of unified world-access, by contrast, is more difficult to confirm or falsify on the basis of observational data.[72] However, there is a promising model of such an idea, recently developed by neuroscientists and specialists about neuro-imaging, and fruitfully picked up by philosophers of mind whom I would call "scientifically or empirically minded". This means that they take the scientific stance towards mind and cognition as natural phenomena, rather than the stance of first-personal analysis. In 4.2.2., I will introduce this promising account of the brain creating unified world access, called "predictive coding" or "predictive processing".

4.2.1. Spontaneity as a subject: Self-understanding in a unified framework

Spontaneity, as Skorupski describes it, thereby drawing from Kant, is a response that comes directly from the nature of the subject (2010: 405). A purely spontaneous response involves no receptivity by the senses and no interference by alien causes. Whether a response is truly spontaneous cannot be established merely based on a current subjective experience. However, spontaneity is, in the last instance, a first-personally assessed phenomenon. As such, it builds on some sort of introspection – in a broad sense of the term to be specified later.[73]

In this brief subsection, I will outline the characteristics of this phenomenon as described by Skorupski. First, we must reconstruct the relevant understanding of subjectivity or being a self. The probably most central feature

[71] See for example Brinck/Liljenfors 2013a.

[72] The question of neurophysiological evidence for the theory is competently discussed in Walsh/McGovern/Clark/O'Connell 2020.

[73] Some empirically oriented philosophers of mind (for example Eric Schwitzgebel in Hurlburt/Schwitzgebel 2007) challenge this kind of assessment as genuinely unreliable and not yielding much valuable insight into the nature of mind. Another philosophical camp, in contradistinction from that, claims that the idea of spontaneity as it was outlined by Kant relates to the theory of the noumenal self, not the empirical self of psychology or social science (see Wood 1999; 2007). Defendants of this view argue that because of this distinction, neither introspection nor empirical observation can challenge the concept of spontaneity or contribute to its understanding. Skorupski distinguishes himself from this type of Kantian view (see 2010: 406). He emphasizes instead that spontaneity relates to the empirical self.

of Skorupski's notion of a human subject is the capacity of self-determination. Self-determination, as the term is used here, is a cognitive capacity that requires what Skorupski calls, with a Kantian term, "unity of apperception" (2010: 464). Apperception, in Skoruspki's understanding, describes a peculiar way of having access to a mental state. To be precise, access to a mental state is a potentially misleading picture, for, in apperception, there is no difference between the apperceptual state and the apperception of it. "Apperceiving a state is just *being* in that state" (ibid.: 459). Apperception is acquiring knowledge "in the first-person way" (ibid.: 461), while this knowledge acquisition does not involve any form of receptivity or "inner perception" (ibid.). This disqualifies naïve pictures of introspection. Apperception, it must be noted, is a state – the state of mind which we must understand first, in order to examine the question of what makes us as human animals capable of achieving such states. Skorupski describes a list of features that a subject must have in order to be capable of self-determination:

> (i) A self-determiner does not just apperceive this and apperceive that, simultaneously or in succession; it apperceives a synchronically and diachronically unified field.
>
> (ii) Within that apperceptual field it discriminates an ostensibly perceived field of objects, and an ostensibly remembered field of objects.
>
> (iii) It places its perceptual and remembered fields within a single spatio-temporal framework of continuant objects, a framework which both unifies them and extends far beyond them.
>
> (iv) It thinks of—in a sense, experiences—objects in this framework as causally unified in some way.
>
> (v) It locates itself and others within its perceptual and memory field and thus within this spatio-temporal and causally unified framework.
>
> (vi) It identifies itself as self-determining, apperceptual subject with itself as spatio-temporally and causally located object.
>
> (vii) It places its epistemic field within the framework. The epistemic field extends beyond the perceptual and memory fields but is still only a sub-part of the framework.

(viii) It responds to and sometimes reflects on its apperceptual field in the light of epistemic, evaluative, and practical reason relations which it takes to obtain and by which it determines its activity. (2010: 464f)

As exemplified by this list, self-determination requires a complex set of capacities. The items (i-v) occur most likely, to different degrees, in some non-human animals, while the items (vi-viii) might be distinctively human (Skorupski 2010: 465). The picture is so complex that it makes sense to examine at least these two steps separately: the basic unity of apperception characterized by items (i-v), consisting of a certain conception of space and time and causality as forming a unified framework in which one is moving; and what Skorupski calls "a higher, more thoroughly articulated and developed unity of apperception" (ibid.: 465), coming into existence with the capacity to reflect on one's activity in normative terms. I start with the idea of unified world access as it is involved on all levels and then proceed to the more complex idea of reflection in terms of reasons.

4.2.2. A neuroscience-based model of access to an intelligible world

The basic feature of being a subject, according to the philosophical theory presupposed by Skorupski, is having a unified perspective on the world – more precisely, a perspective on the world as a spatiotemporally and causally unified and coherent system. In an instance of apperception, mind understands whatever it understands in that very instance as an element within such a spatiotemporally and causally unified system, or simply within a "world". It is important to note that the world in this sense cannot be perceived as such. The presence of the world in the state of apperception is, as it were, indirect. The apperceiving mind does not apperceive the world, it apperceives a certain element within that world *as* within that world. The "world" – the system of space, time and causality – is not an object among other objects but the background system which enables a phenomenal unity of consciousness. We can understand the phenomenal unity of consciousness as constituting a subject. Accordingly, having a unity of consciousness, and thus being a subject, is inextricably linked to the structures of space, time and causality – to the structures of a world that we can be aware of as unified.

As the "world" is not an object to be perceived in addition to the objects within it, it is challenging to give a naturalistic account of the apperception of an object as belonging to a world. A simple model of our contact with external reality would be a stimulus-response model – a model of a biological organism

that responds to objects in its environment on the basis of received sense data. Since we only receive sense data from particular objects with which our senses are currently standing in contact, the simple stimulus-response model has no room for the indirect presence of a "world" in apperception. Philosophy, as opposed to that, offers a traditional alternative to such a stimulus-response model. This alternative account builds on the assumption of apriorical categories inherent in our thinking. If we imagine these apriorical categories to hold independently of particular brain mechanisms, this assumption raises many Darwinian questions, as I discussed in more detail in chapter 3 – questions of why and how brains should have developed these apriorical structures under the evolutionary pressures that explain their specific ways of functioning.

In distinction from that traditional philosophical account, contemporary neuroscience offers a more naturalistic account of what could be identified as apriorical categories. More naturalistically speaking, apriorical structures are nothing but the structures according to which our brain models an unordered sequence of stimuli as a unified and coherent "world". So-called "predictive coding" or "predictive processing" accounts of human cognition[74], rapidly developing within the recent years, say, in a nutshell, that our perception of the world is not composed of sense-data that travel up our receptive apparatus. Instead, our perception of the world is a construction of the brain that is predicting sensory input by help of Bayesian inference. This means that what we take to be the world is, in the first place, a model built by the brain, while the external world interferes with the internal predictive mechanism only in the form of error signals. This seems to make our relation to the external world in some sense indirect.[75]

The structure of active inference necessarily leads to the construction of a causally unified framework, as Hohwy puts it: "The idea here is simple: you cannot simultaneously use two or more different, competing causal hypotheses

[74] Predictive coding theories, in their basic outline dating back to the physician and physicist Helmholtz (1925, originally 1860), are formulated more recently by for example Karl Friston (2009) and Chris Frith (2007), both neuroscientists and leading specialists about brain imaging, and philosophers of mind such as Andy Clark (2016) and Jakob Hohwy (2013). The interest in the approach is currently growing (see e.g. Mendonça/Curado/Gouveia 2020).

[75] Some philosophers see the Kantian theory of aprioricity as a forefather of the idea of predictive coding (e.g. Swanson 2016). As opposed to simple mind-to-world models of perception, Kant, in his theory of the *a priori*, asks "whether we shall not make better progress on the problems of metaphysics if we assume that objects must conform to our cognition'" (Kant 1999/1787, sec. B xvi), rather than vice versa. Kant himself labelled this reversal of the relation between mind and world a "Copernican Revolution" in the understanding of cognition. According to that view, we have no access to the world as it is in itself.

as a basis for sampling the world to minimize prediction error" (2013: 216). So, the brain decides for one hypothesis and, from there on, goes on to spin a model of the world, always deciding for the hypothesis that minimizes prediction error. The resulting model is therefore unified. This unified model, however, does not necessarily depict causal structures in the world – at least not directly. In this sense, the unified world of which we are conscious is a mental construction.

If this type of account is appropriate, it might have important implications for how we conceive of the mind-world relation. Predictive processing theories give prima facie support to the idea that perception is, as has been formulated, "virtual reality" or "controlled hallucination" (Hohwy 2013: 137; Clark 2016: 169) rather than a mind-to-world response understandable according to a simple stimulus-response model. Philosopher of mind Andy Clark, another influential contemporary proponent of predictive processing, cautions against the idea that "the probabilistic inference engine in the brain" is hallucination in the sense that it constitutes "a barrier between agent and world" (2016: 171). He emphasizes, instead, that predictive processing, rather than preclosing us from the world, "delivers a genuine form—perhaps the only genuine form that is naturally possible —of 'openness to the world'" (ibid.: 195). Clark puts the focus on action – on the fact that we are "active world-engaging systems" (ibid.: 188). His account of the predictive brain depicts action as the essential link between the predictive modelling of our brain and the phenomenon of a unified field of consciousness. It "makes adaptive sense" for us to model the world in an unambiguous and unified way. Understanding ourselves as moving in a unified world gives us a "grip upon a structured, organism-salient, external world" (ibid.: 202).

Clark agrees with others that the phenomenon of unity is produced by the mechanisms of our brain but holds in addition that "it is only because the world we encounter must be parsed for action and intervention that we encounter, in experience, a relatively unambiguous, determinate world" (ibid.: 188). It is because cognition is "forged in the presence of, and dedicated to the service of, world-engaging action" (ibid.: 202) that it continuously models the world as unified.[76] Cognition, according to his view, serves the purpose of giving us

[76] With that emphasis, Clark (2016.: 289) places himself in the tradition of what is known as enactivist cognitive science (see for example Varela/Thompson/Rosch 1991). The enactivist approach can be understood as a branch of ecological psychology that will be discussed in the next section. Many authors explicitly combine the ecological and the enactivist tradition, for example Weichold (2018), Ryan and Gallagher (2020) or Heras-Escribano (forthcoming). Others, as Read and Szokolzsky (2020), by contrast, argue that the two traditions share the same over-arching goal, to propose a non-representationalist account of cognition, while they go distinct paths, explaining cognition by perception (ecology) vs.

a grip on the world. “In active animals, that grip is not rooted in some kind of action-neutral image of an objective external reality” (Clark 2016: 202). Instead, the outlook on the world is adaptively tailored for action. Clark uses the formulation “gripping tales” (ibid.) for the mechanisms that our brain develops for accessing the world. Going back to Skorupski’s list, we could say that our orientation in space, our temporal ordering and our understanding of causal chains is, first and foremost, only the most functional way to order the multitude of sense data in a way that facilitates action. It does of course not follow from this that the world as we perceive it is mere fantasy. However, we must acknowledge that it is an active constructive achievement of mind – dependent on our needs as “world-engaging systems”, which shape this specific form of mind.

It is unclear, and not of relevance here, to which extent the categories of space, time and causality are innate and genetically determined. Most likely, some non-human animals share these basic categories and are thus basic subjects in Skorupski’s sense as well. Skorupski, however, distinguishes between features of a more basic form of the unity of apperception and the more developed unity of apperception necessary for self-determination. In the next subsection, I will suggest that, while it is unclear whether the basic features are innate and universal, at least this cognitive upgrade requires more specific environmental conditions. In addition to understanding themselves within a spatiotemporally and causally unified framework, self-determiners understand themselves as agents guiding their activity in light of a unified normative framework that holds independently of them. As we can learn from developmental and evolutionary psychology, communication and sociality are central for this cognitive capacity.

4.2.3. Self-determination in light of reason relations: developmental questions

In addition to locating oneself in a spatiotemporally and causally unified framework, self-determination requires more sophisticated cognitive capacities, according to Skorupski’s list. Skorupski states that “acting for a reason and responding to a unified perceptual field develop together” (2010: 465). As suggested in the previous subsection, an account of predictive processing gives a good scientific explanation of the basic features of

explaining cognition by sensorimotor contingencies (enactivism). I agree with Clark and others that perceptions and sensorimotor potentials are interlocked.

Skorupski's unity of apperception. In a very basic sense, all biological organisms[77] have a unified model of the external world, as described by these theories. In a subject capable of self-determination in light of reason relations, however, this unified model develops further. The self-determining subject is not only conscious of a unified perceptual field, and, as Skorupski sees it, thereby capable of acting for reasons. The self-determining subject is, additionally, conscious of itself as acting for reasons.[78] Skorupski describes the stage of self-determination as a stage in which a subject comes to understand itself as having a particular epistemic field, which is located within the unified framework, but "is still only a sub-part of the framework" (ibid.: 464). Such a subject can finally understand itself as responding to and reflecting on reasons by which it determines its activity. In addition to simply having a unified perspective on the world, self-determination involves an awareness of a world beyond this perspective (item vii on Skorupski's list). Finally, the most crucial characteristic of self-determiners, according to Skorupski (2010: 465 viii), is a sense of deciding for and against particular actions based on reason relations one believes to obtain.

Within developmental psychology, there is some disagreement regarding the first of these faculties – the acknowledgment that there is a world beyond one's own perspective, a faculty called "object permanence" by Piaget (1954). According to Piaget's school, object permanence is a basic sensorimotor faculty that is pre-social, while a tradition established by developmental psychologist Vygotsky (1962; 1978) holds that even these basic processes are

[77] There are controversies about how inclusive this picture should be. At the most extreme end of the spectrum, there are positions that extend the basic approach to each form of cellular organisation, in principle also to unicellular organisms. As an example, we might consider Varela's (1988) proposal of bringing forth meaning through operational closure in bacteria. I thank Martin Weichold for introducing me to this work.

[78] It is Skorupski's picture that a basic unity of apperception allows for acting for reasons, while a more developed unity of apperception allows for self-consciousness as acting for reasons. The following subsection rejects the picture that Skorupski entertains here – a picture that seems to imply the possibility of not fully reflective or implicit reasons, of which we can become reflectively conscious. The revision of this picture also relativizes the classification of self-determination as an "upgrade". While self-determination in light of reason relations is indeed an enrichment of cognitive capacities, it works in analogy to other cognitive capacities, not as a direct development of those. However, the term upgrade of cognitive capacities is innocent if we just understand it as an addition or augmentation of complexity, not as an improvement on a linear scale.

determined by a communicative social space.[79] Brinck and Liljenfors (2013a) have more recently defended the Vygotskyan approach in that respect.[80]

Much less disagreement can be found about the fact that the second part of Skorupski's discription – the sense of deciding on the basis of reasons – is essentially social, a result of the necessity to locate and justify oneself in a social world. We need not take a stance in the dispute about the relevance of the social context for the basic sensorimotor development, involving "object permanence", at this point. What we can, nevertheless, state at this point is that both of the described camps in developmental psychology, as well as a striking number of contemporary cognitive scientists (Tomasello 2001; Boyd 2018; Sperber/Mercier 2017) and evolutionary biologists (Laland 2017; Dunbar 2000; 2016), agree that the social context is essential for the development of the full-fledged capacity of reasoning – of reflecting in normative terms.[81]

---

[79] To make a short excurse into studies of the first faculty, we can start with Piaget who refers to this cognitive faculty as "object permanence" (1954). This faculty can be attributed when a child developed an awareness that objects persist even when it is not looking at them. By "object permanence", Piaget marks the completion of what he calls the sensorimotor stage of cognitive development. Piaget distinguishes the sensorimotor stage, among others, from the formal operational stage of development, which is completed during adolescence and results in the faculty to weigh, compare and endorse life plans and long-term goals that stand in a wider personal and societal context. Some psychologists disagree with Piaget concerning the separability of the sensorimotor stage and the formal operational stage. Vygotsky, for example, claims that both sensorimotor development and the development of reflective assessment and planning take place within a societal and cultural context (Vygotsky 1962; 1978; see also Wertsch 1985). This context, according to Vygotsky, has crucial implications for the direction of development. The context, to which the child is exposed in the form of reactions by other people, determines which actions are conceived as doable. The reactions of other people also determine where to draw the borderline between what is attributable to an agent and what is attributed to external forces. According to Vygotsky, speech, at first only used to address an adult, is gradually "turned inward" (1978:27) and accompanies behaviour intrapersonally. While it first only accompanies behaviour, Vygotsky claims that speech later comes to "precede" action, so that it takes up a "planning function" (ibid.: 28). Thereby, Vygotsky says, "children acquire an independence with respect to their concrete surroundings; they cease to act in the immediately given and evident space" (ibid.).

[80] They, however, depart from Vygotsky in rejecting his picture of "internalization" in favour of a situated approach to cognition. My ecological proposal in the next section is in line with their idea.

[81] While developmental psychology in the tradition of Vygotsky emphasizes the role of verbal communication in shaping action possibilities, more recent research denies the relevance of what we communicate verbally about our actions for the behavioural outcome we produce. A general suspicion against reflectively understood and verbally communicated reasons for action has become widespread in scientifically minded communities. Reasons for action, according to a famous study, are usually nothing but confabulations – not descriptions of what really moves us to action, but rationalizations after we have acted (Johansson *et al* 2005). Various other studies show that what really moves us to action are unconscious

Whatever applies regarding the earlier stages of cognitive development, we can understand the final stage, reflecting about oneself in terms of reasons, as essentially mediated by communicative contexts. The ground-breaking works in developmental psychology by Inhelder and Piaget (1958) link the completion of the formal operational stage, the stage in which full-fledged reflection on abstract matters develops, to a sense of locating and determining oneself in the social world (Inhelder/Piaget 1958: 334-350). Elkind (1967), elaborating on this, coined the terms "imaginary audience" and "personal fable" for phenomena observable during adolescence. Adolescents, according to these observations, typically reflect on their behaviour in front of an "imaginary audience", consisting of a specific social reference group, which is supposed to approve of everything they do. A "personal fable" gives one's actions and plans significance on a large scale. During adolescence, imaginary audience and personal fable are explicit and especially pronounced. In the course of cognitive maturing, they move into the background. However, being capable of these representations is necessary for developing a full-fledged sense that one's actions matter or, put otherwise, that normative considerations concerning one's actions matter and that these normative considerations are valid independently of oneself.

Other psychologists describe the beginning of normative reflection in much earlier stages than adolescence (e.g. Rakoczy *et al* 2008).[82] However, details about child development do not matter here. What matters is the widely established fact that there is a close link between reflecting about oneself in light of reason relations on the one hand and being a member of a community, which communicates and assesses reasons, on the other hand. The completion

mechanisms and biases. The reasons we cite when we reflect on reasons often have little to nothing to do with our actions (see Kahnemann 2011; Haidt 2012).In line with this spirit, Sperber and Mercier (2017) have tried to take reflection in terms of reasons seriously, nevertheless. They propose that the whole activity of reflecting in terms of reasons is not done in order to facilitate and guide action, but for the purpose of justification – justification to oneself and to others. Thinking in terms of reasons, according to this proposal, is nothing but a communicative practice for the purpose of justification – a practice that serves social cooperation. Mercier and Sperber conceptualize the faculty of "reason" as a module, a compartmentalized area within the cognitive system. Modules, they define, "should be viewed as components of larger systems to which they make a distinct contribution" (2017: 73). The distinct function of "reason" within the cognitive system of human animals is, as the two authors suggest, not "to help individuals achieve greater knowledge and make better decisions on their own". Instead, "reasons are commonly used in the pursuit of social interaction goals, in particular to justify oneself and to convince others." (ibid.:175).

[82]A possibility to deal with this disagreement is to distinguish between "moral intuitions" developing early on (see also Turiel 2006), and normative reflection in terms of reasons, developing during adolescence.

of what Piaget calls the formal operational stage can be seen as corresponding to the acquisition of a full-fledged unity of apperception according to Skorupski – satisfying all the eight criteria listed by him, including the capacity to reflect about oneself in light of reason relations that obtain independently and according to which the agent determines her own actions. This and only this stage shall be of relevance in the following. This stage shall be referred to as conscious reflection, or simply reasoning. Although the examination must remain superficial and sketchy at this point, and many of Skorupski's eight criteria are explained in different ways by different researchers, the various cognitive scientists, evolutionary biologists and developmental psychologists mentioned above agree on one basic claim: That the empirical basis of self-determination in terms of reason relations is a genuinely social practice of giving and accepting reasons. While a great variety of intelligent responses to the environment may be independent of social practices, the specific response of reflecting in terms of reasons is dependent on a social environment.

When responding reflectively to a reason as a reason we do not only locate ourselves in a spatiotemporally and causally unified field, we locate ourselves in a social space – in a social space within which certain communicable facts are accepted as justifications, for certain actions in certain contexts and by certain people, and others are not. The brain, according to predictive coding, models the world – and likewise the social space in which we mentally locate ourselves – as unified. This does not mean that it is impossible to believe or expect that particular people will not accept the reasons which oneself takes as justified. However, it means that for believing that oneself is really justified, it is necessary to be able to imagine a community that shares one's judgements, and that is so unified and coherent that the practice is workable and predictable. Private reasoning, accordingly, is derivative on the social practice. "Derivative" here means that the private process rests on the predetermined paths of the social practice. This does not necessarily mean that the social practice determines the private process exhaustively. But it means nevertheless that private thought is dependent on the social practice and cannot autonomously set that practice aside.[83] In the following, I will present a recent proposal by two cognitive scientists, Dan Sperber with a background in cognitive anthropology, and Hugo Mercier with a background in psychology, who portray the faculty of reason as having evolved in a social practice of justification. Moreover, their picture can be neatly mapped on the picture developed by Reasons Fundamentalists in philosophy.

[83] See also the reply by Brinck and Liljenfors (2013b) to Markova and Legerstee.

4.2.4. Justification practice: The evolution of reflecting in terms of reasons

Within the last years, psychologists, cognitive scientists and evolutionary biologists have examined the question of how seemingly "higher" cognitive faculties, such as reflective reasoning, have evolved in the human species.[84] In addition, research has contributed to the emergence of scepticism or even debunking of reflective reasoning by showing that, all things considered, it does not improve our behaviour or gives us a more objective picture of the world (Kahnemann 2011). On the contrary, it mainly serves as an "after the fact" rationalization (see Johansson *et al* 2005). Mercier and Sperber, in their new theory of explicit reasoning, defend the claim that reasoning is, in the first place, a social justification practice.[85] The two authors state the following:

> Reason, we argue, is a mechanism of intuitive inferences about reasons in which logics plays at best a marginal role. Humans use reasons to justify themselves and to convince others, two activities that play an essential role in their cooperation and communication. (2017:107)

The two researchers thereby endorse two distinctive claims. The first claim says that the basic building blocks of reasoning are "reasons", fundamental propositional items on which mind works, rather than logical rules or structural requirements, according to which propositional content is processed. Mercier and Sperber: "There is no general method that we could or should follow when reasoning, either on our own or in dialogue with others." (2017: 174).

The second claim says that these propositional items are items that are exchanged in communication and fulfil a justificatory role. "Reasons", the two authors formulate, "are for social consumption." (2017: 123) The authors speak of communities "producing" reasons. Nevertheless, they concede that a community cannot "build a battery of reasons all of its own" (ibid.: 143). What can function as a reason and what not is determined by "cognitive efficiency" (ibid.: 144) and thereby largely by external constraints and constraints of the species. However, the fact that something that has the potential to function as a reason does indeed function as a reason, i.e. is acceptable as a reason by mind, depends on a community that accepts a particular item as currency in the process of justification. Solitary reflection, on this account, is nothing but a

[84] See e.g. Mercier/Sperber 2017; Laland 2017; Boyd 2018; Dunbar 2016.

[85] I thank Bartłomiej Czajka for pointing me to the work by Mercier and Sperber and for a very rewarding discussion.

mirror of the social practice of justification – driven by the need to justify oneself.

Mercier and Sperber relate their terminology explicitly to philosophical literature on reasons as facts. They define: "Facts, as we are using the term, are true propositions, abstract objects without causal powers." (2017: 111) The first of the described claims can be compared to Scanlon's claim that the idea of a reason must be presupposed as primitive, since there is hardly "any general account" of what makes a reason a reason and of how we arrive at a reason judgment (1998: 17). This is one of the core ideas of Reasons Fundamentalism as discussed in chapters 1 and 2 of this book. We can treat the philosophers Scanlon and Skorupski on the one hand and the cognitive scientists Mercier and Sperber on the other hand as agreeing on the fact that reasoning is primarily a matter of responding to reasons. However, when it comes to the second claim of the empirical researchers, we must treat Scanlon and Skorupski as indifferent towards the issue in question, namely how and why reflecting in terms of reasons has evolved. Scanlon states that he "presuppose[s]that [his] readers are rational" (1998: 17). None of the mentioned philosophers examines the question why we can presuppose that one's fellow humans are responsive to reasons and which evolutionary forces shaped this responsiveness. A primary motivation of this book, defended in the first chapter, is to show that the empirical and genealogical perspective enriches our self-understanding as reflective creatures and contributes to a comprehensive analysis of what reasons are.

The story presented by Mercier and Sperber is especially convincing because it picks up one of the most widespread worries which the current age of neuroscience and behavioural psychology harbours against traditional philosophy. The philosophical emphasis on humans as reflective creatures in contrast to other species that are traditionally envisioned as non-reflective and instinct-driven is more and more under attack. Mercier and Sperber build on exactly these worries but succeed in making sense of the capacity of reasoning as a cognitive faculty in its own right. We can regard this strategy as distinct from debunking strategies – strategies that refuse to consider humans as reflective creatures altogether or at least deny reflection an important role. According to the picture of reasoning as a justificatory practice of communication, reasoning makes an essential contribution to our form of life. Nevertheless, we are invited to gain an appropriate understanding of the origin and role of reasoning in order to appreciate it in the right way.

According to Mercier and Sperber, reason is misunderstood, if we understand it as a "global" faculty – i.e. a faculty that potentially comprises all areas of cognition and can rationally criticise, evaluate and improve all our

cognitive capacities. As opposed to that picture, Mercier and Sperber propose to understand reason as a particular "module" – and, moreover, as one module among others. Mercier and Sperber define modules broadly as "autonomous mechanisms with a history, a function, and procedures appropriate to this function. They should be viewed as components of larger systems to which they each make a distinct contribution." (2017: 73).[86]

The authors view reason as a module whose function it is to produce reasons for justification. This view also questions our view on mind in a fundamental way, at least our view on mind understood as reflective consciousness. If reason is a functionally distinct mechanism in the brain, rather than a global faculty, we might see ourselves forced to give up a popular idea. People, at least since the times in which psychoanalysis in the Freudian style became popular, often imagine that they have a lot of cognitive and emotional states of which they are unaware, while only a few become fully conscious. Between the states of which we are unaware and the states of which we are fully conscious lies a huge continuum on which we can move – depending on our motivation to reveal or to deny states of consciousness.

This picture is seriously challenged by the recent wave of research on reflective consciousness of which Mercier and Sperber are part. The two authors doubt that there is explicit and implicit reasoning in the sense that the former is the conscious version of the latter (2017: 118). Explicit reasoning, according to their picture, is just what it is. The processes of which we are not conscious are entirely different mechanisms, not just the unconscious counterpart of reflective consciousness which we have not accessed. Several neuroscientists confirm this picture of mind. Chater, for example, argues that there is no such thing as unconscious or subconscious mental states (2018: 160). When we take ourselves to realize something new about our mental attitudes, such as our feelings, beliefs or desires, we in fact acquire a new way of interpreting given conditions, according to this view. We are "reinventing ourselves" (2018:220) rather than gaining access to a state of which we have hitherto been unconscious. Another contrasting view to the understanding of reason as a module is the distinction between system 1 and system 2 processing, as defended by Kahnemann (2011). According to this so-called dual processing theory, system 1 responses are intuitive, fast and spontaneous,

[86] The idea of the modularized mind, originally suggested by Fodor (1983), has continuously been developed within neuroscience (see Schlosser/Wagner 2004). Contemporary accounts of neuronal modules have departed from the original idea in that they no longer see a module as being located in a particular area of the brain. Complex modules can have neuronal correlates that span over various areas of the brain. What makes it nevertheless appropriate to speak of distinct modules is their relative functional autonomy and distinctiveness.

while system 2 works deliberatively, slowly and rationally. Mercier and Sperber's investigation, by contrast, suggests that what Kahnemann calls system 2 is as intuitive and spontaneous as any other response, just that it fulfils a different function and is done in different situations – when having the time and opportunity to think about justification for example.

This modular view on reason is further confirmed by the neuroscientists Richard Brown and Joseph LeDoux (2017) who locate something in the brain that might be identifiable as a reason module. More precisely, they develop a "higher-order theory" of what they call phenomenal consciousness. The phenomenon of having reflective access to phenomenal states of consciousness – states we can introspect and report verbally – is a matter of higher-order representations in the cortex, according to their view. When we have a conscious representation, such as "I am seeing red" or "I feel fear", this is due to the activation of particular neuronal circuits in the cortical areas of our brain (2017: E2017). Typically, the activation of these neuronal correlates of conscious experience is coupled to many other functional processes. In the case of fear, for example, the conscious state "I feel fear" is coupled to particular responses in the amygdala, and coupled to this is a variety of physiological responses, such as sweating, trembling, increased heart rate etc. In the case of representations such as "I am seeing red", there is usually activity in the visual cortex and eventually also in areas governing executive or motor functions related to seeing red under specific conditions.[87]

However, clinical research on pathological ways of processing reveals that the different systems can very well work independently of each other. The most convincing example are probably so-called blind-sight patients (Brown/LeDoux 2017: E2017). They can "see" red in the sense that their brain registers the colour and produces various responses that indicate "seeing" it. Nonetheless, the patients report not having any introspective content or consciousness of "red". Similarly, it is possible in the case of fear that there is a clear threat response in the amygdala, also effecting the physiological responses, without producing a conscious state of fear, or even producing a totally different conscious response.

Brown and LeDoux identify links between what they call first-order states, such as for example responses of the amygdala, on the one hand, and higher-order states being the correlates of conscious experience on the other hand. The important point is that the higher-order state correlating with the conscious experience is not the consciousness *of* the (non-conscious) first-order state. Instead, it is a completely distinct type of state that can potentially be attached

[87] I thank Manuel Oliva for a helpful exchange about these matters.

to certain kinds of first-order states. However, there is no in itself correct or incorrect way of attaching partly independent processes to each other. Which higher-order states an individual brain activates in particular situations and conditions is, according to Brown and LeDoux, largely a matter of encultured learning, upbringing and especially linguistic communication (2017: E2022).[88] In other words, so-called higher-order states have a distinctive function in the brain, mostly limited to linguistic communication within social relations. This makes the higher-order states, or simply the conscious mental states, functionally unnecessary for most activities. Instead, we can see them as, in the first place, designed for getting along in a communicative context. This is exactly Mercier and Sperber's picture of reason as a module, rather than a global faculty. We can thus identify the locus of what Brown and LeDoux call "phenomenal consciousness" with the reason module as described by Sperber and Mercier. In fact, the term "reason module" might be a more suitable expression for what is reflectively accessible and verbally expressible than the term "phenomenal consciousness" that some might define broader.[89]

The theories by neuroscientists such as LeDoux, Brown or Chater, backing up Mercier's and Sperber's evolutionary story about reason, thus reject a paradigm in the study of consciousness. According to this paradigm, exemplarily defended by Block (2011), so-called "access consciousness" (Block 1995) is only a restricted part of consciousness. According to the rival theory, there is no such distinction. Our consciousness is simply what we are conscious of – what is "on our mind". The phenomenon of having something "on one's mind" in this sense, is accordingly a very narrow and restricted phenomenon. To those who feel an unease with restricting the term "consciousness" to what we explicitly reflect on, I suggest leaving room for extending terms like "consciousness", "awareness" or "subjectivity" to a broader range of phenomena. This possibility notwithstanding, neuroscience gives us good reason to treat the phenomenon of "conscious reflection" as distinct from possible other phenomena falling under these broad headings. Explicit reflection is not just the conscious version of non-conscious or implicit "reasoning". The relation between reflective states and non-reflective processes is ill described if we describe it as a matter of degree, as the two ends

[88] For a better understanding of the view, it may be helpful to know that LeDoux suggests missing or unfavourable links between reflective consciousness and other brain mechanisms as an important cause for many mental disorders, such as anxiety disorder (LeDoux 2015). The judgment that a particular link is "unfavourable" is, important to note, a personal judgment, not a functional defect of the brain in the strict sense.

[89] I thank Martin Weichold for a critical discussion.

of a continuum. Instead, contemporary neuroscience allows us to describe the human brain in a much more fine-grained way. The picture of the modularized brain allows us to distinguish between a variety of qualitatively and functionally different subsystems of processing, which can be identified as partly autonomous and should therefore be treated as distinct. If we constrain our topic of enquiry to reflective reasoning and call "mind" the occurrence of such processes of reasoning, in distinction from other potential objects of consciousness studies, we can indeed say, following Chater (2018), that mind is not a complex and "dark" entity within us. Instead, mind in this sense is only the specific function of brains to instantiate states of reflective consciousness. Instantiating states of reflective consciousness is, according to Mercier and Sperber, something which individuals do as members of a community in which justifications are demanded and accepted.

There is, to conclude, a relation between being able to instantiate a reason judgment as correct and being able to use the communication of this judgment for justification in a social context. Mapping the philosophical picture by Scanlon and Skorupski on the scientific picture by Mercier and Sperber, we can say that the fundamental items of our reflective thought are the same entities, differently described, as those items functioning in a social justification practice.[90] At the end of the previous chapter stood the idea that all normative reasons there are for us can be modelled as mental states we have a potential to instantiate in a particular way. Now, the theory by Mercier and Sperber gives us an understanding of what defines and determines these potentials: their function and success in a social practice of justification.

It is too early to jump to the conclusion that either reasons, or the normative force of reasons, is a social construct. Moreover, if we develop this conclusion, we must specify the term "social constructionism" and argue why it is suitable to describe the ontology of the fundamental items of our thought. Before I get to this question in section 4.5., I must elaborate on the "mapping" of the two respective pictures on each other. Two central questions must be answered. First, we must bring together the idea that reasoning is justification practice and that there is a unity of apperception, providing access to a world as an

[90] There is a difference between really giving reasons and merely emotionally or demagogically persuading (I thank Johan Brännmark for raising this objection). I think this suggests that not everything that persuades is a reason. But we can still hold, vice versa, that everything that is a reason must have the potential to persuade. Maybe, we could describe the difference between giving reasons and other forms of persuasion as the former being more sustainable and easier to transfer through longer time-spans and to different situations. Thus, the definition of reasons as being something that can be accepted in a social justification practice is not touched by that objection.

intelligible, coherent system, in every instance of reasoning. Second, we must reconcile the fact that communication practice evolved over time, while we take reason relations to hold mind-independently.

In section 4.3., I propose an original way of understanding communication practice as giving rise to an environment in which we can grasp certain facts as reasons and thereby actively locate ourselves in that environment. I will arrive at the idea of a realm of reasons as a subset of the natural environment that human creatures inhabit. In section 4.4., I develop a genealogical picture of reason relations by comparing them to schemata in so-called "schema theory". Thereby, I will argue that the social realm of reasons can be understood as the product of a collective archival process – in other words, as a collective archive of mind.

4.3. An active grasp: reasons as affordances of thought

The previous section suggested that the fundamental items of our thought, reasons as described by Scanlon and Skorupski, can be seen as the product of a social practice of justification, when we study them from a scientific perspective. This section aims to show that we can understand those items as parts of the furniture of our environment – independent of each of us, though not determined by the physical world. Responding to these items can be understood as a process of "grasping" that is perfectly analogous to the sensorimotor process of grasping as a physical process. The tradition of so-called "ecological psychology", to which Clark's predictive coding theory stands in close neighbourhood, provides conceptual tools for understanding the subject's grasp of reasons as a natural process, while preserving the clear neurological distinctness of reflection in terms of reasons and other processes. Ecological psychology describes organisms in relation to their organism-specific environments, constituted by sets of so-called affordances to which they can respond. The conceptual repertoire of ecological psychology, I think, helps us to tie the analysis of reasons back to the Fregean idea of mind-independence discussed in chapter 2. It allows us to adopt such an idea while acknowledging two facts that were established so far. First, that all reasons holding for us are potentials of our mind. Reasons can be described as propositional attitudes we can potentially instantiate. Second, that the potential of our mind to instantiate reason judgments is the product or the mirror of a social justification practice.

In this section, I defend the claim that reasons can be described as a specific type of what ecological psychology calls "affordance". The reasons holding for us are potentials opened up to our mind by a socially cultivated environment. This might seem striking at first, because it has become common to use the concept of an affordance, first employed by Gibson (1966; 1979), to play down or deny the relevance of what some call "higher-order" cognitive functions (Reed 1996; Heft 2001; Chemero 2003; 2009; Rietveld 2008).

That notwithstanding, the philosophers of mind Rietveld and Kiverstein (2014) suggest that all sorts of abilities can be understood as "skilled engagement with affordances" (ibid.: 343) – not only unreflective motor-abilities but also reflective judgments and "practices of giving and asking for reasons" (ibid.). This initial proposal by Rietveld and Kiverstein is what I would like to spell out in more detail here. As opposed to what most ecological psychologists acknowledge, communities of reasoners share a realm of reasons that can be conceived as separate from the environment of motor-action though it is in the same physical world. The idea of a distinctively mental realm as well as the idea of distinctively mental states usually provokes resistance by ecological psychologists (e.g. Reed 1996: 177). It seems that the idea of "the mental" is regarded as opposing the most basic claim of ecological psychology, which is that there is no mind other than the organism that moves as a functional unity within its environment. I think we can defend an idea of a mental realm compatible with this claim when we build on the previously introduced concept of a cognitive module – a functionally distinct type of ability with relative autonomy within the cognitive system. This concept allows us to conceive of "the mental", not as distinct from "the physical", but as a socially created part of the physical world. Subsection 4.3.1. presents the basic concept of an affordance according to ecological psychology. In subsection 4.3.2., I introduce a class of affordances that I call affordances of (reflective) thought. Thereby, I suggest an understanding of the realm of reasons as a distinctive subpart of the ecological niche of humans.[91]

4.3.1. The concept of an affordance in ecological psychology

The account of reasoning as a social justification practice, which was presented in section 4.2., depicts reasoning as a functionally distinct set of cognitive capacities – a module, as Mercier and Sperber call it. The capacity of responding to reasons is functionally distinct in one sense. In another sense,

[91] I am especially indebted to Martin Weichold for introducing me to theories of affordances (see also Weichold 2015; 2018).

however, it is not "special", not "global", not "universal", not "more controlled", but completely on a par with other responses of which human organisms are capable. Likewise, the scientific framework of so-called ecological psychology refuses the distinction between the mental and the behavioural. Edward Reed, a proponent of the view, explains: "From the ecological point of view, in which knowing is not separated from living, cognition might best be defined as an animal's capacity to keep in touch with its surroundings." (1996: 169)

This subsection gives an outline of the basic terms of ecological psychology as they have been originally developed by Gibson (1979) and further modified by others, such as Reed (1996), Chemero (2003; 2009) and Rietveld/Kiverstein (2014). Two central ideas must be understood: the conceptual distinction between the physical world and the environment of an active and mobile organism; and the idea of co-constitution of organism and environment. These ideas will be used in order to account for reasons being mind-independent parts of our environment though not being part of the physical world; and for reason responses constituting our reflective mind, for reason responses being processes of grasping that relate us to a meaningful world.

Gibson first introduces the distinction between the world as physical reality and the world as an "environment". The term "environment", Gibson says, refers "to the surroundings of those organisms that perceive and behave" (1979: 3). This means that there was no environment, only a potential environment, before the development of life (ibid.: 4). Without life, the world is only the reality of physics and geology (ibid.). Only in relation to life, so Gibson's idea, we can speak of an environment. In relation to life, it becomes not only possible, but also necessary to conceptualize the surroundings as an environment, because an animal is "a perceiver of the environment and a behaver in the environment" (ibid.). It is inappropriate to say that it "perceives the world of physics and behaves in the space and time of physics" (ibid.).

While physics speaks about points in space, ecology speaks about "places" that structure the organism-specific activity (ibid.: 29). While physics speaks about points in time, ecology speaks about "events", "processes" and "sequences" that govern subjective time perception (ibid.: 6-8). The concept of environment is tied to the activities typical for a specific organism and, vice versa, the organism is defined via the specific possibilities that an environment offers to it. Gibson calls this phenomenon the "mutuality" or "complementarity" of animal and environment (ibid.: 4). This idea amounts to an ontological proposal with which Gibson wants to overcome the "philosophical dichotomy" between "objective" and "subjective" (ibid.: 35).

In the centre of this alternative ontology is the concept of an "affordance". Gibson says the following about it:

> The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill. The verb to afford is found in the dictionary, the noun affordance is not. I have made it up. I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment. (ibid.: 119)

Gibson points out that, for a living organism, the world does not consist of geological or physical objects. Instead, it is meaningfully furnished with objects for daily routine. These objects directly present themselves to the living being as calling for particular responses. Gibson quotes the Gestalt psychologist Koffka (1935): "Each thing says what it is… fruit says ´eat me´, water says ´drink me´, thunder says ´fear me´ and woman says ´love me´" (Gibson 1979: 129).

As opposed to Gestalt psychology, Gibson, however, does not want to understand these "callings" of objects merely as phenomenological features of subjective experience (ibid.: 130). On the contrary, objects with "Aufforderungscharakter" (Levin), so-called "affordances", are real constituents of the world – of the world understood as environment, in distinction from the world understood in the sense of physics and geology. As such, Gibson emphasizes, affordances do not fall within the dichotomy of objective and subjective, they are "both if you like" (1979: 121).

In the decades after Gibson's seminal work for ecological psychology, the concept of an affordance has been interpreted in different ways. While Reed (1996) interprets it as a property of the environment, a property that constitutes a resource for the organism, Chemero (2003) proposes to understand affordances as relations – relations between features of the environment and abilities of the organism. Rietveld and Kiverstein pick up both interpretations and formulate a definition of affordances as "relations between aspects of a material environment and abilities available in a form of life" (2014: 335), providing the individual organism with a resource it is invited, but not mechanically necessitated to use (ibid.: 327).

With this definition, Rietveld and Kiverstein introduce two novelties. First, as opposed to Chemero (and Gibson), they do not speak about features of the environment but about "aspects of the environment". By introducing this term, they want to emphasize that "in the human case the material environment has been sculpted by our sociocultural practices into a sociomaterial environment" (2014: 335). Speaking about aspects rather than features captures the fact that

"in the process of education of attention the novice learns to selectively pick up some aspects of the environment while ignoring others." (ibid.) While a human face, for example, can be a *feature* in the environment, it might reveal different *aspects* to different onlookers. One onlooker might mainly perceive the aspect of a nice or slightly insecure smile, while another onlooker perceives nothing but the aspect of a darker pigmentation typical for sub-Saharan Africans. The distinction between features and aspects will be of relevance later.

Second, as opposed to Chemero (and Gibson), they relate the aspects of the environment to a form of life, rather than a particular individual organism. Thereby, Rietveld and Kiverstein make room for normativity and a distinction between affordances and solicitations – affordances that actually motivate behaviour, in distinction from affordances that merely invite it (ibid.: 341f). Due to the fact that each single living being is part of a collective form of life, it is subject to norms of appropriateness of which it can fall short (ibid.: 332f). Kiverstein coins the term "situated normativity" (2008). Each norm of appropriateness, according to that picture, relates to more or less specific situations and places.

Ecology works with the concept of a "niche", in distinction from the concept of a "habitat". The habitat of an animal, according to Gibson, is "made up of places" (1979:29) and describes "where" the animal lives. A niche, as opposed to that, "refers more to how rather than to where it lives." (ibid.: 120). Gibson defines the niche as a "set of affordances" (ibid.). The primary focus of Gibson's work are affordances as individual motor possibilities – the ecological niche enables and defines the spectrum of daily actions such as grasping, sitting, hiding, drinking etc.

Rietveld and Kiverstein give more consideration to the fact that an ecological niche – "a network of interrelated affordances" (2014: 330) – "is build and transformed by members of the species through the species' typical ways of acting." (2014: 328). Following the philosophers Wittgenstein and McDowell, Rietveld and Kiverstein "situate affordances in the context of a form of life." (ibid.: 330). They emphasize that "even unreflective human actions are situated within communal practice" (ibid.: 333; McDowell 1998; Rietveld 2008). A "form of life" is a regulated pattern, which determines an individual organism, without being fully determined by the individual organism's intrinsic features.

Sociality, it seems, is relevant to a theory of human affordances in at least two different senses. On the one hand, sociality is relevant in the sense that the cooperative and communicative nature of human forms of life transforms the ecological niche of humans as a whole. On the other hand, there is a narrower

sense in which the sets of affordances for humans contain "social affordances" among others. Social affordances in this narrow sense are, as already mentioned by Gibson (1979: 127), other people who afford behaviour – for example "sexual behaviour, nurturing behaviour, fighting behaviour, cooperative behaviour, economic behaviour, political behaviour" (ibid.). Kiverstein (2016) speaks of "social affordances" as other people whose presence, whose face expressions or culturally assigned social roles constitute affordances, for example through empathy.[92]

The younger generation of ecological psychologists, represented by Rietveld and Kiverstein, seems to put special emphasis on the criticism of labelling reason and rationality as "higher-order" functions or "higher" cognitive capacities. Gibson, as already mentioned, sees his theory of affordances as mainly accounting for motor possibilities and allows for the possibility that affordances can be misperceived and that we must "learn to see what things really are" (1979: 134) – for example to see that the "innocent-looking leaf is really a nettle" or that the "helpful-sounding politician is really a demagogue" (ibid.). Although Gibson does not specify how this is possible, it seems that he still draws a distinction between critical reflection and habitual motion within an ecological niche constituted by affordances. At this point, Rietveld and Kiverstein depart from Gibson explicitly (2014: 325f). They state that, mainly through the richness of culture, the human form of life is characterized by a great variety of abilities – not only the classical motor possibilities. All these different abilities, so the new paradigm, are available only because of a particular ecological niche (ibid.: 326). The culturally shaped and transformed ecological niches of contemporary humans contain, according to Rietveld and Kiverstein, "affordances of higher cognition" (ibid.: 342f). They emphasize that "allegedly ´higher´ human abilities", such as the "practices of giving and asking for reasons", can be described as "skilled engagement with affordances" (ibid.: 343).

Merely regarding the terminology, it seems that the "higher-order" theory of consciousness, defended by Brown/LeDoux (2017) and supporting Mercier/Sperber (2017), and the framework of ecological psychology oppose

[92] Recently, Lo Presti (2020) has suggested to understand persons as distinct affordances – affordances that can be distinguished from the affordances that a human creature offers as a biological organism. Moreover, Weichold and Thonhauser (2020) have proposed a concept of "collective affordances", describing affordances that are genuinely offered to groups as agents rather than to individual biological organisms. Both the concept of "collective affordances" and the concept of "social affordances" in the narrow sense can be distinguished from the kind of affordances that are created in the course of a social or cultural transformation of the environment.

each other.[93] However, I think that the term "higher-order" theory as used by LeDoux and Brown should only be understood as emphasizing the distinctiveness of reflectively conscious states within the brain, not suggest an order of complexity or any other evaluative order. Indeed, the term "higher" could be avoided, if we speak instead, as Mercier and Sperber do, of a functionally distinct module, envisioned as one among others.

As established in subsection 4.2.4., reflective reasoning is not "special" in the sense that it is a "global", all-comprising critical faculty. However, reflexivity is "special" in the sense that it is the upshot of a social justification practice, not the far end of a continuum of consciousness which all psychological processes possess to a certain degree. Instead, reasoning as a reflective stream of thought is one cognitive faculty among others. As such it can be accounted for with the same basic framework as any other non-reflective form of behaviour. Both the aspects of an environment and the abilities of a form of life are, in the case of humans, profoundly shaped by culture and communication. Communication in terms of reasons seems to be one specific ability contained in the collective form of life of humans. In the following, I will give a detailed account of reasons as affordances of reflective thought. A major challenge in this endeavour is posed by the fact that the "realm of reasons" seems to be more separated from the material environment than other types of affordances, so that reasoning – moving in the realm of reasons – can be done independently of moving around in the world.

4.3.2. Affordances of thought: the realm of reasons in the human environment

Human activity, according to ecological psychology, is enabled and shaped by so-called affordances. Affordances are ability-related aspects of the environment, aspects that invite and can solicit particular responses by participants in a form of life. The network of affordances that shape a particular form of life is called a niche in ecology.

The probably most challenging task that ecological psychologists have set themselves within the recent years is to account for the ability to reflect in terms of reasons. Rietveld and Kiverstein are committed to the claim that this ability "can also be made sense of in terms of skilled engagement with affordances" (2014: 343). Following that intuition, this subsection provides an

[93] I thank Martin Weichold for pointing out these worries to me in much detail.

original understanding of entrenched structures of communication forming a distinctive subset of affordances: Affordances of reflective thought.

There is already one proposal for understanding the "stream of thought" within an ecological framework, which is offered by Reed (1996: 169). According to this proposal, "the stream of thought emerges from prospective awareness" (ibid.: 175). This means that when we reason, we are not moving in a separate mental realm. We only anticipate or imagine the physical world in which we will or could move. Against Reed, I will argue that, in the reflective stream of thought, we do not respond to an imagination of the world, which poses the same affordances for us as the real world would do if we in fact encountered it. As opposed to that, we can in fact say that, in reasoning, we move in a separate, socially created realm of reasons. This realm is a subpart of the natural environment we inhabit. Nevertheless, it can be described as a distinct set of affordances: Affordances of thought that invite particular responses of mind. Mind, according to Mercier and Sperber´s definition delineated earlier, is the module of the human brain that can deal with reasons – facts that justify, to oneself as well as to others. With the concept of a module as a functionally distinct and partly autonomous subsystem within human cognition, we can get to the idea of a specific type of ability – the ability to deal with reasons. In accordance with this well distinguished ability, we can distinguish a particular type of affordance – a particular type of aspects of the environment to which the ability to respond to it as a reason is related.

The distinction between "aspects" and "features", introduced by Rietveld and Kiverstein (2014: 335), is of crucial relevance for understanding why the stream of reflective thought is not just prospective awareness of the environment. As Rietveld and Kiverstein hold, the environment of a form of life does not simply have features. As opposed to that, the features in the surrounding can be individuated and made sense of in different ways, depending on culture. This opens up the possibility that any given set of features and object constellations can have different aspects for differently encultured beings. However, it also opens up the possibility that a situation can have different aspects for one and the same being – relating to different kinds of abilities of this being. These different aspects can be responded to alternatively, sometimes simultaneously. Thus, a situation can have an aspect that affords a reflective response of the reason module and at the same time an aspect that affords a non-reflective response of another kind. Sometimes these different "aspects" can be one and the same "feature", as will be shown.

In some cases, the different responses may be perceived as perfectly congruent, in others we might perceive a tension. A glass of water, for example, can constitute a classical Gibsonian motor affordance to drink from

the glass. At the same time, a thought of the form "This is water" can function as a reason to drink in the reasoning process. The thought *that this is water* gets available in certain circumstances. It constitutes an affordance for treating it as a reason to drink. However, this affordance is not the same as the motor affordance to drink, which can solicit drinking completely without reflective awareness or reasoning. Instead, the affordance for the thought is a type of affordance of its own, affording a type of activity of its own – the activity of reflecting in terms of reasons. There seem to be "places" in our processing were thoughts present themselves, where they "pop up". We do not seem capable of deciding to form a particular thought. We seem even less capable to create the thought how we want it. At best, we can focus more consciously on our thoughts and thereby intensify a sense of guidance. Nevertheless, the thoughts are presenting themselves and they present themselves at "places" where they "fit" or "suggest themselves". Later on, we will take a closer look at these "places" at which thoughts suggest themselves and normative conclusions are formed inevitably.

Let us first continue the argument for why affordances of thought are distinct from other affordances. If there are no reasons against drinking from a particular glass, the motor affordance for drinking and the thought that serves as a reason for drinking might be perfectly congruent. So congruent that we usually do not need to form the thought! We might, however, have a reason not to drink from the glass if it contains petrol instead of water. It is not necessary that the reason pops up on our mind for it to be possible to speak about there being such a reason. Nevertheless, we can only speak about the consideration being a reason, if the mind responds to it in the right way *when* it is confronted with it. If we are simultaneously confronted with the reason or thought *that this is petrol* and with the motor affordance to drink because it looks like water, our environment presents aspects that afford contradictory activities. Since we cannot both drink and not drink, we must override one affordance – hopefully the affordance to drink.

There are examples in which it is not as easy as that to override an affordance standing in a potential tension with another. Such cases show exemplarily why affordances for reflective thought and affordances for non-reflective behaviour or physiological responses are functionally distinct and relatively autonomous. Let us first look at a case where a reflective thought stands in tension with a non-reflective behaviour. We can consider what is labelled "aversive racism" (Gaertner/Dovidio 2000). Aversive racists are people who explicitly endorse egalitarian values and would never take discriminations based on race as justified. Nevertheless, empirical observations show that the very same people show aversive behaviour against members of certain racialized groups, such as

avoiding eye contact, changing sides of the street or neglecting them when offering help or distributing benefits. A viable interpretation of this phenomenon is that, while a certain skin colour or other racialized feature is not an aspect that can function as a valid reason for discrimination in a reflective process, it can very well serve as an affordance for aversive behaviour. In the reflective process, by contrast, the *thought that all humans are equal* might present itself as a reason and might outweigh all other considerations that pop up in relation to this topic.

Such incongruence, important to note, is not a cognitive dysfunction. It can at best create a psychological tension. As normatively reflecting persons, we are invited to reflect on the question how to solve this tension. The outcome of this kind of reflection is, we must understand, a matter of which kind of thoughts pop up for our reflective mind and present themselves as justifying. Which thoughts, or reasons, this are, depends, as I will suggest in the following, on the communication practice in which somebody's mind is embedded.

Let us consider some cases that show that affordances for reflective thought can be different for differently embedded minds: Left-wing liberal Europeans, for example, might be responsive to some aspects of people as favouring in-group behaviour, while as reflective reasoners, they are not responsive to any of these aspects as a reason. On the contrary, they are responsive to reasons *against* in-group bias. As opposed to that, we can conceive of reasoners, encultured differently from left-wing liberal Europeans, who would indeed count the thought *that this person is related by blood* as a reason to assign her a higher moral status than a non-relative. Moreover, this reasoner may find it utterly repulsive to ignore the force of this reason.

To be sure, there are usually certain links between the reflective thoughts to which one's mind is responsive and the other aspects of the environment to which the organism is responsive. This, of course, reduces incongruence and tension and makes life easier. However, the systems are functionally distinct, and a tension is not strictly speaking a dysfunction. It is, for example, common that humans accept the fact *that this is my child* as a reason to care for it more than for other people, while they at the same time show non-reflective behaviour of care and nurturing. This is, however, only a natural link, not a necessary connection. The two types of responses are clearly separable. We could conceive of somebody, for example some radical Marxist, who is not responsive to the fact *that this is my child* as a reason for privileging it in any way, although the Marxist might still be responsive to the aspects that solicit loving and nurturing behaviour. This, to be sure, might make the Marxist unhappy or even sick. The point is, however, that normative justification is a separate question. In a process of normative justification, the Marxist's

reflection might not even be responsive to the fact *that this makes me unhappy*. Alternatively, this fact might present itself, but as having less weight than the fact *that this contributes to abolishing a society build on blood bonds*.

This shows that the facts or thoughts that present themselves to the reflective mind in the process of justification are distinct objects. The subject is spontaneously responsive to them, it does usually not derive them from itself and its other responses to the environment. These objects, it seems, form a rigid structure of its own – a distinctive subsystem of the environment. The schema theoretic approach in the next section will have a closer look at the genesis of such an environment by communication. At this point, we must understand that it is in fact a rigid environment, independent of the individual, but also distinct from the world of motor-behaviour, though we need not assume anything "immaterial". We can for example consider the fact that it is possible for two persons to cognitively move in different communicative environments, although they are currently standing in the same place on earth and talking to each other. Accordingly, the "realm of reasons" is indeed a subsystem of the environment, which is not coextensive with the geological environment.[94]

Understood correctly, this picture is compatible with ecological psychology's usual emphasis on the unity of the ecological niche. Ecological psychologists favour the idea that all the affordances relating to our activities are "stored" within one and the same material environment. Gibson is exemplary for this when he rejects the idea of a distinction between a natural and an artificial or cultural environment. There is only one environment – the encultured environment is "the same old environment modified by men" (1979: 122). According to Gibson, it is "a mistake to separate the cultural environment from the natural environment, as if there were a world of mental products distinct from the world of material products." (ibid.) Likewise, Reed wants to account for the possibility of a stream of thought – which for him is the upshot of guided and organized prospective awareness, the upshot of a form of encultured learning – without assuming "mental states transmitted from one person to another, or from one generation to another" (1996: 177). The idea of a mental state, it seems, is associated with the assumption of a reality "over and above" the reality of a human animal moving in a material surrounding.

This fear of "mental states", however, seems ungrounded if we understand "mental states" simply as instantiations of reason judgments – instantiations that form parts of a justification practice entertained within a material community in a material world. The "mental", as proposed in 4.2.4., is neither

[94] Mind the Gibsonian distinction between an "environment" and the "world of physics and geology", and furthermore the distinction between a "niche" and a "habitat".

an entity that gives individuals "inner depth" nor an entity that gives an individual autonomous control, thereby elevating it over the merely instinct-driven parts of its behaviour. As opposed to that view, the proposed theory depicts the instantiation of mental states as a response to an affordance in qualitatively the same way in which grasping something graspable is a response to an affordance. This means that we can accept the idea of the unity of the ecological niche as a material environment in principle. Doing so, however, is compatible with distinguishing between different subsystems within the range of our abilities and, accordingly, different subtypes of affordances. Reasoning, we can maintain, is still a process within the one and only material environment.

However, the way in which reasoning transforms the material environment is in a sense profoundly different from other ways in which collective synchronisation of behaviour or encultured guidance of awareness transform the material environment. When a culture engages in explicit justification practices, it not only changes the awareness and normative significance of the environment, it adds a new type of furniture to the environment: *Statements that can be accepted as justifications*. These statements sometimes refer to material objects or their properties, sometimes to statements made by others. However, the statements are objects themselves – objects that are not mysterious. We need not assume anything "immaterial". Reason statements are simply objects in our environment. Stating the fact *that this is my child* as a reason to care for it is not the same as, for example, responding to a visually or auditorily perceivable feature with caring behaviour. When people come up with a reflective thought – a proposition they understand as justifying – they add something to the environment, to people and objects of daily routine.

In order to understand such statements as effectively shaping our environment, however, we must understand them as building an entrenched practice – iterated and systematized over time and stored within interrelated nervous systems. Initially, to be sure, there were no entrenched potentialities in the surrounding. At some point, our ancestors must have started to just come up with reasons. However, before the practice is entrenched and taken seriously, there is no human being having a set of reasons at its reflective disposal and a pronounced sense of their weight in relation to each other. This full-fledged phenomenon is only possible as the upshot of an environment in which reasoning practices are deeply entrenched and taken seriously. The only creature that is responsive to reasons as having "weight" and "normative force" is a human being embedded in an entrenched reasoning practice. When the individual gets something right, the mind typically exhibits a distinctive phenomenology – the phenomenology of being a self-determined reasoner in

a causally, spatiotemporally and normatively unified framework. For a reflective creature, there is a potentiality of reasons that "work" in this way – however unknown these particular reasons may be in some cases. Individuals may be ignorant of these potentialities. They may also misconceive them. However, they are there. There is nothing scientifically problematic in saying that they are there, as potentialities in interrelated nervous systems, materially, even if there are no concrete physical objects we can point at.

Each reasoner in a given situation, to sum up the idea, is surrounded by an extensive set of aspects or considerations that can function as reasons in that situation. The situation encompasses everything from the state of the world to the state of the individual organism. What distinguishes this picture from the picture Mercier and Sperber evoke, when they talk about mind "producing" reasons, is the idea that reasons *surround* the individual in each given situation. The individual *responds* to something in its surrounding. This picture makes understandable why a thought, as described above, usually "pops up" at a "place" in the reflective process in which it fits. If we understand reflective processes as moving within entrenched communication practices, we see why thoughts have their places and why some fit better into these places than others. According to this picture, it is possible that an individual, being embedded in a particular communication practice, is dislocated and communicates with individuals being embedded in other practices. Unless the involved people find *some* common ground and can engage in real communication from thereon, it seems unmysterious to say that they are partly moving in different environments, although they currently share the same geological location.[95]

This section showed how we can envision the realm of potential mental states as the scope of reason statements which a community allows as justifications. It further showed how we can envision these potentialities as something that is there, at the reasoner's disposal, and that constitutes an affordance for thought – an invitation and sometimes solicitation to instantiate the mental state of recognizing it as a reason. The key to locating the realm of reasons in a natural environment lies in the crucial fact that mind in an individual – mind as instantiating reason judgments or mental states – is only the module that serves as a presupposition for participating in a communicative practice. The full-fledged phenomenon of mind, we could say, lies in the communicative network itself. We need the communicative network not only

[95] However, it is important to note that, in the case of permanent dislocation, the reasoning process itself becomes dysfunctional, since "reason" is not entirely "internal" to the individual, but consists in relation to the relevant environment (see also Brinck/Liljenfors 2013a)

to develop the capacity of reasoning to begin with, but also in order to evaluate and make sense of what is instantiated in a particular individual.[96]

The realm of reasons, accordingly, is not "in the head" of an individual. However, the realm of reasons must be recognized as separate from the environmental objects to which we sometimes point as reasons.[97] For example, if we point to the fact *that the hat is yellow* as a reason to buy it or to the fact *that the sun has risen* as a reason to get out of bed, we point to "aspects" of the environment – "aspects" picking out "features" that could as well serve as affordances for motor actions or other forms of behaviour. If these "aspects" serve as affordances for thought, they are, however, not identical to the "aspects" that may serve as motor affordances, though they might be related to the same "feature". The fact *that the sun has risen* is not the affordance that makes non-human, pre-linguistic animals rise from sleep. It is a proposition. In the reading proposed in chapter 3, being a proposition is simply being a thinkable or graspable thought, i.e. a potential mental state - in light of the empirical research explored in this chapter, a potentiality in an entrenched communicative system.

Accordingly, the basic Fregean rationale, slightly modified in the previous chapter, turns out to be compatible with an ecological framework speaking of thoughts as affordances of mind. Frege's rationale is that when I grasp the thought *that the sun has risen* I am not grasping a physical object emitting rays, but an ontologically different object (1956: 292). The ontological difference, this chapter has suggested, is the difference between an affordance for thought on the one hand, and an affordance for motor behaviour on the other. Both types of affordances, however, are parts of the ecological niche that humans have carved out for themselves.

If we want to call this a dualism between motor behaviour and thinking, we must at least call it a very harmless form of dualism. Thinking, according to this view, is a form of behaviour as well – not "special" or "higher" or more

[96] For the first point, the ability to reason, memory traces relating to a communicative network stored within an individual brain may be enough. For the second point, the evaluation of a mental state, its content and implications, we need the network itself. See also Tyler Burge's (1979) view on mental states. Moreover, my view is compatible with, but not dependent on a so-called "open-mind view", as it is defended by Prinz (2013). For discussions of this work see Blomberg (2013) or Wringe (2013).

[97] At this point, we can draw the connection to the Fregean idea of a realm of thought as a "third realm", in between the physical world and the world of subjective imaginations. Chapter 3, following a neo-Fregean interpretation, interpreted the realm of thought as a realm of potential mental states – states that are instantiable, available for someone with a particular form of mind.

"autonomous" than other forms of behaviour. The realm of reasons, we can say, is material as well – entrenched and stored in closely interrelated nervous systems. As such, a reason is not a mental item or a "subjective notion" (Reed 1996: 177) in the sense that is criticized by ecological psychologists and Fregean philosophers alike. Moving in the realm of reasons can be described as a way for the human animal to "keep in touch with its surroundings", to use Reed's ecological definition of cognition (1996: 169). When reasoning, we are not in touch with other people as "social affordances" – we are in touch with a created, but independently holding structure that determines us but also gives us new possibilities and enriches our "landscape of affordances".

4.4. Establishing an archive: reason relations as schemata

The previous section described reasons as affordances of reflective thought. The ecological concept of an affordance helps us to develop a naturalistic conception of reflecting in terms of reasons – a conception that is in line with both the theory that reasons are the products of a social justification practice and the idea that, by grasping reasons, we get access to a unified and meaningfully organized world. I argued for a realm of reasons as a subpart of our shared, socially cultivated environment.

Now we might ask: how can "realms of reasons" emerge within communities? How exactly are we supposed to envision the historical communicative process of establishing reasons? Psychology offers a theoretical framework for understanding the cultural acquisition of habits of thought that is compatible with the ecological framework: so-called "schema theory". This psychological approach can help to make the collective creation of affordances of thought understandable. In this section, I connect the idea of reasons as affordances to the resources of schema theory. Thereby, it shall become clear why the realm of reasons can be understood as created in a collective archival process. Moreover, I will discuss some additional advantages of the metaphor of an "archive".

Subsection 4.4.1. gives a short introduction to the basics of schema theory, while 4.4.2. specifies the idea of a story schema and its role for understanding ourselves as acting meaningfully in a unified and coherent world. I suggest the term *collective fictions* for story schemata that shape individual mind. This term can be contrasted to Skorupski's narrow definition of fictions as

subjective irreals. While Skorupski describes reasons as objective irreals, they can still be fictions according to the proposed understanding.

In subsection 4.4.3., I contrast the idea of a story schema directly to the idea of a reason relation used by philosophers such as Scanlon and Skorupski.[98] Finally, in subsection 4.4.4., I defend the idea that the reflection of a solitary mind – an apparently self-sufficient and autonomous activity – is moving in an archive, a fund of normatively salient considerations inherited by cultural ancestors and used with others.

4.4.1. Schemata as the basic building blocks of cognition

Schema theory is an approach in cognitive psychology that deals with the acquired "habits of mind", to use a formulation by Cristina Bicchieri (2005: 55). A schema is typically defined as a small complex of information in which a response – an action or attitude – is linked to specific situational features that make it appropriate or meaningful. The acknowledgment that situational aspects suggest us the actions we can do, even the mental attitudes we form, is characteristic for ecological psychology and schema theory alike. Unlike ecological psychology, which encompasses cognition in a broader sense, including both inborn and acquired features, schema theory concentrates entirely on the acquired structures of cognition. It emphasizes the aspect of how cultural communities create, shape and reinforce the embodied knowledge of situational cues.

Bicchieri, in her book *The Grammar of Society* (2005), describes the observation that social norms "become salient and active only under certain situational conditions" (ibid.).[99] Even our preferences and beliefs, she observes, "are sensitive to situational cues" (ibid.: 56). Apart from social theorists such as Bicchieri, the already mentioned neuroscientists Brown and LeDoux (2017) and the developmental psychologist Piaget (1954), among others, work with the assumption of schemata. David Rumelhart refers to schemata as the "basic building blocks of cognition" (1980). This builds on studies, such as the memory studies by George Mandler (1967), which suggest that without schematic structures we are not able to memorize – neither

[98] I thank Peter Railton (met 2018 in Kent) and Elijah Millgram (met 2018 in Parma) for detailed discussions about interpreting reason relations as schemata and encouraging the approach.

[99] I benefitted greatly from the workshop "Norms and Change" with Cristina Bicchieri, organized by Wolfgang Huemer and Carla Bagnoli in Parma (2018). I further thank Eric Brandtstedt for suggesting me to look deeper into the work by Bicchieri.

experienced events nor written texts. Memory, it seems, plays a crucial role for an individual's capacity to think reflectively.

Brown and LeDoux (2017) explicitly assume a schema theory about the conscious mental states, which I identified as instantiations of reason judgements. A first systematic outline of the schema theoretical approach is provided by Jean Matter Mandler, with his book *Stories, Scripts, Scenes: Aspects of Schema Theory* (1984). Schema theory, as already mentioned, builds on the assumption that we structure our perceptions according to acquired bodies of information – "small networks of information that become activated as we experience […] things" (Mandler 1984: 3). Such bodies of information can embrace "scenes", (representing "places, rooms, streets in which workaday routines take place"), "scripts", (habitualized "routines of workaday events") or "stories" ("literary expressions we read or hear") (Mandler 1984:1).

There are huge overlaps between the ideas of schema theorists and parts of ecological psychology, such as, for example, Reed's idea of how human caregivers "introduce infants to the daily routines and everyday places of their culture through intense repetition and rhythm of experience" (1996: 173). According to Reed, "ideas are not mental states transmitted from one person to another". Instead, ideas are stored within the environment, in the form of "environmental regularities", shared by people in the form of "regularities promoted in their populated environment" (ibid.: 177). Likewise, the "bodies of information" referred to as schemata are not a content of consciousness but rather the structures that underlie and shape consciousness. "Consciousness is a construction" (Mandler 1984:33), according to schema theory. Usually, our consciousness follows schema structures, while we cannot "observe the procedure at work" (ibid.).

Schema structures, typically without our explicit knowledge, "organize our spatial and/or temporal knowledge about objects, events and places" (Mandler 1984: 4). Schemata can also be described as "sets of expectations" (ibid.: 13), containing information about what specific persons in specific situations are going to do or, normatively, ought to do. Such expectations are created in a process of cultural learning through repetition and acquiring routine. This holds for both routines of bodily motion and routines of reflective thought. According to the proposed view, bodily motion and reflective thought are structurally analogous in the way they work. While it is, however, easy to imagine how bodily skills, paradigmatically dancing, are acquired by repetition, this view of learning is more difficult to defend for the case of reflecting in terms of reasons.

In order to understand how we acquire schemata of reflection, or more precisely schemata of reason-giving, we have to get familiar with a specific type of schema – the "story schema", as defined by Mandler, who has a special interest in the phenomenon of a "folktale" (1984: 17). Folktales exemplarily describe and organize normatively salient situations. They shape the human sense of what matters in each situation, of what matters for a particular type of person and of which attitudes and behaviours are expectable of that person. The following subsection explores the idea of a story schema and the role of narratives for understanding, especially for the type of understanding that is involved in apperception as characterized by Skorupski.

4.4.2. Collective fictions and the understanding of self and world

Skorupski formulates that establishing warrant about reason relations is in practice a "collective, archival process" (2010: 124). Ontologically, however, he describes reason relations as "objective irreals" (ibid.: 420). As such, they are distinguished from "subjective irreals" (ibid.: 425). In the category of subjective irreals, as Skorupski defines it, are fictions or putative irreals. Subjective irreals are "anchored" in imaginations or plans dependent on someone's mind, whereas objective irreals, like Fregean thoughts, must be taken by mind as they are. They are recognized, rather than created.

But are fictions, as Skorupski stipulates, always subjective irreals, anchored in the imagination of an individual? Is this the only kind of fictions we can identify? Or could there be fictions not dependent on an individual's mind? Mandler's concept of a "story schema", based on investigations of folktales, invites us to develop such a concept. Mandler's investigations illustrate how fictions transport relational information that shape individuals, more than particular individuals can ever shape them. I suggest speaking of *collective fictions* – fictions that are objective, rather than subjective irreals.[100]

It seems clear, as also suggested by the large extent of cross-cultural overlap between folktales, that such tales are not arbitrary inventions. They respond to frequently occurring problems of human existence in the world. However, the stories themselves, as well as the significance they attribute to particular

[100] I thank the audience and the organizing team of the *Graduate Workshop in Fiction and Philosophy* (Lund 2018), especially Frits Gåvertsson, Thérèse Söderström, Nora Hämäläinen, Maria Green and Nils Franzén. Moreover, I thank the fiction reading group from the spring term 2019 at LOGOS (Barcelona), especially Enrico Terrone, Filippo Contesi, Manuel García-Carpintero and Aarón Álvarez González.

features of existence, have to count as "myths" or some sort of "fictions". It is undeniable that mind, in making sense of events that affect it, is creative. Mandler builds on studies according to which the human mind imposes structure and meaning even on meaningless series of digits (1984: 20). Likewise, Gregory Currie stresses the role of (inventory) narratives in human perception and control of the world:

> Narratives encourage us to make sense of the world by telling of the ways in which mind controls it. Perhaps narrative encourages us to think of the mind as more structured, more orderly, more robustly in control of circumstance than it really is (Currie 2010: 187).

Stories we tell, according to that picture, give us a grip on the world that affects us. Clark, in his enactivist-based predictive coding account of the brain, also uses the term "gripping tales" (2016: 202) for affordances that enable unified responses of an organism. In the case of story schemata, we can understand this formulation very literally. We could describe story schemata as "gripping tales" tailored to a social world – a social world in which we orient and locate ourselves. Story schemata do not only spatiotemporally and causally unify our perception, they unify and structure the normative world in which we lead and plan our lives as self-determining creatures.

Velleman argues that narratives provide us with a genuine type of explanation. Narratives are a "genre of explanation" (2003: 1) as he calls it. Narratives, Velleman claims, are "conveying not just information but also understanding." (ibid.) Thereby, he sets himself off from an account of narratives as representing chains of events as being causally linked, one event necessarily following from an antecedent (defended by e.g. Carroll 2001; Currie 2006). As opposed to that picture, Velleman claims, narratives present us the events in a way that makes sense, building "arcs of emotion" (2003: 17) and making use of "emotional cadences" (ibid.: 18), thereby achieving what he calls "emotional closure" (ibid.: 20).[101]

Velleman presents an account of understanding that provides valuable insights into the nature of apperception – the distinctive way of recognizing

[101] It is unclear, however, whether causal links can be understood independently of a framework in which they make sense at all. On this distinction hinges also the distinction between "intellectual closure" and "emotional closure", which Velleman stresses in his paper (2003: 20). In light of the frameworks for understanding cognition suggested so far, it is questionable whether we can recognize causal links, and identify them as such, without having what Velleman calls understanding. In the modern natural sciences, there may be possibilities to establish connections based on evidence, which mind cannot directly understand. This can neither be denied nor established at this point.

reason relations, characterized by understanding the single act or mental state within a coherent and unified framework, spatiotemporally, causally and normatively. Velleman develops by reference to the philosophy of history, especially to Mink's (1972; 1987) approach to history as a mode of comprehension, that, in understanding, events are not represented in an interlinked series. In genuine understanding, by contrast, we represent "actions and events … surveyed as it were in a single glance as bound together in an order of significance." (Velleman 2003: 18) Understanding, according to this view, is an instance of "grasping together", rather than grasping successively.

Going back to the previously discussed concept of apperception, this does not mean that several actions or events are the content of our mental state when we apperceive them. Rather, we apperceive one single fact, but *as* standing within a unified whole. The concept of apperception, we can conclude, is basically build on the same idea as Mink's "grasping together" in historical comprehension.

Velleman identifies this idea as the distinctive feature of narratives. The human tendency of narrativity manifests the cognitive capacity of "grasping together" – a way of using our cognition that gives us a sense of standing in a systematic and unified context, a way of using our cognition that also gives us a sense of "closure". We could understand a sense of "closure" here as a sense that the world is clear to us and that all (relevant) questions are answered.[102] There seems to be an analogy between the concept of "closure" and Skorupski's concept of "warrant" (see Skorupski 2010: 108ff).[103] According to the concept of "warrant", there is in principle an answer to every normative question. The concept assumes that we do not only recognize reasons spontaneously. In the same way in which we recognize reasons, we also recognize when it is enough to inquire for further reasons. Even if we can never reach the point where we understand the whole world, we can still reach the point at which we know everything of relevance for our situation. Neither "warrant" nor "closure" requires omniscience. As opposed to that, these

[102]Velleman distinguishes between "intellectual closure" and "emotional closure". However, the distinction between "emotion" and "intellect" is probably not that sharp. This is also backed up by Brown and LeDoux (2017) who identify the phenomenon of conscious emotions with the phenomenon of conscious reflective states.

[103] In order to be true to Skorupski, we must concede that his picture is a bit more fine-grained, mainly because he distinguishes between "warrant" and "justification" and, accordingly, between "Kantian autonomy" and "Hegelian autonomy". Chapter 3 argues that this distinction is not tenable in light of the modified reading of mind-independence suggested against the radical Fregean reading. The picture discussed here assumes the modified reading of Skorupski's theory defended previously.

concepts only require us to have a clear grasp of the world as it is relevant in a particular moment for a particular subject.

As previously developed, the phenomena of "closure" and "grasping together" are ways of representing a world, which in fact only stands in contact with us via a series of sensory inputs. It is not difficult to see that events we "grasp together" do not really occur together and that "closure" does not really represent the boundaries of the physical world. As the predictive coding framework presented in 4.2.2. suggests, the apperception of a unified world is an active creation of the brain. The possibility of other forms of unity is conceivable, though these other forms are not themselves conceivable for our mind.

The particular responsiveness of our spontaneous subjectivity, or our reason modules, is qualitatively the same as our responsiveness to narratives. One way of interpreting this phenomenon is to conclude that we respond to particular narratives with understanding because they match a more basic responsiveness of us as subjects. Another way of interpreting the same phenomenon is to assume that we acquire responsiveness as subjects because we have narratives. The first interpretation would amount to a classical objectivist view of reason responsiveness. The second interpretation, by contrast, amounts to a theory of reasons as collectively produced in the most radical sense. All the reason relations to which we are responsive, according to the latter view, are schemata that emerged from narratives which were repetitively told until they became distinctive stories with recognizable schema patterns. Both of these views, it seems, are too radical. It seems doubtful that we can decide for one of these interpretations on good grounds.

An alternative suggestion, that still seems enough for speaking of reasons as collective fictions, is the idea that narratives and normative responsiveness co-emerged under mutual formative influence. We need not endorse the extreme claim that societies started to construct arbitrary narratives out of pieces that were previously meaningless and mere semantic "noise". However, we must acknowledge that any full-fledged exemplification of a reason responsive subject is to a large extent shaped by the narrative practices in which it learned to think and weigh reasons. Narratives, as mentioned above, have similar contents across cultures and partly respond to genetically entrenched basics of human nature, to be sure. However, the encultured practice of telling stories plays itself a crucial role in shaping and enforcing our responsiveness.[104]

[104] The narrativity that shapes our responsiveness need, of course, not only consist of "folktales" in the classical sense. Especially in modern times, the "gripping tales" we learn and use for developing understanding, come from various sources, shaped by families, cultural and

Stories that are made up, even if not arbitrarily, shape the way in which we speak about actions and situations. It shapes what we can understand and what we can accept as justification.

Story schemata, we can paraphrase, are active collective creations that are central to both communal and individual life and that are passed on from generation to generation. They open up the social space in which individual agents can think of themselves as self-determining on the basis of reasons. We can now draw a connection between the concept of a schema, especially the concept of a story schema, and the concept of a reason relation.

4.4.3. Reason relations first-personally – schemata third-personally

The concept of a schema from psychology is comparable to the concept of a reason relation from normative theory. This subsection will discuss the analogies, but also the differences in depth. What we recognize as objectively holding reason relations from the normative first-person standpoint, it seems, can be identified as schema structures working on our cognition, when we investigate them from a detached third-person standpoint.[105]

There seem to be obvious differences between schemata and reason relations, but also similarities in the basic structure of both. A schema is an acquired body of information activated by certain experiential or perceptual

religious communities and, to a growing extent, social media. The role of media in transporting narratives would be a fruitful and promising research topic of its own.

[105] I have already used the distinction between first-personal and third-personal standpoint or perspective in the first chapter. The third-personal view on ourselves shall be understood as a view on ourselves as purely natural creatures, abstaining from all normative evaluation and identification with those creatures. The term "third-personal" in this context can easily be misunderstood, since it could also mean to think about a third person in normative terms or to explain another person's judgments in normative terms (I thank Björn Petersson for pointing this out). Likewise, what I call "first-personal" reflection can be understood as including genuinely normative judgments as well as more detached or scientific evaluations of what is the case, in so far as those judgements flow naturally into our thinking about our actions and the world (I thank Johan Brännmark for making me aware of this). Whether a view is "first-" or "third-personal" in my sense depends on whether we are normatively interested and identify with the creature we view. Another terminology, such as "practical" and "theoretical" standpoint (Kant and contemporary Kantians such as for example Korsgaard 1996, O'Neill 1989 or Kleingeld 1998), or "reactive" and "objective" stance (Strawson 1962) could also be helpful here. Each of these terms, however, comes with problems of its own, why I choose to employ "first-" and "third-personal" as technical terms in this context. The distinction resonates with the distinction between conceptual and material analysis of reasons, also introduced in the first chapter.

key features. Typically, a schema is connecting a situation and a person with an action or attitude, which is considered appropriate in that connection. A reason relation, by contrast, holds between a propositionally represented fact and a subject, an action or attitude, a time (Skorupski) or circumstance (Scanlon) factor and (only in Skorupski) a degree of strength. The most crucial distinction between reason relations and schemata is probably the clarity and determinateness of the normative information yielded by the reason relation. For a schema, by contrast, it is characteristic that it favours particular connections, but does not determine thought and action exhaustively. Mandler emphasizes the "multiplicity of relations", which accounts for the "richness and flexibility in our thought" (1984: 6). As opposed to that, it is characteristic of the reason relation that it objectively determines the situation. Skorupski depicts the normative domain as perfectly unified.

The structure of a schema network is fragmented, perhaps even internally incoherent. The domain of reasons, by contrast, is systematic, unified, coherent. This contrast seems striking at first glance. However, the contrast disappears if we take into account that the unification and systematicity of the domain of reasons is not just there. We must understand it as actively achieved in every instance of apperception. The unity of the normative landscape must be achieved in exactly the same way as the unity of the causal and spatiotemporal field in which we understand ourselves.

The apperception of a causally and spatiotemporally unified field, as developed in 4.2.2., is not a depiction of what is out there, but a model of the brain to predict sensory input and minimize prediction error. The causally and spatiotemporally unified field can thus be classified as an irreal structure. Likewise, the unity of the normative landscape must be understood as a model of the brain – a model that enables self-determination in light of reasons. Depicting the landscape as unified is what cognitive processing must achieve in order to proceed as an activity of self-determination. In other words, the unified fields – the causal, the spatiotemporal and the normative one – are phenomena that characterize the instance of apperception.

To understand that idea, let us consider in more detail what it means for a particular reason relation to hold definitely and unambiguously, by looking at the structure of a reason relation again: R (π, x, t, d, φ). What is crucially involved in making π a reason, is the factor t. It means that the normativity of π is not only relative to a particular empirical subject, but also relative to a particular point in time. The reason relations that hold for a subject at any time t may cease to hold or be entirely different at another time. When we apperceive a reason, we have actual first-personal insight into a relation that holds in a specific moment specifically with respect to us. What is necessary

for any reason relation R to hold definitely and unambiguously for us in a specific moment is just that this relation R is systematically connected with all other reason relations holding at that particular time t. Systematically connected here means that, considering all reason relations that hold with regard to their degree of strength, we arrive at a clear order and an unambiguous judgement of which reason is strongest. Otherwise, we couldn't understand ourselves as acting on the basis of reasons. Our actions would not have the quality of self-determination.

Achieving unity, we can say, is an inherent need of self-determination. In every act of self-determination, there is a virtual unity of reasons involved.[106] That, it seems, is in principle all that is involved, and can be reasonably claimed, with regard to the alleged unity and systematicity of the domain of reasons. The perfect unity that Skorupski claims for the domain of reasons can thus be explained as inherent in cognition and is not incompatible with the claim that the empirically, or third-personally, observable building blocks of our cognition as such are forming open networks, which need not be coherent in all respects. The fact that those schema networks are nevertheless coherent to a high extent and show loose tendencies of systematicity can be explained by the fact that the schema networks we use for determining our normative landscapes must function in greater societies and over a long time. Schema networks might therefore approximate unity. Moreover, it might be only due to these collective networks approximating unity that individuals can achieve instantaneous unity in their reasoning. Impairments or even fragmentations of reason are common experiences for people living between differing normative landscapes.

Successful self-determination depends on the possibility to apperceive facts first-personally as standing in unified contexts, spatiotemporally, causally and normatively. Moreover, these unified contexts are not just there. They must be actively achieved, created if you want, by mind. An individual mind must rely on a communicative structure. An individual is irritated, and experiences a pull to re-evaluation and adjustment, if its own normative judgment is not shared or received as expected.

---

[106] In so far as we conceive of ourselves as temporally extended subjects, we of course have a need to conceive of the normative landscape as stable over time. However, for *being* a subject, being it for one instance is enough. It is conceivable that we change our very way of being a subject, become different subjects, over time. A change of the normative landscape would, of course, come with a break in subjectivity, eventually it would slowly give rise to a new subject. The subject as a temporally extended entity is, as Parfit (1984) suggests, a vague notion. As opposed to that notion, it might be better to define a subject as an instantaneous form of consciousness (see also Shoemaker 1996; Brinck 1997).

Third-personally observed, all that there is, all that the mind is doing, is activating acquired schemata. Each instance of considering a fact as a reason, weighing a consideration normatively and understanding it as favouring a particular response is an activation of a schema that relates persons, actions and situations. Individual mind is constituted by these kinds of structures, it does not of its own make them. The graspability of a reason depends on the reason being part of a communicative tradition if we define graspability according to the concept of grasping as "grasping together", as subjective understanding within a unified field that gives us the power of self-determination. In every moment in history, for every natural person, there is a finite totality of reasons that can be grasped *in this particular way*. This totality, I propose in the next subsection, can be conceptualized as an archive.

4.4.4. The solitary mind: moving in an archive?

As established so far, we understand ourselves as self-determined in a spatiotemporally, causally and normatively unified framework by responding to reasons. This first-personal experience can third-personally be described as locating oneself in a social realm – a realm we create by making sense of the world through collective fictions. Skorupski, as mentioned before, speaks of establishing warrant as a "collective, archival process" (2010: 124). My own account holds that this process also defines the ontology of reasons, that the realm of reasons *is* a collective archive, a mind-independent fund accessible to all the minds within a community.

But is it really the case that individual mind is determined and forced by tradition in this way? Isn't mind typically more flexible, if not in inventing reasons, then at least in weighing reasons? In what sense is our common fund of reasons really an *archive* rather than just a historical accumulation of cultural commons from which each of us draws creatively, in different ways and to different extents? Archives are typically more than just loose accumulations or natural historical sedimentations. Typically, archives are *intended storages* of records, arranged by *designated people*. An archive is usually *thematically defined* and clearly *separated* from other archives. Moreover, archives are explicitly *temporally indexed*. They do not just *have* history but are inherently *backward-looking* in their organization.[107]

[107] I take these characterisations literally from Johan Brännmark's presentation of my work as an opponent in my final seminar. I think this collection of features describes very well what the picture of an archive stands for.

In this subsection, I will defend the adequacy of the "archive"-metaphor, while still acknowledging the valid objections to the implied view on cognition. To begin with, I agree that cognition broadly conceived, including all the capacities of our brains, is not determined by history in that strong and exhaustive way. However, I hold that the specific self-understanding of being self-determined in a unified and coherent intelligible world takes place in an archive. Especially the activity of the solitary mind in performing what Skorupski calls "self-audit" (2010: 109f, 121f) is, I think, best understood as a motion on the grounds of a collective archive. Self-audit as Skorupski defines it is first-personal scrutiny of one's judgments with the aim to establish warrant. Warranted reason judgments are not in need of further questioning. They form the basis for autonomy[108] and are extremely difficult to achieve. Nevertheless, it must in principle be possible to gain first-personal access to all warrants, and also to determine when warrant is achieved entirely with the means of spontaneity and convergence. All self-determination, according to the relevant definition, aims at autonomy, and thus at warranted reason beliefs.

Let us now have a look at what this specific picture of reflection presupposes. How would we describe what is the case when a subject has achieved warrant from what I called the third-personal or scientific perspective? If all there is are different *forms of cognition* that create an organism-specific access to the world, it means that the warranted subject can only be described as applying its cognition correctly (and as having all the relevant information of course). Applying one's cognition correctly, however, can only mean conforming to the standard of a collectively shared *form of cognition*, given that there are no radically mind-independent structures that every cognitive creature is supposed to grasp. In section 3.4., I rejected the latter possibility.

Nevertheless, if Skorupski's picture of self-determination is adequate, the individual thinker needs mind-independent and clearly determined truths that are not up to her. But such truths can only hold in relation to an individual with a particular form of mind in a particular moment. What the form of mind in question is must be clearly determined. It must not be up to the subject who is defined by its form of mind. The only possibility to account for such an independent determination of the form of mind is to assume a capacity to rigidify, to clearly define the potentials of a form of mind at least for an

[108] Note that Skorupski distinguishes between Kantian and Hegelian autonomy, and also between "warrant" and "justification" (2010: 124f). In section 3.4., I argue that Kantian autonomy ultimately collapses into Hegelian autonomy unless we assume the radical version of mind-independence that runs into a Darwinian Dilemma.

instance of thought. Otherwise, the described kind of thought would not be possible. Clearly defining the potentials of a form of mind involves furthermore to clearly define from which point on a mind belongs to, or exhibits, a different form of mind.

We can acknowledge that minds change through history and through the course of evolution. This makes it, from a scientific perspective, difficult to decide what counts as a potential of our mind. How different from the actual world can the possible world, in which I have realized the potential, be in order for me to be entitled to say that I have the potential? Taking a detached scientific stance, I think there is no definite answer. However, the first-personal reasoner, for engaging in serious normative self-audit, needs a clear understanding of the limits of her mind. She must exclude the possibility that she could have a fundamentally different mind. She must understand the normative domain as finite and clearly defined – at least for every moment of apperception and self-determination to succeed.

As suggested above, mind actively creates the unity of the normative domain by predictive coding. However, the mechanism only works if expectations regarding convergence with others are sufficiently often fulfilled or can be updated easily by going back with some new information. For this mechanism to succeed, we need tradition. Of course, it is possible that an individual instantiates a mental state for the first time. It is even possible that the expression of it inspires collective re-evaluation, opening up a chance for social change. The next chapter will shed a light on that possibility. However, it is unlikely that this mental state, when instantiated for the first time, bears the characteristics of a reason response – genuine understanding and apperception in a unified field. Even if the individual in question exhibits the phenomenology typical for reason responses, it would ultimately be indistinguishable from an instance of delusion or madness. Whether madness can sometimes play a transformative role, when a community receives it well, is a question that will also be investigated in the next chapter. *For* the individual in its current situation, however, there is no difference between mere madness and madness that turns out as playing a positive role in virtue of factors beyond the individual's own understanding. To conclude, the reason relations, in order to enable warrant on the basis of both spontaneity and convergence, must form a rigid and unambiguously structured totality – a totality we can only understand as delineated and rigidified by tradition.

Now, we might simply conclude that this picture of cognition is inadequate and demand a picture that puts more emphasis on openness and creativity. As mentioned before, I do not reject such a picture but stress that a rigidified totality of reasons is involved in self-determination as Skorupski describes it.

So, should we reject this picture of self-determination? This, I think, is a legitimate question, although I doubt that it has a true answer. Self-determination as Skorupski describes it is a strong and influential idea. Many philosophers, indeed many human beings taking normativity seriously, understand themselves in this way, as far as I can see. Not only people taking normativity seriously engage in self-audit and understand themselves as self-determined in the proposed way. It seems obvious to me that cognition as self-determined thought has personal, intellectual and even aesthetic value for many. For many, the exercise of this cognitive capacity contributes to mental health and success in life. It seems first-personally obvious that the exercise of self-audit can be more or less successful, more or less stable and that it is worthwhile to try to find out what reasons one really has.

Now, I have suggested by reference to several scientific theories that all reasons we can hope to discover in this particular way, with this particular quality, are items of collective tradition. Moreover, the very possibility of grasping them in the right way, as full-fledged self-determiners, requires the rigidification, the closure of this tradition. What happens to human creativity in this picture? To the openness of human development that is not only obvious, but that we also tend to appreciate and try to foster?

I think, this value can indeed stand in tension with the ideal of self-determination, with the ideal of the solitary mind engaging in self-audit. Our collective funds are dynamic in nature – they are fluent, they can gain new items all the time, the funds of one community can overlap with the funds of another community, so that, for any individual, there is not really one clearly defined community of mind. When a reflective reasoner, however, grasps reasons as reasons in the full-fledged sense, she accesses not just *a loosely related fund*. By contrast, what she immediately accesses in this very instance is better described as an *archive – closed*, *arranged*, *cultivated*; perfectly ordered and unchangeable in the instance of apperception; containing a shared heritage to which others can refer in the same way; a collective archive in which mind moves; an archive which enables mind to move in this way.

My book neither destroys nor propagates the value of moving in that way. It merely argues that the activity of self-determined thought, constituted by reason responses, depends on the instantaneous closure of a collective fund. Thereby, it might make us aware of the implications and dangers of this activity. An astonishing insight might reveal to us when we arrive at the conclusion that the realm of reasons is a collective archive, that what we took to be a genuinely self-determined judgment, a judgment we arrived at with the power of our own reasoning, is nothing but compliance to the structures of an archived tradition. Still, it does not follow from this insight that our reason

beliefs are false or illusionary. Nevertheless, we can benefit from explicitly acknowledging the origin and genealogy of our reasons. Finally, we can deliberately construct the fund of the normative domain as an "archive". By doing so, there are two distinct paths opening up for us: First, we can understand the need to *actively cultivate* our traditions, to *actively arrange* them as an archive, if we want to stabilize the conditions for successful self-determination. We might do so by, for example, establishing, cultivating and teaching a *canon of art and literature* or by practicing *religious rituals*. We might strengthen these institutions by *designating specific people* for guarding, spreading, and cultivating them.[109] We might further support this process by *dedicating specific buildings or places* to this task. But second, we might also recognize the *back-ward looking* and restrictive character of these institutions. We might want to put the elements we inherited into their place and promote their *opening* and *transformation* instead of their preservation and guarding. The Collective Archive View's position on self-determination shows what is at stake when we choose that path and what it might eventually involve for individual subjects. A step towards transformation, as I will further develop in chapter 5, might require departure from the ideal of self-determination. Self-determined thought, we could say, is *frictionless motion in a collective archive of mind*. Transforming the archival tradition, by contrast, requires friction. Such frictions can come up between individuals and their society, but also between different parts of a society. Transforming tradition, finally, might require the destruction of the stability and harmony of the solitary mind. It might require the destruction of the basis on which the solitary mind can understand itself as aiming at autonomy out of its own intellectual power, the basis on which the solitary mind can inhabit a unified and unambiguous world of meaning.

[109]Particularly interesting cases can arise when a community has explicitly constructed tradition as an archive and guards it, while some parts of the community make use of the tradition in completely different ways. Johan Brännmark pointed out an interesting case in my final seminar: Mitt Romney reasoning about the Trump impeachment, thereby being firmly grounded in American tradition, while being out of line with others using that tradition. I do not have a definite opinion about this case, and there is unfortunately not much space for analysing the case in detail. What it illustrates, nevertheless, is the impossibility to decide in a universally valid way whether guarding or reinterpreting a tradition is better.

4.5. The Collective Archive View: a social ontology of reasons?

In this chapter, I developed what we can label the Collective Archive View – the view that reasons are items in a collective archive of mind. When we reflect in terms of reasons, when we recognize reasons, our mind accesses a storage of considerations that was developed and cultivated in a communal tradition of thought. What we experience as the process of free thought, the self-determined activity of our mind, is dependent on a shared tradition. When we engage in free thought, however, we respond to reasons as fundamental, as irreducibly normative. We respond spontaneously and we need no further explanation, once we have understood that something is a reason.

Thereby, the Collective Archive View unites two theses: The thesis of metaethical Reasons Fundamentalism, according to which reasoners must simply understand what it is to be a reason and cannot derive normative force from something non-normative; and an ontological thesis that explains what it is to be a reason in social terms. As defended in subsection 1.4.3., metaethics narrowly conceived, an analysis of the concept of a reason, is not sufficient for giving a full-fledged account of normativity but needs what I call a material analysis of what we are doing when we respond to reasons.

Moreover, Reasons Fundamentalism, as argued in 2.4.1., is a weak thesis that is compatible with a variety of ontological positions. The proposed ontology in this chapter is a social ontology. Reasons, according to the Collective Archive View, are socially constructed facts. This last section specifies what this means and compares the proposal to an influential social constructionist position that is mainly concerned with categories such as sex, gender or race.

Subsection 4.5.1. introduces the distinction between being socially constructed in the *causal* sense and being socially constructed in the *constitutive* sense. I will argue that while reasons are at least in part socially constructed in the causal sense, they are definitely so constructed in the constitutive sense. Subsection 4.5.2. brings up Haslanger's proposal of social constructionism as "critical realism". I will show what the analysis of normativity can gain from this ontological proposal but also point out why social constructionism in the specific case of the normative domain presents us with challenges that social constructionism about other entities lacks.

4.5.1. Socially constructed facts: causal and constitutive construction

The field of social ontology offers elaborated theoretical work that can be used to understand the ontology of reasons suggested by the Collective Archive View. As Epstein (2018) notes, the term "social ontology" has become common to describe philosophical analysis of social reality – of, for example, the nature of groups, group agency, organizations, institutions such as money, marriage or the law, or categories such as gender, race or disability. Scholars who made the term famous introduced theories of social or institutional facts brought into existence by collective intentionality (e.g. Gilbert 1989; Pettit 1993; Searle 1995; 2010).[110] In more recent years, the idea of facts brought into existence by social creatures is fruitfully used in order to argue that certain parts of our reality are not inevitable (Haslanger 2012; Díaz-León 2015) but determined by factors that could be otherwise.[111]

It has become common to distinguish between what is *causally* socially constructed and what is *constitutively* socially constructed (Mallon 2019; Haslanger 2012: 86f; Díaz-León 2015: 1141). Haslanger gives the following definitions:

> *Causal construction*: Something is causally constructed iff social factors play a role in bringing it into existence or, to some substantial extent, in its being the way it is.
>
> *Constitutive construction*: Something is constitutively constructed iff in defining it we must make reference to social factors. (Haslanger 2012: 87)

In the case of "gender", a frequent example in the debate, we might find elements of both types of social construction. If labelling and treating people in a certain way causes them to develop certain traits, this is a matter of causal construction.[112] As opposed to that, many philosophers argue that gender is a social category "whose definition makes reference to a broad network of social

[110] I thank the audience of my talk at *ENSO V* (in Lund 2017), especially Åsa Burman, for feedback on an earlier version of this work, in which I built on Searle's concept of "collective intentionality", which I later found less helpful in accounting for the creation of reasons. See also the criticism of intentionality-focused social ontology by Epstein (2015).

[111] I thank the audience of my talk at *Social Ontology 2020* for feedback on the idea of reasons as socially constructed facts. Furthermore, I thank the social metaphysics group in Barcelona, especially Esa Díaz-León and Dan López de Sa.

[112] See also research about what is known as the "looping effect" (Hacking 1995).

relations, and it is not simply a matter of anatomical differences." (Haslanger 2012: 86f). This means that, for example, "being a woman" is defined as standing in particular relations to others, as bearing a particular social meaning or as having a particular social status, as Àsta (2018) formulates it.

I think that both senses of social construction are applicable to reasons according to the Collective Archive View. However, they must be sharply distinguished and can, potentially, stand in a tension, which gives rise to interesting questions – and, moreover, helps to understand frictions or disruptions of thought as described in the first chapter.

So, first, what would it mean for reasons to be causally socially constructed? According to Haslanger's definition, this would be the case if social factors were relevant for bringing reasons into existence and making them the way they are. As argued earlier, a reason relation can only "exist", or better "hold", if mind has a potential to respond to it. Following the evolutionary theory by Mercier and Sperber (2017), what mind does in reasoning is trying to justify itself in a social context. The attempt to justify is targeted at a specific group and thus to a large extent determined by what this group collectively accepts. But does this make the group the cause of those reasons? In other words, is the group the cause of the facts that an individual person can see as reasons? Does it bring these facts into existence? Does it determine the way they are?

To assess these questions, remember first that the relevant facts are not entities in the physical world, but communicable conceptualizations of states of affairs in the physical world – propositions, or senses in Fregean terms. If the presented evolutionary and neuroscientific story about our capacity to form such specific "higher-order" conceptualizations is correct, it is in a trivial sense true that the potential of our mind to deal with reasons is caused by a social practice. But does the theory allow us to say that we are responsive to these reasons and not others as a product of the specific social practice which gave rise to our potentials? This seems a reasonable expectation of a distinctive social constructionist position about reasons. I think we can defend a moderate version of this position. Accounts such as the one by Sperber and Mercier or by various schema theorists do not rule out that some reasons will inevitably be treated as reasons because of basic biological or anthropological constituents. This does not mean that such constituents, such as for example a drive to survive or to have offspring, are themselves reasons. A fact is only a reason, as the term has been developed in this book, if we have formed a conceptualization of it that we use in a specific form of communication and reflection. However, anthropological factors may secure that certain conceptualizations will always play a role in the relevant form of communication and reflection. As Mercier and Sperber note, a community

cannot create a "battery of reasons of its own" (2017: 143) but is constrained by "cognitive efficiency" (ibid.: 144). While some reasons might indeed be the local products of a specific community, others might not be caused by social factors in an interesting sense. They may be shaped to an extent but not be directly and exclusively caused by the particular social group in which they function as currency for justification. This brings me to the conclusion that reasons are in part, or to an extent, but not entirely socially constructed in the causal sense.

Now, second, what would it mean for reasons to be socially constructed in the constitutive sense? Being a reason, according to this position, would be a "social status" – a property that nothing in itself *has* but that we *confer* on certain things by ascribing it.[113] In other words, "being a reason" is "being considered or accepted as a reason". While the causes for people's acceptance of something as a reason may be a hybrid of social and non-social factors, something, however, can positively be a reason if and only if a group of people accepts it as a reason. The obvious question at this point is certainly "Which group of people is relevant if groups differ?" Before turning to this tricky question, let us assume the simple case in which a person lives in a normatively homogenous community. What if this person comes up with a reason which this group does not accept? For the individual person, the problem can be on two levels. First, the person might have psychological difficulties to seriously come up with this reason to begin with, since our very own private judgment mirrors the social practice and is influenced by our expectations about the social practice – this is, our expectation about whether we will find convergence with others. As an example, a homosexual can have difficulties to fully believe that she has reason for gay marriage if it can be expected that everybody denies it. Now suppose the person does come up with such a reason, nevertheless. As to be expected, she fails to justify herself to others. Following the epistemological twin pillars of spontaneity and convergence, a lack of convergence always challenges the individual's spontaneous judgment. Either the judgment is given up or the lack of convergence can be explained with good reasons. Alternatively, the self-determined process of thought breaks down.

Suppose that the person in question maintains her judgment despite of the lack of convergence. Which sources of reasons to explain the lack of convergence could she have? There are only two possibilities: Either she figures them out by herself or she "borrows" them from minority groups that she knows in person or that she might also be acquainted with from books or

[113] See Ásta's (2018) conferralism about social properties.

other forms of human testimony. In real life, the latter is probably what is most often the case when individuals go against the mainstream in their culture without breaking down. I would say that, in this case, the dissident in one community just relates to another community in which she finds convergence. How small or remote that community can be in order to still give support for a successful reasoning process probably depends on individual psychological factors such as self-confidence, courage or tolerance for emotional conflict and isolation and, most certainly, also on how strong the surrounding community sanctions the dissident, apart from denying her reason claims. Still, we can maintain also for this case that a person can only understand as a reason what is accepted as a reason by a community.

Now, let's finally turn to the most extreme case in which a person holds a reason judgment despite of a lack of convergence and a total lack of any external source for reasons that justify her persistence. I think that there is nothing that distinguishes this person from a mad person. The person being capable of such an inflation of one pillar of normative epistemology, while at the same time completely denying and ignoring the other pillar, is spinning in the solipsistic universe of her own mind. Even if this person exhibits the subjective phenomenology of self-determined thought, I would, by definition, not call this an instance of self-determined thinking.[114]

A final point should be noted: I allow for the possibility that a community that comes later in time might develop an understanding of this lonely dissident's activity, might be able to see justifications for her decisions and judgements. Indeed, we might even consider the possibility that the observation of this very person inspired the social changes that brought about this new communal understanding. In the last chapter, I will introduce this as the possibility of "retrospective justification" of the initially unjustified. However, I would like to insist that, in the moment of being made, the judgement was unjustified. There were no reasons that could justify it. We simply lack the theoretical means to call something a reason for somebody if this person cannot possibly arrive at its recognition by means of proper use of the inherently connected twin pillars of reason epistemology – spontaneity and convergence. Thus, I conclude that something is a reason if and only if it is accepted as a reason by a community. Reasons, accordingly, are socially constructed in the constitutive sense.

[114] Also feminist philosophers such as Calhoun (2015) share my intuition that people who constantly ignore a substantial divergence, when they cannot resolve it, are problematic from a genuinely moral perspective – even though Calhoun holds that people in this situation *can* get it right, which I doubt (at least people in this situation cannot get it right in the proper sense).

To sum up, the fundamental items of our thought have a social ontology – they are constitutively, and in part also causally socially constructed. Given that we must accept reasons as fundamental when we reason, while at the same time we can infer from the proposed investigations that they are socially constructed, and from that perspective not fundamental, we end up with a tension when analysing reasons. Which is the relevant community that must accept our reasons? Are there reasons that make some communities appropriate referents while others are less appropriate?

These questions may naturally suggest themselves, but I think they ultimately rest on a misunderstanding. At this point, the distinction between two fundamentally different but ultimately compatible perspectives on reasons becomes relevant. When asking such questions, we must realize that we take the first-personal stance of normatively interested reasoners. As such, we do not directly think about communities at all – we only think carefully about reasons and try to use our reasoning capacity in the best possible way. If we do so, however, we move in what I call a collective archive of mind – we depend on socially constructed facts. There are no reasons, and accordingly no normative truth, from the perspective from which we describe our activity in that way. Accordingly, there is no better or worse community beyond the standpoint of a communally embedded first-personal reasoner. I take this to be one of the core insights of the Collective Archive View. The last chapter will expand a little bit more on these points.

4.5.2. Haslanger's Critical Realism applied to the domain of reasons

In the previous subsection, I argued that reasons are socially constructed facts. Accordingly, the collective archive in which we move when we reason is a socially constructed reality. Haslanger examines social constructionism as a "debunking project" (2012: 113-138) and defends an ontological position about socially constructed facts that she calls "critical realism" (2012: 183-218). In this subsection, I want to apply Haslanger's ontological proposal to the domain of reasons.

Haslanger's outline of social constructionism as a debunking project adds to a list about different kinds of social constructionist positions, different kinds of motivations that can drive them, that has been formulated by Hacking (1999). Hacking distinguishes the historical, the ironic, the reformist, the unmasking, the rebellious and the revolutionary constructionist (1999: 19-20). Their projects all embrace 1.) "X need not have existed, or need not be at all as it is. X or X as it is at present, is not determined by the nature of things; it is

not inevitable." (Hacking 1999: 6) To varying degrees, they also embrace 2.) "X is quite bad as it is" and 3.) "We would be much better off if X were done away with, or at least radically transformed." (ibid.). Haslanger's "debunking project", by contrast, is more modest regarding normative and practical conclusions as exemplified by 2.) and 3.).

As opposed to constructionists with for example reformist, rebellious or revolutionary intentions, all the debunking constructionist wants to say in the first place is that there is a "theoretically important" kind that has "not been adequately acknowledged to be social" (Haslanger 2012: 137). As Haslanger describes, debunking constructionists "may seem to be offering radical and implausible `analyses´ of our ordinary concepts, in fact they can be better understood as working within a semantic externalist model that looks to social theory" (ibid.). Thus, debunking constructionists bring something to mind that is hidden in our ordinary talk, thought and practice.[115] In other words, they make us better understand what we are doing when we use certain concepts or categories.

It is important to note that, in this context, "debunking" does not mean "destroying" or "annihilating", it means showing with sober analysis how something is build that we usually take for granted and that plays a fundamental, usually unquestioned role in our lives – gender in Haslanger's case, reasons in my case. I think that the ontological status of both kinds is analogous to a large extent. Haslanger also calls her social constructionist position "critical realism" (2012: 197). Thereby, she refuses the idea that social constructionism equates anti-realism or anti-naturalism. The point of saying that something is socially constructed is not to say that it is not real, or that there is no objective knowledge about it that could be good to have (ibid.: 198). Instead, social constructionism specifies the structure and creation of a particular part of reality. Moreover, saying that something is socially constructed is not to say that it cannot be a natural kind. Instead, Haslanger emphasizes that the social is part of our nature and that nature consists of interdependent social and non-social factors (ibid.: 212f). Finally, Haslanger describes the point of social constructionism as follows:

> if we are going to change the world, we need to know how the problematic parts are created and maintained. We need to find the *levers for change*. And if the goal is to find the levers for change, then it is important to understand ways in which the social and non-social are interdependent. (2012: 215)

[115] See also the term "covert construction" (Mallon 2019).

We might be able to transport some of this spirit into the analysis of reasons, but we also encounter a problem at this point – a problem that occurs for social constructionism about reasons, but not for a social constructionism about gender or race. If we know how our gender categories are socially constructed, we can think about reasons for or against certain interventions or changes. When we know that our reasons are socially constructed, by contrast, how shall we decide to go on? How shall we evaluate the status quo if not in terms of reasons?

In some cases, showing the history of our normative beliefs might just make us see that these beliefs are not warranted. However, the Collective Archive View is more radical than that. It holds that all reason relations are parts of a socially constructed fund – also those which we would only discover if we had full knowledge about the world and our history. In other words, an omniscient computation machine would have no reasons at all unless it is a social creature that takes part in a specific communicative practice, in which all, even the most basic, reason beliefs originate – even the belief that life is worthwhile, or that we have reason to protect life. Without this communicative dimension, I take it, an organism could be driven to maintain itself, but not believe that it has reason to be alive. We are, in fact, given a reason to be alive only by a community. This means that all normative conclusions we could possibly draw from the Collective Archive View are themselves based on a social construct. It means, in other words, that we cannot reflectively escape from the social constructs that constitute our thought.

This, it seems, is a peculiarity that distinguishes social constructionism about reasons from social constructionism about other entities. Nevertheless, I think, that the analogies outweigh the differences and that the distinction is, after all, not that deep. Consider, concerning the domain of gender, the psychological restrictions to thinking outside of deeply entrenched gender roles, that were, for example, pointed out by theorists such as Butler (1990). The socially constructed is, if we want it or not, a reality that constrains our minds. Then consider, concerning the domain of reasons, the psychological flexibility with which we can relate to our beliefs about what we have reason to do. We can pursue self-determination in terms of reasons, while we can also pursue, or at least allow for, states of mind that are less than self-determined thought but still bring about changes in the world and in ourselves that we can value. There is an intriguing question that opens up when we thoroughly think about the implications of the Collective Archive View: Can we act without or against reason and thereby bring about states for which we later start to see reasons? New states of being a subject? New "forms of mind", as it was called in chapter 3? New forms of mind that are responsive to new reasons? To

reasons that were not available, and therefore could not have been said to hold, for our earlier selves?

I will dedicate the last chapter to examining this possibility. Understood along the lines of Haslanger's critical realism, the Collective Archive View can claim its own stance in metaethics - between simple realism, normative constructivism and error theory. In the beginning of the last chapter, I will take up these positions again and summarize my responses to each of them. The Collective Archive View, to conclude here, is an attempt to lead metaethics into social ontology, or to introduce a social constructionist position into metaethics.

4.6. Conclusion

This chapter has reconstructed the psychological activity of responding to reasons, as it is described in Skorupski's philosophical account, from a perspective that views our psychological activities as natural processes – as processes that can be described scientifically, that have an evolutionary history, a neurophysiological structure, and can be explained with frameworks used in psychology and the cognitive sciences. The difference between the philosophical and the scientific approach, as I use the terms in this context, is that the first analyses reasons from the first-personal perspective of normatively engaged reasoners. It asks for the conceptualisation of reasons that is presupposed in our reflective activity and for our relation to reasons during this activity. The scientific approach, by contrast, brackets normative questions and views the same activity in a detached way, as a third-personally observable process. To be sure, we often enough view ourselves and others third-personally when reasoning about normative questions in everyday life. However, the "third-personal" or scientific stance, as I defined it, detaches further than that. It asks what a creature responding to reasons is doing without identifying with this kind of creature.

What a creature achieving self-determined thought by responding to reasons is doing, according to the analysis offered in this chapter, is moving successfully in a collective archive of mind. The collective archive is a shared fund of considerations that count as reasons – a fund that has been established in a communicative process and must receive effective unification and closure for self-determination to succeed. We can arrive at this reconfiguration of the domain of reasons by looking for a scientific explanation of the complex unity of apperception, which is characteristic of responding to reasons – the complex

phenomenon of understanding oneself as self-determined in a causally, spatiotemporally and normatively unified framework holding independently of us. We can call such a unified framework a "world" – a world as an intelligible system which we intellectually penetrate. When we grasp a reason, we grasp a feature of a world as meaningfully standing within that world, which endows us with the feeling of intellectual power and individual independence.

As presented in section 4.2., neuroscientists have developed an approach to describe the mind-world relation coming with this kind of access to a unified world. That approach, called "predictive coding" or "predictive processing", receives growing interest from scientifically oriented philosophers of mind. It can be fruitfully combined with insights about the genuinely social development of the specific unity of apperception that comes with self-determination in terms of reasons. The conceptual framework of ecological psychology, presented in section 4.3., offers further help in understanding reasons as parts of our specifically human environment without being entities in the physical world – parts which are nevertheless rigid and independent of our individual mind. Parts we can "grasp" in a sense that is, after all, not that different from grasping as a sensorimotor process – a process that gives us active world access, orientation and meaning. Again, we can rely on many previous interactions between philosophy of mind and the empirical sciences for justifying and strengthening this approach. Section 4.4. has expanded on the genealogy, the creation or establishment, of reason relations in a collective process by drawing from "schema theory" in cognitive psychology. It has identified reasons as collective fictions that serve our self-understanding in a meaningful world. Finally, it has motivated the description of the domain of reasons as an archive – a tradition that is actively given closure, unity and systematicity.

Section 4.5. has argued that reasons, being items in a collective archive of mind, can be considered as socially constructed facts. After specifying the notion of social construction, I have proposed social constructionism as critical realism, following Haslanger, in order to frame the ontology of reasons. Critical realism acknowledges the reality of what is socially constructed and allows for the socially constructed to be a fundamental part of our nature. However, critical realism emphasizes the changeability of reality and aims at showing what a change of reality would require and involve.

If reasons are socially constructed facts, forming a collective archive in which our solitary mind can gain autonomy and exert the power of self-determined thought, this brings a fascinating twist into our self-understanding: We may find it striking that what we see as our intellectual power and independence is in fact the upshot of behaving correctly in a socially created

environment. Furthermore, we may understand that normative change – genuine normative change – cannot be based on a fully self-determined decision, warranted by reasons. Indeed, active change might require giving up the ideal of self-determined thought, giving up the high philosophical aspiration of the solitary thinker, the aspiration to achieve normative warrant. At least, normative change requires risking the stability of self-determined thought, making the achievement of a new stability dependent on others who must change with us.

Chapter 5
Losing one's mind
Madness, subversion and normative transformation

Everyone who has hitherto overthrown a law of established morality has always at first been considered as a wicked man: but when it was afterwards found impossible to re–establish the law, and people gradually became accustomed to the change, the epithet was changed by slow degrees. History deals almost exclusively with these wicked men, who later on came to be recognised as good men.

(Nietzsche, The Dawn of Day I, 20.)

5.1. Introduction

This chapter brings the implications and theoretical advantages of the Collective Archive View to the fore. The most distinctive advantage is the possibility to conceive of genuine normative change or transformation; a transformative process the possibility of which stands in a tension with the ideal of self-determined thought; a transformative process that either happens very slowly and unnoticed, or that we can choose to embark for. When we choose to do so, however, the outcome is beyond our control. More severely even, we are in danger of losing our mind. The sanity of our mind, I will point out, can be described as having ownership of our thoughts – a status that has subjective as well as intersubjective components.

By providing these resources, the Collective Archive View can make sense of what the classical metaethical positions discussed in chapter 1 cannot account for in a convincing and satisfying way: the experience of disruptions of thought. Experiences that reveal the limits of our thinking and that suggest that reasons are not simply there for every intelligent brain to grasp. This final

chapter goes back to situations of disrupted thought, or disrupted reason, and analyses them as opening a possibility for transformation.

Section 5.2. relates the Collective Archive View to the metaethical positions discussed in the beginning. Thereby, it shows how the combination of what I called a metaethical approach narrowly conceived, a first-personal analysis of the concept of a reason, and what I called a material approach, a third-personal empirical analysis of what we are doing when we use that concept, contributes to our understanding of reasons. It shows how the comprehensive picture resulting from the integration of two in themselves neatly distinguished approaches is better than the discussed metaethical positions – better in terms of conceptual fruitfulness, practical adequacy and clarity. The more comprehensive picture captures the most convincing intuition in each metaethical position, while avoiding their flaws. In the end, a full understanding of the picture raises the question whether we should do what we have reason to do.

Section 5.3. discusses the costs and gains of doing what one has reason to do. It suggests the term "ownership of thought" and discusses two different forms of how people could react to a loss of ownership. They could either disintegrate as subjects, as Simone de Beauvoir in the case discussed in chapter 1, or they can inflate their own subjective responses and discount the essential twin pillar of subjective spontaneity, convergence with others.

Finally, section 5.4. presents the idea of transformation and offers a possibility to quasi-rationalize what we cannot really rationalize: giving up the ownership of our thought for the exploration of better ways of thinking. I present an example of normative pioneering that comes at the cost of a healthy mind but can nevertheless be worth it. If the experiment is socially unsuccessful, the person will simply end up in madness. If it is successful, however, we might consider the act of pioneering as retrospectively justified – as given a reason after the fact that the person at the time could not own.[116]

5.2. The Collective Archive View between realism, constructivism and error theory

In the first chapter, I discussed three different ways in which metaethics can account for our belief that there is objective normative truth. Realism,

[116] I thank Björn Petersson, Frits Gåvertsson, Nora Hämäläinen, Dan Egonsson and Cathrine Felix, among others, for encouraging my interest in social change.

exemplified by Scanlon (2014), says that there are irreducibly normative facts. Constructivism, as exemplified by Korsgaard (1996), in an alternative version defended by Street (2008a), says that objective truth is a function or construct of something internal to us, desires or rational principles. Error theory, exemplified by Mackie (1977), says that there is nothing that makes unconditional normative judgments true and that all our unconditional normative judgments are therefore false. The Collective Archive View shares a basic intuition with each of these theories but avoids their flaws or shortcomings. The strength of the view is mainly due to its sharp distinction between the conceptual and the material approach. After discussing the Collective Archive View in comparison to each of the mentioned metaethical positions in the subsections 5.2.1. to 5.2.3., I will point out a question raised by the approach. The possibility of this question, formulated in subsection 5.2.4., indicates that, although reasons are fundamental to self-determined thought, we can distance from them by giving up self-determined thought.

5.2.1. Reasons Fundamentalism beyond simple realism

The Collective Archive View shares several commitments with realism as it is defended by Scanlon (2014). Most importantly, it shares Reasons Fundamentalism – the view that the concept of a reason must be presupposed. It is non-analysable and cannot be explained further in non-normative terms. What, then, is the thesis that reasons are items in a collective archive of mind if not an analysis or explanation of what it is to be a reason? This is where the distinction between the conceptual and the material approach, postulated in chapter 1, shows its merits. In chapter 2, I argued that Reasons Fundamentalism is only a narrow conceptual thesis, which does not amount to a strong ontological thesis. Accordingly, the thesis of Reasons Fundamentalism is compatible with a social ontology. Let me now justify the idea that a concept can be fundamental, while at the same time being socially constructed, and defend it against a few objections.

The first obvious objection is, as already mentioned, that a social constructionist theory of reasons is an analysis of reasons and that, therefore, reasons cannot be fundamental. Of course, conceptualizing reasons as items in a collective archive of mind is to give and explain a concept as well. However, it is a way of conceptualizing something that we can observe and study as a phenomenon in space and time, a phenomenon in the natural world, described by the empirical sciences. The philosophical concept of a reason, by contrast, is the concept that we use when reflecting in terms of reasons. Now should we conclude that there are simply two distinct concepts of a “reason” that classical

metaethicists and scientifically oriented theorists respectively work with? Maybe we could say so, but we should keep in mind, if we say so, that these two concepts are not referring to two separate ontological kinds. As I strengthened especially in chapter 3, philosophers defending Reasons Fundamentalism and psychologists examining reasons empirically are talking about one and the same world – a world in which natural creatures are responsive to reasons. They, however, talk about these creatures from a different standpoint, with a different interest. From the scientifically oriented standpoint, it does not matter whether we ourselves are the creatures that we examine. It might, of course, influence our theories indirectly, but it does not matter by definition. This book says nothing about any of these standpoints being superior. It only says that we can take them both and that we can gain interesting insights by taking each of them separately and bringing the insights together into a single picture. The Collective Archive View offers such a picture and states that reasons present themselves as irreducibly normative and not in need of further questioning – at least some of them – from the first-personal perspective of reasoners, while being recognizable as socially constructed when examined from the third-personal scientific perspective that was taken in chapter 4.

Scanlon, as pointed out in 2.4.1., jumps from Reasons Fundamentalism directly to realism. He employs the idea of different metaphysical domains in order to stress that not only physical entities are real. In principle, the Collective Archive View is compatible with that claim. As emphasized in 4.5.2., the proposed social constructionism can be read as a form of realism – agreeing that different ontological kinds can count as "real" and that reasons can be accepted as parts of our reality. However, social constructionism, as distinct from Scanlon's simple realism, can count as critical realism (see Haslanger 2012) – acknowledging something as real, but showing how it was created and why it is the way it is; acknowledging, finally, that it could be different if we were different. It is not clear, but very likely, that Scanlon would agree with that. However, he is not interested in the possibility that we could be different subjects. Eventually, it seems, he is not even sufficiently aware that there could be genuinely different ways of being a subject, as opposed to subjects having different needs and preferences. As Skorupski, he seems to accept what section 3.4. criticized as Reason Monolithism, a scientifically inadequate view according to which "reason" is one distinctive faculty, alike in all creatures in the universe that develop it, rather than a variable way of cognitively processing external stimuli. It is a fair point to find reflections about being a different type of subject too deep, too remote, or simply irrelevant for a common-sense analysis. Nevertheless, I think it is a worthwhile

reflection for philosophers to engage in and sections 5.3. and 5.4. will discuss human predicaments in which this reflection may become existentially relevant. I think we should accept that, insofar as our first-personal thought is a natural phenomenon, what is fundamental to our thought can be explained further.

This brings us to a second obvious objection: What are we doing when assessing our thought scientifically? Aren't we thinking and responding to reasons as well? Aren't reasons, as traditional Reasons Fundamentalists hold, indeed more fundamental than all scientific explanations, because they must be presupposed for each of these explanations? To be sure, we also reflect in terms of reasons – reasons for scientific belief – when we try to explain phenomena scientifically. Nevertheless, I think we can make a distinction between engaging in all-things-considered normative reflection and using our reflection in other ways. This is a distinction that Reasons Fundamentalists in contemporary metaethics have dedicated themselves to extinguish. Some might hold that the very point of this movement in metaethics is to show that moral reasoning is after all not that different from other forms of reasoning – reasoning in science, mathematics or practical contexts other than morality. At this point, I think we can make a distinction. I agree with this movement in so far as each of these activities is governed by reasons. I further allow for the possibility – but do not take a qualified stance about it here – that all these reasons are items in a collective archive of mind. Studies of the development of scientific practice by researchers such as Dunbar (1996) or Nersessian (2010) suggest that. Kusch's (2002) relativist and communitarian conception of knowledge might represent such a position applicable also to the domain of science. However, I do not touch these debates. Instead, I hold that no matter where the reasons for our scientific beliefs come from, they are not automatically *all-things-considered normative beliefs*. We can distinguish the latter from all other applications of normativity. This makes the Collective Archive View compatible with but not dependent on a form of scientific realism. As distinct from normativity in the sciences, I think, we have a conception of what we might call "central normativity" – reasons to do or to believe something all-things-considered, or unconditionally. Some Kantians, such as Korsgaard, call this "morality". Since most people associate something narrower, or explicitly interpersonal, with that term, I prefer "central normativity" instead. Questions for what reasons you have, period.

In this sense, I think it is obvious that not everybody can have all-things-considered reasons to form particular scientific beliefs – or particular beliefs about what is the case in the world. Even though there might *be* such reasons, they might not be reasons for a particular person, with limited memory and

processing capacity, and reasons to prioritize the formation of other beliefs.[117] As biological organisms, we have limited processing capacities and limited world access. According to the approach presented in chapter 4, we have no world access at all if we cannot carve out a way of modelling and unifying the stimuli relevant for us. We need to carve out a finite totality of normatively relevant considerations – a finite totality of reasons that are really normative for us in the central sense. This carving out, I suggested, is done by help of collective tradition. Centrally normative reasons in this sense are provided by the collective archives of our mind. Although we can never get rid of all normativity at the same time, we can step out of the central normativity that we usually take as unconditionally guiding our lives. The central normativity, the standards that regulate what we are to do or to believe all-things-considered, unconditionally, can be bracketed. It can be examined how we make centrally normative judgments in this sense. The reasons for believing the Collective Archive View are probably not central – as are the reasons for believing most scientific theories. While people committed to or dependent on finding the truth in a particular, highly complex and specialized scientific discipline might have reasons to form the relevant beliefs, not every human agent can have reasons to do the same. My claim is that what *central reasons* an individual can have, can discover and judge herself to have, is determined by the collective archives.

There is a last objection to the approach, which I would like to address. When we reflect about what we have reason to do, we often shift between a first-personal and third-personal view on ourselves. Doesn't this mean that the third-personal view has a bearing on normative questions? My answer here is very brief. The third-personal view in this case, I would hold, is within the normatively interested and engaged standpoint. The third-personal stance from which the material analysis is made, by contrast, is more detached, more remote, bracketing all judgment and identification with the examined activity. Although we never in fact reach a god's-eye-view, we can take a detached view on ourselves. If we took *only* this view, all the time, we would not be able to arrive at any centrally normative judgment. Thus, the sphere of normative reasoning is limited, but can be subject to scrutiny that is not clearly limited. We could stop at any point or refuse to engage in that kind of scrutiny altogether.

[117] Even if we could make sense of the idea of an omniscient and perfect computation machine that could state and process everything that is the case in the universe – I am not sure if we can make sense of that – such a machine would have no reasons. It could stop processing. It could process incorrectly. We typically speak of correct processing, when the process can reach an inherent goal or maintain a particular form that is distinctive of it.

To sum up, I have argued that there is a domain of reasons that we can recognize as holding for us in the central, all-things-considered sense, a domain of reasons that we can recognize by help of spontaneity and convergence and that enables self-determination, the understanding of one's own actions as warranted by reasons within a causally, spatiotemporally and normatively unified world. This domain of reasons, I have further argued, is a collective archive of mind. As such, it is an important and fundamental part of reality. But it is also a socially constructed reality that could be otherwise. This makes the Collective Archive View, embracing critical realism in Haslanger's sense, more comprehensive and more interesting than realism simpliciter. It represents a critical view that goes beyond simple metaethical realism.

At this point, I can align with Skorupki, who also labels his view a "critical stance". Skorupski departs from realism to irrealism, because he wants a distinction between causally effective objects in the world and abstract objects. The Collective Archive View, by contrast, has resources, such as predictive coding and ecological psychology, to speak about reality in a way that includes more than physical objects, without embracing orthodox Fregeanism. In the end, however, it is just a matter of terminology whether we want to count socially constructed abstract entities as part of reality or as irreal structures. If we read Scanlon's realism and Skorupski's irrealism as similar in spirit, as Olson (2018) suggests, we can conclude that the Collective Archive View goes beyond both in taking the critical analysis further, and relativizing the scope of the fundamentalist claim radically.

5.2.2. Social constructionism vs. normative constructivism

The Collective Archive View, with its social constructionism about reasons, shares some intuitions with constructivism as it was introduced in the first chapter. The view is, however, different from constructivism in a crucial respect. Constructivism says that there is objective truth about what reasons we have, but that normative facts are not just there as parts of the world. The Collective Archive View is compatible with that so far. Now, constructivists try to account for the fact that reasons, in an important sense, depend on us, while not being completely up to us, by depicting them as the constructed results of something internal to us. According to Korsgaard's constructivism (1996), normative facts are what can be constructed according to the principles of practical reason. In Street's version of constructivism (2008a), normative facts are a function of our desires reflected on in light of each other.

In chapter 1, I arrived at the conclusion that constructivism is either a normative, rather than a metaethical position, or it is based on speculations that

ultimately depend on empirical claims. The Collective Archive View, by contrast, with its strict distinction between the first-personal and the third-personal perspective on normativity, can avoid normative assumptions, while assigning the right status to the other claims. The normative assumption in Street's constructivism is that we ought to best satisfy our desires, while the normative assumption in Korsgaard's constructivism is that we ought to follow the principles of practical reason, which are constitutive of being an agent, or constitutive of our practical identity. To be sure, there may be good arguments for each of these claims. However, these arguments are directed to normatively interested first-personal creatures. It is conceivable that somebody is responsive to a reason that cannot be traced back to either a desire or a presupposition for agency. Likewise, if somebody is responsive to constructivist arguments, this is a basic responsiveness that cannot be justified any further. By contrast, the basic responsiveness to some considerations as reasons must be presupposed. Some reasons can be derived from other reasons, but ultimately reasons are underived – irreducibly normative, non-derivable from non-normative facts such as desires or principles internal to our consciousness. This is the claim of Reasons Fundamentalism, which the Collective Archive View holds up against constructivism.

There is an objection to my reading of constructivism which I like to consider at this point. Defendants of either a desire- or a rationality-based version of constructivism might hold that what they point out as the basis of normativity is intrinsically normative, and that it is the only thing in the universe that is intrinsically normative. Desires, you might say, are intrinsically normative because they can prompt you to action without further justification. I think this idea rests on an unclarity in using the concept of a "desire", which the Collective Archive View avoids. As I showed, we must distinguish between two things. On the one hand, there are *causes that bring about behavioural outcomes*. If these causes can be conceived as sufficiently "internal" to an organism, a person with behaviouralist leanings may be inclined to call it a "desire". On the other hand, however, we *conceptualize* states in the world, internal or external to ourselves, *as desirable*. On the material level, these *conceptualizations*, as developed at length in chapter 4, correlate to *"higher-order" brain states* that we instantiate in a social context, as parts of a communicative practice. They are clearly distinct from whatever prompts or causes our behaviour. Since we often speak about desires and give desires as reasons, we should reserve the word "desire" for exactly this – for a *conceptualization*, for something we instantiate when we say "I desire…". While it may be trivially true that we can only do something if we have "internal causes" that prompt us to do it, it is far from obvious that we always

conceptualize our actions as following a desire. In some contexts, a "desire" is not the right conceptual element to function as a reason, even though everything that is a reason for you corresponds to a possibility to understand yourself as motivated by it. The Collective Archive View with its essential distinction avoids all conflations here.

There is a similar way to defend Korsgaard's position, discussed as constitutivist constructivism in chapter 1, against the accusation of making normative assumptions. This defence strengthens the fact that constitutivist constructivism argues for no particular reasons but only points out that all our reasons are related to our practical identity, and that we can find them out by reflecting about our identity in the right way. The Collective Archive View denies the latter and confirms the former in a more qualified way. I agree that being able to understand the reasons relating to one's situation strengthens something that we could call "practical identity". I called it "self-determination", or the state of being a "self-determined subject", thereby following Skorupski. However, I do not see why reflecting on this very phenomenon should always help us to assess reason claims. As opposed to Korsgaard's rather speculative position about practical identity and self-constitution, the Collective Archive View develops an empirically founded explanation for why we can indeed say that reasoning relates to being a self – to being a socially recognized person capable of justifying herself to others. As opposed to Korsgaard's normatively interested appeal to a first-personal agent, the Collective Archive View acknowledges that there is absolutely nothing left to say to a creature not responsive to reasons.

More radically than the discussed constructivist positions, the Collective Archive View acknowledges the consequences of a view on the world in which nothing is a reason. However, it does not follow that this view on the world is the only true view we can take. As opposed to that, we can take the view of normative creatures who *are*, as a matter of fact, responsive to reasons. As I will develop in the next subsection, when discussing error theory, there is nothing, in a world without reasons, that forbids biological creatures to develop a communicative environment in which communicated considerations function as reasons. Acknowledging that the domain of reasons is a communicatively developed part of our socially transformed environment, a socially constructed reality that gives rise to what we describe as self-determined thought, helps us to reconcile the two perspectives, while not conflating them. Constructivists, it seems, want to hold that the modern sciences forbid us to assume something intrinsically normative in the universe, while still holding that our desires, or our rational will, are intrinsically normative. The Collective Archive View avoids this contradiction. On the one

hand, it takes the scientific perspective more seriously than constructivism does, when it says that, from this perspective, absolutely nothing, not even our desires, our rationality, our lives or what have you, are normative – i.e. provide a reason. On the other hand, the Collective Archive View emancipates the normative perspective when it says that there is nothing in itself wrong with socially transforming our reality and constructing realms of reasons. However, we should acknowledge that the domain of reasons is a socially constructed reality that could be otherwise and that is presumably still changing – a reality that can comprise the value of life, the value of consciousness, but also various religious, aesthetic or perfectionist values. The Collective Archive View can thereby explain why Korsgaard's argument from "practical identity" may appeal more to modern readers than for example the realist views that Korsgaard criticizes. However, my view does not support the belief that there is in itself more justification for reasons derived from practical identity than for reasons that are taken to hold simpliciter. In fact, all reasons must be taken to hold simpliciter – or be derivable from reasons that are taken to hold simpliciter.

Justification in terms of reasons is essentially a social phenomenon, dependent on a socially constructed realm. Outside of such socially constructed realms, we encounter what Velleman called "moral black holes" (2013: 71) – spaces where we can recognize no reasons, and consequently suffer an implosion of our self-determined reasoning. In this sense, social constructionism as part of the Collective Archive View, is a relativist position. The Collective Archive View is, however, not relativist in a normative sense. It does not raise the normative claim that everybody ought to do what his or her community judges best.[118] Nor does it state with normative authority that all reasons are community-relative, or that all communities having reasons are equally good. There is a subtle difference between raising the latter claim as a normative judgment and judging as a critical scientist that all creatures who make normative judgments are necessarily moving in a collective archive of mind. Whether one community is better than another is not a question that I have attempted to answer. Instead, I have argued that whatever we judge, whether we judge that all communities are equally good or that our own is better, this judgement is a response to an item that stands in a collective archive – at least when it is a full-fledged reason judgment that comes with self-

[118]My view is very different from normative versions of "communitarianism" in political philosophy (see for example Bellah *et al* 1986). Other versions of communitarianism, such as Sandel's (1984; 1998), MacIntyre's (1983) or Taylor's (1989) might be compatible with some of my descriptive claims. However, I explicitly stress that *no* normative conclusions follow from the fact that normativity, according to my proposal, is a social phenomenon.

determination, based on spontaneity and convergence. Thereby, the Collective Archive View examines the limits and presuppositions of normativity without making itself a normative claim.

5.2.3. Without reasons, no errors

As the previous subsection pointed out, the Collective Archive View shows the limits and presuppositions of normativity, while raising no normative claims of its own. This implies that, strictly speaking, the Collective Archive View does not even say that we have a reason to believe the view itself. Arguing for a view, while at the same time saying that we have no reason to believe the view, seems to be paradoxical. Contemporary error theorists, in particular Bart Streumer (2017), argue, more globally than Mackie, that all normative judgments are false. Streumer argues, moreover, that we are incapable of believing the error theory and can therefore have no reason to believe this theory. While I share some intuitions in the tradition of error theory, I reject the most distinctive claim of error theory, namely that normative judgments rest on an error and are therefore false. The characteristic distinction made by the Collective Archive View, between a perspective from which normative claims can be true or false, and a perspective from which the concepts of normative truth or falsity are inapplicable, refutes error theory, as I will show in this subsection.

To begin with, the Collective Archive View seems to be in a similar vein as error theory, which is supported by various aspects of error theoretic proposals. For example, Mackie (1977), whose approach to morality I presented in the first chapter, explains the illusion of objective moral truth with a cultural process of "objectification". At first sight, this resembles the idea that communities established practices of justification that became entrenched and rigid over time. Likewise, when I argue that reasons are a sort of collective fiction, my view reminds of error theories embracing moral fictionalism, such as Joyce's theory of the "myth of morality" (2001). As distinct from error theorists about morality in the narrow sense, Streumer applies the error theory to normativity broadly conceived and argues that all normative judgments are false. All normative judgments, so the idea, ascribe normative properties, which do not exist. Since nothing has the property of being a reason, Streumer argues, even Mackie's argument from queerness fails to give us a reason for believing in the error theory. As opposed to that, we can only infer the truth of error theory from our incapacity to believe it. The Collective Archive View, as I will show, can reject Streumer's worry by emphasizing its sharp distinction between reflecting in terms of reasons and responding to aspects or

considerations in qualitatively different ways. Let us first consider Streumer's argument for why we cannot believe the error theory:

> Since judgments about reasons for belief are normative judgments, the property of being a reason for belief is a normative property. The error theory therefore entails that there is no reason to believe this theory. And anyone who understands the theory well enough to believe it knows that it entails this. (2017: 137)

From this, Streumer concludes that we cannot believe the error theory, since, as he has established before (ibid.: 134), we cannot believe something while believing that there is no reason to believe it. Accordingly, it is impossible to fully understand the error theory and believe it at the same time. We can thus only believe it partially, Streumer says. We can "believe different parts of the error theory at different times while implicitly changing some of our other beliefs" (ibid.: 171). Our cognitive inability to fully believe the error theory as a whole, Streumer claims, does not count as evidence against the theory. Instead, it makes it more likely to be true (ibid.: 127). Moreover, Streumer holds, our inability to believe it makes the error theory "benign rather than malignant" (ibid.: 176), since the main worry with error theory simply disappears. The main worry of philosophers who think that error theory is true is that it might undermine our deepest convictions. However, if we are incapable of believing that error theory is true, the theory is incapable of undermining any of our beliefs.

Indeed, I am very intrigued by Streumer's idea that we cannot believe all parts of the error theory simultaneously. Likewise, my Collective Archive View consists of two parts which are difficult to believe simultaneously. One part is the thesis that the concept of a reason is fundamental in the sense of inexplicable and non-analysable. The other part is that being a reason can be analysed as being an item in a collective archive of mind. As summarized in subsection 5.2.1., these two different views on reasons can be reconciled by assuming two different perspectives – the first-personal perspective of reason-responsive creatures and the third-personal perspective of scientific investigators of our reason-responsiveness. It seems that we are indeed incapable of taking both perspectives *simultaneously*. As opposed to error theory, however, the Collective Archive View recognizes that, apart from these two distinct perspectives, there is no standpoint from which to judge that one perspective is faulty, or that one perspective refutes the other. To be sure, the incapability of taking both perspectives simultaneously indicates a potential tension in our cognitive faculties. However, it does not make the cognitive faculty of responding to facts as reasons faulty. As opposed to that, it indicates

that frictionless self-understanding – understanding ourselves and the world as a whole, without a cognitive tension – is only a local phenomenon. As chapter 4 suggested, the phenomenon of unified self-understanding is a predictive model of the brain. Moreover, unified self-understanding as determined by reasons is a model that we can only achieve by moving in collective archives. Only collective archives, so the idea, enable what is called a unity of apperception, a cognitive state that subsection 4.4.2. referred to as "grasping simultaneously" or "grasping together". This means that the possibility of frictionless self-understanding is not given by the logic of the universe. Instead, it is an achievement of human cognition. Accordingly, the tension between the two proposed perspectives only indicates the limitation of our cognition. It does not indicate that we make a mistake.

The view I proposed confirms philosophers who hold that we can respond to evidence in ways that are different from responding to it as a reason. Such responses can induce changes in our behaviour for which we cannot coherently give reasons. Olson, in his objection to Streumer, claims that we *can* believe the error theory without making a mistake. I find his objection helpful, without ultimately supporting the label "error theory". Olson says:

> I can base my belief that the error theory is true on the argument from queerness, without judging that his argument favours my attitude that the error theory is true. I can thus maintain that while there are arguments on which I base my belief that the error theory is true, that there are no irreducibly normative reasons for the attitude of believing that the error theory is true. Hence, we can indeed believe the error theory. (Olson 2014: 117f)

Olson supports the idea that we can believe things without believing that we have a reason to believe them with the observation of children and non-human animals who can hold basic beliefs, such as for example perceptual beliefs, without having the capacity to entertain thoughts about reasons. According to Rietveld (2008), a philosopher following ecological psychology cited in chapter 4, these forms of cognition involve normativity as well. However, chapter 4 confirmed Olson in distinguishing sharply between the normativity of reflecting in terms of reasons and responding to aspects of the environment in other ways, which some people might label normativity as well. Streumer, in his reply to Olson's objection, agrees that "we can base a belief on a consideration without believing that this consideration is a reason for this belief" (2017: 143). He holds nevertheless that this "does not mean that we can base a belief on a consideration while believing that this consideration is *no* reason for this belief" (ibid.: 144).

Here, it seems, is the point at which Streumer makes a mistake. He is certainly right that we cannot form a belief based on reasons if we believe that there are reasons against this belief. It is, however, not a problem to have a belief based on reasons in a specific sense, and at the same time a belief that, from another perspective, there are no reasons in this sense. This is what the Collective Archive View proposes. The Collective Archive View holds that as first-personal normative reasoners we respond to certain facts as reasons, while as third-personal scientific reasoners we acknowledge that there are no such reasons. These two perspectives are *compatible*, though not entertainable *simultaneously*. This means that their co-existence is not inconceivable as such – even though their co-existence introduces a tension into our existence, because we cannot take both perspectives simultaneously. However, the fact that we analyse reasons differently when we are considering the natural and collective history of our reason-responsiveness, than we do when we reason *on* them, does not warrant the claim that reasoning on them involves an error.

Acknowledging that our reason-responsiveness is the upshot of communal tradition does not itself give us a reason to reject our responsiveness. In itself, this insight gives us no reason – not even a reason to take an interest in it or to form the beliefs it suggests. It offers us a perspective from which there are no reasons at all – from which we bracket the idea of what I previously called *central reasons*, reasons that hold unconditionally and all-things-considered. From the latter perspective, we could say, the question which reasons we have is inapplicable. From this perspective, we might say that there are no reasons. With this idea, however, the Collective Archive View raises a claim that is different from Streumer's claim, which is that understanding the error theory involves understanding that there is no reason to believe it. Streumer says that the judgment that there is no reason makes us incapable of a reason-based response. To be precise, however, we must replace this *no reason* by *a reason-question inapplicable*. A *reason-question inapplicable*, as opposed to a normative *no reason*, leaves our responses, as far as we apply the reason question, untouched.[119]

[119] Streumer formulates a reply to Cuneo, who raises a similar objection, arguing that an error theory which holds that there are no reasons to believe it is "toothless" (Cuneo 2007: 117). Streumer replies that it is Cuneo's objection which is toothless (Streumer 2017: 172). Streumer assumes that Cuneo's objection establishes the claim that the error theory is false, which he cannot establish because its truth, as Streumer holds against him, is indicated by our incapacity to believe it (for a critical argument against Streumer see also Bruno 2020). However, the error theory as designed by Streumer could be both true and toothless. It seems that neither Streumer's nor Cuneo's claim can be established on the basis of reasons. The Collective Archive View, by contrast, offers a way of explaining why we can believe that something is a reason and believe that there are no reasons at all at the same time – though

From the perspective from which the reason-question is inapplicable, we do not ask which reasons we have; we ask what we do when we ask and answer questions of that kind. Whether we have reason to pose the latter question is a question of the former kind. It seems that we can pose that question irrespective of a belief that we have reason to pose it. Sometimes, such a question just "pops" up. There might be sociological or psychological reasons why it pops up at one time rather than another. However, these reasons are just scientific explanations, not normative justifications. The two kinds of questions – and accordingly the two kinds of perspectives – are so different that the view from one is incapable of undermining the view from the other. Streumer seems to refer to exactly this phenomenon when he presents the idea that we are incapable of believing in the truth of error theory to the effect that it undermines our fundamental normative beliefs. However, I think that the Collective Archive View offers a better explanation for this tension in understanding reasons. There is a tension in our understanding because we, the creatures who take normativity seriously, can also examine what creatures taking normativity seriously are doing as if we were not ourselves such creatures.

Thereby, the Collective Archive View can avoid the odd picture of falsely ascribing properties that do not exist. This is the picture warranting the label "error theory". Speaking of normative properties may be innocent if we just mean the property of a fact to stand in a relation to us that makes it a reason. However, according to the view of reason relations advocated in this book, we do not ascribe the property of standing within that relation when we make a normative judgment. The property of standing in that relation – if we want to speak of a property – only secures that our mind, when coming to see all the relevant considerations in the right way, will respond to the fact in question as a reason. The fact that this fact stands in a relation, which enables a reflective mind to respond like this, is not a fact we usually consider or reflect on. It is not a property we actively ascribe or respond to. We *respond to the fact* as a reason in virtue of the normative relation, but we do *not respond to the normative relation* itself. We do not see the relation at all. We do not ascribe the property of standing within that relation. We see facts as reasons in virtue of the relations in which they stand.

---

it is psychologically difficult to *actively entertain* both beliefs simultaneously. The Collective Archive View solves the puzzle by understanding the difference between the perspectives from which we say that something is a reason, on the one hand, and that there are no reasons at all, on the other hand.

This is a description of the way we work – of the way normative evaluation works. It is a mistake to evaluate this description normatively. In so far as there are norms for a description, they must be distinguished from normativity as reflecting in terms of reasons. There is obviously a tension between, on the hand, simply taking something as a reason, and, on the other hand, realizing that is not simply a reason, but that it is a reason only because of a collective history. This tension between humans who reason and humans who investigate their reasoning does, however, not establish a ground for the claim that it is an error to respond to something as a reason.

To sum up, the Collective Archive View does not necessarily undermine our normative practice. As Streumer's error theory, it is "benign rather than malignant" (Streumer 2017: 176). Likewise, the Collective Archive View shares the basic intuition reflected in Streumer's idea that we can partially believe the error theory, even if we cannot fully believe all parts of it at the same time (2017: 171). The Collective Archive View indeed acknowledges that our cognition is complex and that it gives us ways of accessing the world which stand in a tension. This, however, does not mean that normative judgments, as a particular way of accessing the world, are false. It only shows that the complexity of human cognition makes us vulnerable to tensions in our self-understanding, which we can only avoid by closing off certain perspectives. Whether we open up these perspectives or close them off for the sake of frictionless self-determination in the collective archives of mind is a question which has no predetermined answer. Neither choice can be made on the basis of reasons but, vice versa, neither choice involves an error. This choice, for which there are no clear reasons, is the topic of the next subsection.

5.2.4. An opening question: embracing or rejecting the traditions we depend on?

According to the Collective Archive View, what reasons there are for us depends on a collective tradition of thought. What the view, as opposed to that, does not establish, is that we are right or wrong in following the reasons archived within our communities. There are no reasons over and above our collective archives, no reasons for or against complying with the archives themselves. However, as we are able to gain insight into the genesis of our reasons, insight into what shaped our reason-responsiveness, we are able to reject or embrace these reasons. But based on what reasons? This subsection argues that there are no predetermined reasons in this respect. Elaborating on

an idea by Risberg (2020), I argue that we are capable of asking a deliberative question that has no true answer.[120]

When we recognize that the normative reasons we can discover by thinking correctly about all the relevant evidence are items in a collective archive of mind, we are left with two options. One option is to embrace our dependence on collective archives and to value reasons as what they are. This option seems to transform Reasons Fundamentalism. Reasons, it seems, are no longer simply fundamental. The idea that a realm of reasons in which mind can move is a communal achievement becomes integrated into the way we relate to reasons. Rather than being fundamental in an absolute sense, they are, in a qualified sense, fundamental items within a collective archive. Another option, in distinction from that, is to reject the collective archive as a legitimate foundation of reasons. Sometimes, it might be shattering to realize that, as a matter of fact, you just think the way your ancestral community used to think and you could think in an entirely different way if you were the product of a different community. People who pride themselves for having gained their convictions as a result of their own critical reflection may start to doubt their convictions if they find out about this source. Many tend to reject communal tradition as a source of reasons and emphasize critical distance from this source. Such distance, a common belief suggests, liberates the individual and makes collective harm less likely to happen.

Indeed, the main application of the psychological approach of schema theory, used to support the Collective Archive View in section 4.4., is usually to liberate people from acquired reactive patterns that are bad for them or that are no longer adaptive. Gaining insight into the schemata working on one's cognition, so a common thinking, enables people to relate freely to their behaviour and to choose reactions on the basis of reasons. Against this idea, the here proposed view holds that our responsiveness to reasons is a matter of acquired schemata in pretty much the same way. As a matter of fact, there seems to be no way out of the collective schemata working on us. More precisely, there seems to be no way to reject our reasons *on the basis of reasons*.

Most of the time, this is also unnecessary. Most of the time, we can just revise our reason judgements on the basis of the very same reasons to which we are already responsive. There is another explanation for most of the cases in which people claim that considerations, which certain groups of people believe to be reasons, do not in fact count in favour of a particular action. Most dissenters for good reasons do not have to leave the collective archive

[120] I thank Wlodek Rabinowicz for pointing me to Risberg´s article in this context.

determining the common reason-responsiveness to make their seemingly independent judgments. As an example, we could take the case of a group of people who believe that they have a reason to burn witches. The dissenter in this group must point out the mistake in classifying women as witches – by, for example, showing that an alleged causal link between an adverse event and a woman does not hold. This might be psychologically and practically difficult in many ways – but the difficulty is usually not that there is no common reason-responsiveness *if* the relevant reasons are seen. The discrepancy between the group and the dissenter, in such cases, is not that they move in different archives of mind and, therefore, cannot access the same thoughts.[121] Rather, they have different information about the external world, which determines whether they can apply a thought in the given situation.

In distinction from these more common cases, we can conceive of cases in which two individuals differ with respect to their reason-responsiveness itself. Sometimes, two individuals – or an individual and a group – have insight into the same state of the world and, nevertheless, make different reason judgments. As an example, we can take a fictional painter *Gauguin* who has both a strong personal longing to leave his family in order to realize a creative project, and a duty to care for his wife and children.[122] One way of evaluating this situation is to judge that *Gauguin* has an all-things-considered reason to sacrifice his creative project for his family, since this project weighs less in comparison to the family duties. Another way of evaluating the same situation is to judge that the reasons in favour of aesthetic perfection are weightier than the reasons in favour of self-sacrifice for a family. In this case, there is a genuine difference in the way people respond to the reasons at stake – in the way they weigh aesthetic perfection in relation to caring for a family.

While in the case of *Gauguin*, most people presumably judge that *Gauguin* is not immediately justified in following his artistic ambition at the cost of neglecting his family duties, judgments might differ more widely if we change the case a bit. As opposed to a male "genius" who leaves his economic dependents for good, we might choose the case of a modern woman renegotiating the extent of self-sacrifice a family requires, in relation to other life goals. In such cases, the judgments in modern societies typically differ widely, for example between so-called progressive feminists and religious conservatives. In such cases, the difference in judgment is not entirely due to

[121] As developed in chapter 3, the term „thoughts", as in a (modified) Fregean reading of „senses", and the term „reasons" is used interchangeably throughout this book.

[122] This example was first discussed by Williams (1981). Later, I will pick up Williams' view in more detail and offer a reinterpretation.

different states of knowledge. We cannot generally accuse people who advocate more extensive family duties of not knowing that it makes people dependent or that it hinders people to realize potentials, for example in art or science. It is possible to know all this and still judge that these potentials are of lesser value compared to the persistence of a certain type of family life. Moreover, we cannot reduce all these differences to differences in personal taste or inclination. Many kinds of reason judgments run against personal inclinations. People sacrifice personal inclinations *because* they see good reasons for it. The Collective Archive View holds that we can only see such reasons – reasons that hold independently of our personal inclinations or desires – in virtue of our embeddedness in a community that we can expect to share our judgment.

For example, there might be a community in which scientific or aesthetic interests weigh less than the needs of people. A person embedded in such a community, call her *Eve*, cannot believe – believe as a self-determined reasoner – that she has a reason to engage in a scientific endeavour, which requires at least temporary indifference to the needs of people. As opposed to that, we can imagine a community in which scientific or aesthetic interests may sometimes trump reasons for serving the needs of people. Persons embedded in this community may think that a world in which serving the needs of others is the highest value would be a poor place.[123] A person embedded in this community, call her *Lilith*, may, in distinction from *Eve*, be capable of thinking as a self-determined reasoner that sometimes neglecting the needs of people in order to engage in an aesthetic or scientific project is fully justified. The example of *Eve* and *Lilith* portrays two genuinely different types of reason-responsiveness. Each attempt to defend one or the other seems to be dependent on a collective archive of which the defendant is part. If we embrace one archive, as opposed to the other, we may see no reasons for changing to the other. But is it possible to think about changing to the other? Is it possible to ask oneself whether one should change one's own normative psychology even though one does not see reasons for it?

Risberg (2020) argues that it is in principle always possible to ask oneself whether one should do what one has reason to do. While he holds that answers to the question of what we *ought*[124] to do can be true or false, he denies that the

[123] See for example Wolf's (1982) criticism of "moral saints".

[124] For my argument, I take the concept of "ought" and the concept of a "reason" to be equivalent, thereby setting the ongoing debates about the relation between these two normative concepts aside.

question of what we *shall*[125] do has a true answer. Risberg prefers to state the question of what we shall do as the question *what to do* (2020: Essay IV, p. 20). Thereby, Risberg combines cognitivism about *ought* with non-cognitivism about what he calls the *central deliberative question*. This existentially open question, it seems, is always possible. It is always possible to consider the possibility to act without a reason – or even against reason. Although Risberg's position is compatible with normative realism, Risberg speculates that his insight might "undermine an important argument for the realist position" (ibid.: 26). If the deliberative question is open in the proposed way, normative truths, even though they exist, "seem much less interesting than we often take them to be" (ibid.). At this point, the Collective Archive View has more to say. As argued previously, the possibility to respond to mind-independent normative facts is constitutive of the activity of free thought – or more precisely, of self-determined thought. Acting from first-personal insight into reasons, I pointed out, enables us to understand ourselves as acting meaningfully in a coherent unified system – as acting in a "world".

Examining what it is like to act without or even against reason, I think, confirms this view. While such action is clearly not impossible, it lacks a certain quality – a quality that is typically important to us. Let us go back to my previous example and imagine that *Eve*, being confronted with *Lilith*, starts to feel that the normative foundations her life rests on give her life a lesser quality. This is especially likely if *Lilith* gets more respect or access to public goods than *Eve*. *Eve* could start to feel that being a person who sees no reason to sacrifice scientific interests for the needs of people, a person who sees nothing problematic in being indifferent to the needs of people, as long as one is justified by a scientific project, is somehow better or superior to her. *Eve* might not be able to justify this feeling in terms of reasons, because she genuinely believes that she has all-things-considered reason to care for others and that *Lilith*'s way of life is wrong. At the same time, *Eve* might find herself incapable of fully embracing her own way of life – of wholeheartedly sticking to her warranted reason beliefs. At some point, she might realize that this is the way her reasoning works; but that she rejects the way her reasoning works without being able to fully embrace different reasons. She seems to reject what constitutes her. Although this rejection has a particular first-personal authority[126], it has not the quality of a full-fledged first-personal normative

[125] See also Dreier's (2015: 172) discussion of Reasons Fundamentalism and the "normative question".

[126] Johan Brännmark, as also Risberg in his discussion of Korsgaard, suggested to understand the open question as the truly first-personal normative question. I hope to show that the first-personal quality of doing what one believes that one has reason to do is different from the

insight. Such an insight presupposes that *Eve* has full access to the collective archive in which *Lilith*'s mind moves. Still, *Eve* does not have such access, as indicated by her incapacity of fully embracing the reasons to which *Lilith*, with the natural spontaneity and normative harmony of the self-determined reasoner, is responding. As opposed to that, *Eve*'s rejection of her own tradition of thought happens on a psychological level that is different from the level of reflecting in terms of reasons. Nevertheless, her case suggests that such rejection is possible. Obviously, societal, economic and political conditions can play a role in making people reject the traditions on which they depend. However, this kind of rejection is not always a matter of a position being refuted by reasons.

Accordingly, the Collective Archive View allows for a distinction between people who move on their collective archival grounds without problems, who succeed in what we could call *frictionless self-determination*, and people who are less than self-determined agents, who *lose their mind* – even though they obviously cannot do so on the basis of reasons. In section 5.3., I elaborate on the idea of frictionless self-determination, which I identify as a case of owning one's thinking, and thereby owning the justification for one's choices. We can contrast this to cases of "losing one's mind", of losing the ownership of thought and justification. For the individual person, I suggest, this ultimately amounts to a problem of psychological health and mental sanity.

quality of deciding to act against reason. Some authors, such as Copp (1997) or Tiffany (2007) doubt that the concept of an "all-things-considered" normative question, as Risberg formulates it, is comprehensible. My account, as distinct from Risberg's, could accept that the question is barely comprehensible. However, the question is still possible. As distinct from Risberg, I use the term *"all-things-considered" or central reasons* as describing the normative reasons we "really" have (thereby, I could probably agree with Copp who argues (1997) that it is a question of reasons how to weigh different oughts, except for the fact that, on my account, the "really" in question is understood in light of the socially constructed facts on which our mind depends). What Risberg calls the *central deliberative question* is something different from that. I think two things are possible: Asking what reasons one has all-things-considered, and asking whether one should do what one has all-things-considered reasons to do.

5.3. The costs and gains of embracing the collective archive of our mind

In the previous section, I showed the possibility to either reject or embrace our thoughts, when we realize that they are dependent on a tradition of thought, dependent in a way that is not always directly visible to us and that is not subject to our reflective control. We cannot choose to think differently based on reasons, but we can reject the way we think, nevertheless. We might now wonder what the point of normative facts at all is if we can choose to disregard them – if the deliberative question whether to comply with them is still open. In this section, I will make clearer what we get out of this compliance, and what happens when we refuse to comply or simply fail to comply.

In subsection 5.3.1., I explain what a thinker gains when she makes a full-fledged judgement on the basis of the twin pillars spontaneity and convergence. I describe what she achieves as *ownership of thought* – a state characterized by having one's thoughts at one's full cognitive disposal, understanding them as one's own and being able to effectively use them in communication and justification. This state presupposes the active achievement of cognitive closure, or of uncritical immersion in a collective archive. When we are in this state, we can understand ourselves as fully self-determined, without any frictions, problems or doubts in this self-understanding. No matter how many challenges we face, we are sure that we can in principle always reach a point where no further questions arise. We enjoy normative stability. In subsections 5.3.2. and 5.3.3., I discuss two different states in which we can end up if our condition falls short of this ideal of normative stability. The first case describes the disintegration of the subject, the loss of its spontaneity, as a response to normative instability, while the second case describes a psychotic inflation of subjective spontaneity. The section aims to show that the ability to embrace a collective archive of mind gives us something that is best described as mental sanity, while we have to pay it with uncritical immersion in the archive and closing down the radical existential openness of the deliberative question.

5.3.1. Ownership of thought and normative stability

In this subsection, I will suggest the term "ownership of thought" as a description of what is the case when one makes a full-fledged reason judgment. Making a reason judgment, according to the picture developed in this book, has both a subjective and an intersubjective component. It rests equally on

spontaneity as a subject and convergence with other subjects. Moreover, the spontaneous responses of a subject have the quality of self-determination in a unified framework, aiming at "getting it right" and thereby achieving autonomy. While, according to the theory as it is proposed by Skorupski, certainly autonomy can be reached to different degrees, it is more difficult to understand self-determination as succeeding to a fuller or lesser extent. With the term "ownership", I suggest, we can provide such an understanding. Such an understanding, finally, reveals the implications of the fact that mind, for self-determination to succeed, must achieve the closure of a unified and systematic archive, while the traditions of thought that shape us are not always as unified and systematic as mind needs them to be. Instead, the real world allows for self-determination to succeed to a fuller or lesser degree, depending on the extent to which we own our thoughts.

In its general meaning, the term "ownership" refers to possession legitimized by an authority. It comprises the aspects of having something in one's control and of having a recognized right to this control. In stable conditions, the authority issuing such rights is clearly defined and universally recognized. The rights are clearly codified, and everybody has access to this codex. When conditions destabilize, however, some of these aspects might break away or become less clear. I think, we can gain much from applying the term "ownership" to the conditions under which we make use of reasons. Our use of reasons is characterized by capacities of subjective control, both limited and legitimized by intersubjectively valid authority.

Moreover, the term "ownership" as applied to thought processes, occurs in psychiatric literature (Frith 1992; Cahill 1996; Campbell 1999), where it mainly describes a first-personal quality of the activity of thinking. The term "ownership of thoughts" is used to distinguish thoughts which the subject attributes to itself from thoughts which the subjects experiences as coming from outside, such as in the pathological phenomenon of "thought insertion", occurring sometimes during schizophrenia. In this sense, "ownership of thoughts" is involved when a subject is responding to a reason in the way that Skorupski describes as spontaneity – a response that comes from one's very nature as a subject. In cases of thought insertion, by contrast, reporting one's thoughts is more like reporting perceptions or observations (see e.g. Ratcliffe 2017), rather than the expression of an activity coming from one's very own nature.

The basic quality of thoughts to come from one's very own nature as a subject is clearly a part of ownership but not all there is to this phenomenon. First-personal ownership of a thought requires having a full grasp of it, in the almost literal sense defended in this book – having a grasp of it as standing in

complex relations within a coherent and meaningful world, having the thought, we could say, at our full cognitive disposal. What having a good grasp of a consideration gives to the subject is usually the capacity to use the consideration flexibly, to use it in argumentation, to defend it against potential reasons that stand against it and to understand its implications, i.e. to understand what it means for other situations and what it excludes for other situations. In non-pathological reasoning, having a good grasp of our reasons and having a sense that it is oneself who is grasping them, come together. This is, to begin with, a matter of the subjective phenomenology of grasping. But it also goes beyond subjective phenomenology. For complete health of reasoning, people typically also require an effective capacity to communicate and justify one's thoughts to others. What is an "effective capacity to communicate and justify"?

While, as an ideal, this picture is clear, there are, in reality, many different possibilities of what a communicative situation could look like.[127] Only in the most radical case, a person cannot communicate that p is a reason at all. In another case, the person can communicate that p is *a* reason, but not that it is a reason that justifies. In yet another case, she can convey that it justifies in a respect, but not all-things-considered. Moreover, there is a significant difference between *systematic divergence*, a situation in which whole world views are rejected, and *single-item divergence*, where we find an idea about a specific question rejected.[128] Last but not least, it might matter to the person whether the people rejecting her views are friends that usually support her thinking or come from groups that are expected to disagree. As we can see, the effective capacity to communicate and justify can be limited in many different ways and to different extents. When does the lack of that capacity affect ownership?

Facing disagreement is not itself incompatible with owning one's thought. However, disagreement *can* affect the status of one's ownership, depending on the quality and extent of the disagreement as well as the authority of the people you disagree with. Likewise, the subjective components of ownership, personal grasp and control, can be affected by disagreement, depending on psychological factors as well as maturity and experience. For example, a subject whose thinking is limited to the cognitive traditions of a small, parochial community might lose the subjective ownership of its thought easier than a subject trained to move in the rich funds of a large and diverse group. Likewise, an insecure, anxious or depressed personality may lose it easier than

[127] I thank Johan Brännmark for showing the need to differentiate and offering some distinctions.

[128] I owe these two terms to Johan Brännmark.

a calm, secure and stable personality. What matters for the subjective component to succeed is the capacity to achieve closure in making a judgment – a mental process that requires a degree of predictive certainty and an absence of factors that require us to question our judgments further. When a subject successfully achieves closure, it is in a state of normative harmony and stability. While disagreement always challenges this harmony, disagreement is only a threat to ownership if the harmony cannot be regained.

It seems we are required to choose a trade-off between avoiding challenges and training to meet challenges. While harmony is eventually less frequently challenged in small, parochial communities, it is more severely challenged when confronted with new information in these circumstances. In other words, while parochiality may facilitate closure, this closure may eventually be less stable. These questions, however, depend on many external conditions. What we can state at this point is that ownership of thought correlates with closure. We own our thoughts to the extent we can use them in a predictable unified and coherent framework that holds independently of us. For this purpose, we must succeed in what we can call *uncritical immersion in a collective archive of mind*. When we are uncritically immersed in a collective archive, we understand ourselves as self-determined to the fullest degree – we enjoy all the aspects of ownership, can achieve responses that are truly spontaneous and not in need of further questioning. We encounter no frictions, problems or doubts in our self-understanding as responding to reasons. We see the potential to arrive at warranted judgements that are truly our own in all respects.

To sum up, ownership of thought, as ownership in general, involves control and disposal as well as an externally recognized status. When we own our thoughts, we have them at our full cognitive disposal, can use them for explanation and justification in different circumstances and can reliably predict that others will receive them as we intend – that others thereby recognize us as the owners of our thoughts, as reasonable, mature personalities knowing their reasons. Thus, ownership of thought seems, among other things, to be an aspect of what we value as mental sanity. The following two subsections will portray two different subjective responses to the partial loss of ownership. When we find ourselves incapable to respond to reasons on the basis of spontaneity and convergence, we can either discount or inflate our own spontaneity.

5.3.2. Normative instability with subjective disintegration

Full-fledged reason judgements, as suggested above, are characterized by the subject having ownership of her thought. Ownership of thought is characterized by normative stability – a state in which the subject has a full

grasp of her reasons and a reliable possibility to use those reasons effectively in communication and justification. The possibility of normative stability depends on the capacity to achieve cognitive closure in making a judgement. There may be different causes for a subject's failure to achieve or maintain normative stability. Potential causes may reach from oppressive structures in the community to personal psychological hindrances. Psychological hindrances can presumably be either developmental or innate or a mixture of both. I cannot investigate all causes and contributing factors at this point. However, I will give a general description of the loss of ownership and of how an individual can respond to this loss.

There are different ways for subjects to react to a threat to their normative stability. One way is to doubt one's own spontaneity up to a point at which self-scrutiny leads to disintegration as a subject. Another way, which shall be discussed in the next subsection, is to inflate one's own spontaneity and become indifferent to the responses within the community. Subjects living under conditions of normative instability can presumably also exhibit mixtures of both types of reactions.

As an example for disintegration as a subject in the face of failing convergence, we can cite de Beauvoir's description of her philosophical discussions with Sartre, introduced in chapter 1, in which her radically different point of view was not acknowledged as even basically reasonable. As an upshot of this failed conversation, de Beauvoir writes: "I'm no longer sure what I think, or even if I think at all." (1959: 344) The conversation can be described as failed because it does not amount to an exchange of two reasoners enriching each other's perspectives. This can only succeed if the two reasoners exchange thoughts which they can both grasp in the same way. If the weight or force of a particular item is not seen at all by one party in the conversation, the conversation is unlikely to bring together all considerations into a richer picture shareable by both parties. As a result, one party in the conversation, in this case de Beauvoir, doubts her own spontaneity as a subject – her very own capacity of reasoning. She is neither convinced by her counterpart, following his reasoning with the power of her own mind, nor capable of insisting on her own reasoning and showing why it refutes the arguments broad up by her counterpart.

Abramson suggests that this situation is an instance of "gaslighting" (2014: 4). "Gaslighting" as a technical term in psychotherapy refers to a strategy in communication, which aims at making the other person question her sanity – by methods such as denying obvious facts and falsely accusing the victim. At

this point, it is not of relevance whether this practice is done intentionally[129] or whether the victim is a whole social group or part of an intimate relationship. What should be brought to attention at this point is the deeply disturbing and shattering effect that such behaviour has on the psychology of the person exposed to it. What is shown by "gaslighting" is that an incapacity to communicate one's reasoning can cause a subject to disintegrate as a subject. The subject loses trust in its most basic responses. It loses the sense of understanding its surrounding. It ceases to have a clear sense of what perceived facts call for because others who seem perfectly stable and reasonable interpret these facts in a completely different way. The very capacity of reasoning is undermined.

Disintegration as a subject may further have some similarities to what is known as "depersonalisation disorder" in psychopathology. The DSM-5 manual of mental disorders (American Psychiatric Association 2013: 302-306) describes one of the most central features of depersonalisation as longstanding or recurring feelings of being detached from one's mental processes or body, as if one is observing them from the outside or in a dream. Both the incapacity to trust one's basic responses as a subject and the incapacity to identify with one's mental processes as one's own damage the quality of full-fledged reasoning. Though depersonalisation turns out to enhance the accuracy of certain types of perception (Michal *et al* 2014), it may be identified as an impairment of what we value as the reasoning that gives us a sense of self-determination.[130] Moreover, there is evidence that social connection and affirmation, respectively touch, helps to alleviate the symptoms of depersonalisation (Ciaunica/Fotopoulou 2017; also Michal *et al* 2007; O'Sullivan *et al* 2018). By contrast, influences fostering mistrust and disconnection from one's basic responses as a subject promote a partial or complete disowning of one's thinking. Disintegration of the subject is one way of reacting to normative instability, to unavailable convergence with a community. The route that is chosen is to question subjective spontaneity when spontaneity and convergence cannot be realized. An alternative route will be discussed in the next subsection.

[129] The very term „gaslighting" might invoke the assumption that it is an intentional manipulation. As a technical term in psychology, however, the term describes a behaviour, the description of which is not dependent on intention. Thanks to David Alm for pointing out this potential misunderstanding to me.

[130] I am grateful for the exchange with Anna Ciaunica during her talk at LOGOS (Barcelona) in spring 2019.

5.3.3. Normative instability with subjective inflation

While some subjects, failing to achieve normative stability in their environment, start to question their own responses, other subjects inflate the validity of their own responses as a reaction to normative instability. While the former subjects disintegrate as subjects, the latter subjects discount convergence and maintain a state that has the quality of psychotic certainty. Psychotic certainty is a quality that typically accompanies delusions (Jaspers 1963; Binswanger 1963; Sass 1992; 1994a,b). Delusion is often characterized by incorrigibility and immunity to doubt (Unterthurner/ Kadi 2012). As some psychiatrists hold, delusion as a psychiatric symptom is not so much a matter of what is believed, or that what is believed is false, but of how it is believed – namely with an absolute certainty in the face of counterevidence that reasonable people hardly claim (Kadi 2012).

Other psychiatrists take psychotic certainty to be the result of "background foundations" that are "shaken" (Rhodes/ Gipps 2008: 301), "shifted" or "eroded" (ibid. 299). An approach to delusion, building on Wittgenstein's *On Certainty* (1969), assumes that people capable of getting along in the world must presuppose a bedrock of certainties which are never questioned and which lack justification (Campbell 2001; Rhodes/Gipps 2008).

According to the Collective Archive View, this function is fulfilled by the fundamental items of thought to which subjects respond. Only if an item of thought is part of a collective archive, only if it has a tradition of being shared within a community, there is a chance for a subject to respond to it in a process of healthy reasoning – a reasoning taking place in a dialectic between subjective spontaneity and communal convergence. If the subject, by contrast, responds to items of thought to which its social surrounding cannot respond in the same way, it can only preserve its spontaneity if it cuts off the communicative part of the dialectic. It enters a psychotic universe that is immune to doubt and cannot be challenged by the objections of other people.

In this state of self-inflation, the subject may conceive of itself as self-determined, as owning its thought – not just in the ordinary fallible sense, but even in a heightened infallible sense. However, this subject cannot count as in fact owning its thought because it is not capable of effectively using it for communication and justification. It might subjectively possess its thought processes. Nevertheless, the subject does not enjoy all the aspects of ownership – aspects we enjoy in virtue of our possession being socially recognized and legitimized. When a subject inflates its subjective responses in the absence of collective convergence, it is not the owner of its thought. Inflation as lack of ownership is another way, in addition to subjective disintegration, of reacting to normative instability. Thus, we can conclude that the possibility of true

normative stability in convergence with one's surrounding secures the sanity of reasoning, while this sanity is endangered when we cannot reach such stability.

Both inflation and disintegration must count as pathological ways of reasoning – ways of thinking that are less than full-fledged thinking, types of "madness" that we usually try to escape. Most people are obviously frightened by the prospect of becoming mad. Nevertheless, the Collective Archive View does not offer a basis for judging madness bad and sanity good. If sanity is defined as nothing but normative stability, subjective spontaneity in convergence with one's surrounding, it is even a common philosophical instinct to deny that sanity is necessarily good. Sometimes, we might intuitively hold, the subject that disintegrates because it fails to embrace a horrible tradition of thought, is better than a healthy and harmonious subject, being uncritically immersed in its tradition.

This idea, however, involves an external normative standpoint – a standpoint for which the Collective Archive View, depicting normativity as the psychological phenomenon of moving in a collective archive of mind, leaves no room. The last section of this chapter suggests a way of conceptualizing a standpoint outside of the collective archive in which someone is currently moving – even though it is not independent of *any* collective archive.

5.4. Normative transformation and retrospective justification

In the previous section, I spelled out the costs and gains of embracing the collective archive of one's mind. Succeeding to do so gives us ownership of our thought – a full grasp of our reasons and the secured possibility to use them in communication and justification. The cost we are required to pay, by contrast, is the uncritical immersion in the archive that enables our thought. We must achieve cognitive closure in every reason judgment for it to be the response of a self-determined subject – a subject that owns its reasons.

Now, this perspective on the foundations of our thinking opens up an intriguing possibility. It enables us to ask questions that, from a purely logical point of view, seem impossible to ask. For example, if we know that our reasoning is determined by collective archives that could always be otherwise, we might ask if the archive we have is the best or whether there would be a better one. Likewise, we might ask whether our collective archives are valuable at all or, to put it otherwise, whether we ought to preserve these archives, and

with them the ownership of our thinking. Maybe, we might ask, it would be better to reject our tradition of thought, even if we lose the ownership of our thought if we do so.

Subsection 5.4.1. points out why we might think that these questions must be logically impossible to ask and why, as a matter of fact, we seem to be able to ask them, nevertheless. Asking the question whether we should reject the foundations of our thinking for something better makes the possibility of normative transformation conceivable. Normative transformation, as it is understood here, is the change in shape of a collective archive of mind up to a point where the subjects, which are defined by their capacity to respond to reasons, are different subjects. Subsection 5.4.2. brings up the idea that we might have "reason" to allow ourselves to become different subjects – to develop collective archives that enable us to become different subjects. Such a reason, however, is not a reason we can own. It is a reason that we can only be ascribed retrospectively, as subsection 5.4.3. will explain.

5.4.1. Why value ownership of thought?

Having ownership of thought is to have a good grasp of one's thoughts and a sense that the thinking is one's own activity. Moreover, it is to have an effective capacity to justify one's thinking to others. I have argued that we can achieve full ownership of thought only if we move in a collective archive of mind and comply to its normative structures without questioning them. In subsection 5.2.4., I have shown, by reference to Risberg (2020), that it is always an open question whether to do what one has reason to do – although the decision cannot be based on normative reason beliefs.

Now, the Collective Archive View, as opposed to a simple normative realism also considered by Risberg, offers a seemingly absurd possibility in addition to that. Suppose we recognize what we have reason to do, using our cognition correctly according to the outlined criteria. Moreover, we realize, in accordance with the Collective Archive View, that we are responsive to *this* reason and not another one because of the way our cognition works, while our cognition could always work differently. We understand further that we must use our cognition correctly in order to be the owners of our thought. But, as suggested in the previous section, the Collective Archive View gives us a critical perspective on the ownership of thought. Ownership requires cognitive closure, which requires uncritical immersion in a collective tradition.

Could it be, we might ask now, that the reasons available in the collective archive of our mind are not the best ones? Could it be that we can gain access to archives that are better than the ones in which we currently move? Of course,

we might be aware that everything we can figure out with some degree of certainty by using our own thinking will necessarily be part of the archives in which we move. However, it seems, the question if there *are* better archives becomes possible when we fully understand the implications of the Collective Archive View. The idea turns up that factors beyond our control might induce profound changes of our normative cognition. Factors contributing to such changes might be socio-economic or technological changes.[131] They might also be cultural factors or factors of our very own personal history. We can at best influence such changes indirectly, by taking risks or by relying more heavily on non-reflective parts of our psychology[132]. Still, such changes can, in the long run, establish new archival grounds on which future people or our future selves can move in exactly the same way as we can move in our current archives.

At first glance, asking such questions might seem absurd because it obviously involves a break with the logics of our thought. Applying strict logics, we must find such questions impossible to ask. Terms like "good" or "valuable" are terms obviously belonging to the normative domain.[133] The Collective Archive View depicts the normative domain as the domain of reasons, which is characterized as a collective archive. Asking whether our collective archives are good, or whether ownership of thought is valuable, seems to be like asking whether there is reason to respond to reasons. This is more than Risberg's question whether to do what one has reason to do. It is the question for *reasons* against reasons. It seems that we must be responsive to reasons to give an answer to this question and all reasons we are responsive to, I argued, are items in a collective archive of mind.

At this point, we can differentiate. Firstly, we cannot "have" any reason to which we are not responsive. "Having" a reason in this sense means that there is a reason relation holding for us that we can discover by using our cognition correctly. The reason, in other words, is part of our potential mental states. It is part of the potentials that define our way of being a subject. Secondly, however, we can conceive of reason relations that are not holding for us but

[131] See for example Dewey's and Tuft's (1909) analysis of the development of ethics driven by occurring "frictions" in societies leading to change.

[132] See also the example of Huckleberry Finn discussed by Arpaly (2002).

[133] This might be a point at which it is beneficial to claim a possibility to conceptually distinguish between "being valuable" or "being good" on the one hand, and "being a reason" on the other hand (and thereby departing from the Reasons First Approach that is sometimes connected to Reasons Fundamentalism). However, I would not know how to fill the former concepts with any content.

that would hold for us if we were different kinds of subjects. To be precise, we cannot conceive of these reason relations, but we can conceive of the possibility that we could conceive of them. This is made possible by the sciences that offer us a perspective from which we can understand our way of being a subject as a natural phenomenon that is constantly changing and could be otherwise. We gain an understanding of the genealogy and the limits of our "having" reasons. Such an understanding opens a stunning possibility to our thinking – the possibility to think about reasons for rejecting the normative judgments made by our own mind. Obviously, these reasons are not holding for us, we do not "have" them, and we can never own them as the subjects we currently are. However, we could develop into different subjects – a development during which new reason relations are created. If we reject the foundations of our thought with this possibility in mind, rejecting them can ultimately be more than just the exercise of a radical existential freedom, which it tends to be in the first place.

If we view the phenomenon of normativity primarily as having to do with the possibility of mental sanity through owning one's thought, there is a question naturally suggesting itself – the question whether normativity can sometimes be bad for us. If descriptions like those by de Beauvoir or Bachmann from the beginning of this book – descriptions that feminists clearly see as indicating a moral problem – are paradigmatic cases of lost ownership, it seems that people's need for ownership of thought makes them vulnerable to many forms of oppression. The most radical version of this idea is to claim that the collective archives of mind are necessarily oppressive.[134] They enforce compliance by tying compliance to psychological well-being, and by sanctioning deviance with the loss of central presuppositions for mental health. This radical conclusion, however, seems unwarranted to me – at least if we use the term "oppression", as we commonly do, as naming something that damages the oppressed in her potentials. According to the Collective Archive View, the existence of mind-independent forces is not in itself damaging us. By contrast, it is only the social construction of a collective archive of mind that gives us the potential to understand ourselves as self-determined in a meaningful world.

Nevertheless, it seems obvious to me that we could find some collectively created ways of being a subject worse than others. I do not think that there is an absolute truth, but it seems common though to value certain changes as progress. There might be an objection at this point. Why, one might wonder,

[134] I especially thank Federica Berdini for discussions about the inadequacy of "unity" as a concept in normative theory, and the relation between "unity" and oppression (see also Berdini 2019).

should we not advocate an absolutely mind-independent truth, while depicting the collective archive as the part of the truth that we can currently access? Indeed, many feminist philosophers choose this path. Calhoun (2015), for example, distinguishes between "getting it right" and "practicing morality successfully with others" as two separate, but equally important aims of morality. While this distinction is obviously very suitable for activist purposes, I do not think that it is sound as a philosophical view answering deep metaethical questions. As argued in chapter 3, it is difficult to make sense of "getting it right" in a way that is not dependent on a socially determined way of being a subject. Arguing that there is an ultimate end at which the development of all subjectivity in the universe aims is simply too demanding and too implausible from a naturalistic perspective on evolution.

So, do we not have a reason to bring about the changes in thinking that we typically call "moral progress"? Unless the changes are not just more accurate realizations of what is already part of our current moral psychology, I would say "no", we do not "have" such a reason. However, fundamental changes in psychology could occur and our future selves may come to have such reasons. Does this have any implications for us now at all, we might wonder. No implications that we have *reason* to care for, but still implications that we *might* care for, is the answer I can come up with. As with Risberg's central deliberative question, there is no "predetermined" or "true" answer to the question whether to value ownership of thought. Still, it is possible to not place highest value on it and instead go against one's own normative psychology. Furthermore, it is possible that our future self or future generations will find themselves with reasons for what we did – reasons, however, which our current selves cannot be the owners of. All these possibilities are open but far beyond our control. In the following, I will go back to the case of *Eve* and *Lilith*, introduced in 5.2.4., and consider the possibility of *Eve* risking normative stability, giving away the ownership of thought, for the exploration of better alternatives – without owning a reason for the belief that they are in fact better.

5.4.2. Risking normative stability for the exploration of better alternatives?

When we enjoy mental sanity, we make the experience of frictionless self-determination in a collective archive. We are stable owners of our thought. The question whether to disidentify with our thoughts or to reject their normative force does not arise. Responding to reasons, we could say, is a matter of mental sanity, while the question whether there is reason for mental sanity, or whether mental sanity is valuable, is not relevant.

Sometimes, however, this question becomes relevant. Let us turn back to the example of *Eve* and *Lilith*, introduced in subsection 5.2.4.: *Eve* asks herself the question whether her self-sacrifice for others is really necessary, even though her mind still weighs the reasons provided by the needs of others as weightier than the reasons supporting her scientific endeavours – the endeavours that would require her to neglect the needs of others to a certain degree. She cannot revise her judgment. Her mind reasons that way. However, she might feel a discontent with the way her mind reasons – a discontent that is difficult to spell out and justify in terms of reasons.

Eve has now two options. She can follow her judgment, because she cannot fully spell out the reasons for changing it. She cannot change her judgment based on reasons she fully owns. As opposed to that, however, she could be bold and act despite of a lack of ownership. It can be because she is just callous, or because she has a suspicion that *Lilith*, who seems competent in many other respects, might see reasons which she will also see when she develops as a subject. Suppose, *Eve* chooses the latter option and suffers a loss of ownership of thought. In any case, this means that she can no longer effectively justify herself in the community in which she is embedded. To an extent that can vary with her personality traits, she might also suffer some disintegration as a subject.

At this point, the evaluation of *Eve* depends on processes that *Eve* can neither control nor predict. We can imagine two different courses of development, leading to *Eve-1* and *Eve-2*. *Eve-1*'s rejection of the collective archive is followed by other women and is accompanied by other changes that *Eve* could neither control nor predict. At the end, the whole community in which *Eve-1* lives arrives at a state of development in which survival and flourishing is secured without everyone sacrificing completely for the needs of others. As opposed to that, it even turns out that the various endeavours of artists and scientists contributed a great deal to the achievement of this state. The artists and scientists, of which *Eve-1* was one, were at least partly insensitive to the immediate needs of people, when they chose to prioritize their personal endeavours over immediately serving others. Finally, they brought about a state in which everybody sees good reason for dedicating a fair amount of time and resources to endeavours that do not immediately serve the needs of people.

Eve-2, by contrast, violates obligations to her family and ends up in isolation and mental illness. It is not only that she herself does not succeed in making a causal contribution to change. In the world of *Eve-2*, it is, by contrast, never the case that society changes its view on which obligations a woman has towards her family, or its view on the general question whether such obligations can be outweighed by other considerations.

Was *Eve* justified in acting as she did? It seems that this depends on whether *Eve-1* or *Eve-2* is realized. What *Eve* has done was giving up the ownership of thought, and thereby also the ownership of justification, for discovering new grounds. Either she will discover a justification on this new ground, or she will never be justified. If she successfully discovers a new ground, a ground she can share with others, it might be warranted to say that she transformed herself as a subject, together with the archive of reasons that also transformed. Sometimes, we might intuitively hold, such transformations can be good things – good in the intuitive sense that we would regret not having undergone them. So, can such transformations ever be supported by reasons? Is there anything we can say about the normative situation in which *Eve* is at the moment of choice?

Williams, in his seminal paper *Moral Luck* (1981), suggests a conception of retrospective justification. Justification according to this conception is nothing we can own. However, retrospective justification seems to be an acceptable form of justification, without which, we may find, humanity would indeed be very poor and limited. It seems that frictionless self-determination in the collective archives of mind is not the only interesting resource of human mind. The openness for change and continuous transformation, with which this conception of self-determination stands in tension, seems at least equally important in the history and evolution of humanity.

5.4.3. Retrospectively justifying transformative explorers

The previous subsection distinguished between two possible scenarios in which a subject, *Eve*, losing the ownership of thought, can end up. According to one scenario, *Eve-1* turns out to play a pioneering role in the development of thought – a transformative role in the collective archives of mind, which can later be appreciated by the collective as a whole. In the other scenario, by contrast, what is happening to *Eve-2* is "just" a form of mental illness and personal failure.

The crucial point is that, in many cases, we can only distinguish those scenarios in the aftermath. At the time at which they happen, the situation may look the same. If somebody just got mad, or lost ground for something that turns out to be recognizable as a reason at a later point can only be recognized at the later point. This lies in the very nature of transformation which is that it restructures a collective archive from within. At the end of the transformative process, the normative landscape looks different. The reason relations holding at the end of the process are different from those that were holding at the beginning. The process is not taking place within the hermetic systematicity of

a unified normative framework. Instead, the hermetic structure breaks up and the framework itself takes a new shape. Consequently, it is impossible for a subject to own the reasons for transformation, to play the transformative role as a self-determined agent. Necessarily, it seems, the subject must give up its cognitive and psychological need for self-determination. It must leave the sphere of frictionless self-determination and survive in the tension of a reflectively groundless existence.

It seems clear that, as an individual, you can never have a full-fledged reason to prefer this process to mental sanity and successful self-determination. However, when we consider the many cases of normative pioneering which brought us intellectual and cultural achievements that we wouldn't like to miss, it seems problematic to judge that the transformative individuals should not have any reasons that justify them. Even though they could not see the reasons in favour of their behaviour themselves, we might object that these reason relations have been holding. The Collective Archive View, as opposed to that, holds that there cannot be any reason relation holding for a subject at a particular point in time when this subject – given the form of subjectivity it has, given the foundations of thought that define it as a subject – is incapable of responding to it. In the transformative process, a subject becomes a new subject. It turns out to have reasons which it previously didn't have.

The transformation, indeed, seems to involve a moment of indeterminacy – a moment of radical openness, which the Collective Archive View captures. The question whether the process that starts without reasons turns out to establish new reasons, is not predetermined. In the cases in which an individual that lost ground is justified at the end of a transformative process, the justification is what Williams, in his seminal essay *Moral Luck*, calls "essentially retrospective" (1981:24). This means that whether the individual will be justified or not lies beyond its control and possibilities of assessment. Moreover, it means that the justificatory status is not even determined at the point of action. William's paradigmatic case for retrospective justification is the case of a fictional painter *Gauguin* who violates his duties towards his family in embarking on a journey that is supposed to make him exceptional as an artist. Only an exceptional success as an artist, so Williams idea, can justify *Gauguin*'s decision. Accordingly, whether *Gauguin* is justified in his choice depends on the unpredictable outcome of a future development. If he fails as an artist, he has "no basis for the thought that he was justified in acting as he did" (1981: 23).

We could, of course, debate whether exceptional aesthetic achievements can ever justify the violation of family duties. Alternatively, we might criticize Williams' idea for making justification at the moment of action impossible as

such, since any action could fail, and the result would in principle be the same as the result of a failed *Gauguin*. Accordingly, we might want to reject the very concept of retrospective justification. We might hold that either everything is only justified when its success is secured, which is absurd, as it makes justification of most future-directed actions impossible, or that *Gauguin*, having a realistic belief in his success, has been justified from the beginning, irrespective of his success.

I would like to suggest a way of interpreting Williams' *Gauguin* that can deal with this objection. It is probably not the mainstream interpretation, but I think it is the best one for emphasizing the distinctive aspect of justification being *retrospective*, instead of the justification just basing on risky or controversial assumptions. My interpretation appeals to an aspect that can only be found in truly exceptional developments. We can set aside the question what Williams himself originally intended. In cases of exceptional developments, it sometimes occurs that the actual experience of something we have never seen and couldn't foresee makes our reproaches silent and makes us accept justifications we usually wouldn't have accepted. When we actually see the result, we accept justifications we wouldn't have accepted if we were just pointed out and explained the possibilities. This is a phenomenon we could identify as making human existence what it is. It is an important feature of human existence that our practices of justification sometimes face a limit. Sometimes, we see ourselves compelled to value and appreciate things made possible only by behaviour that we were incapable of justifying before we actually saw the compelling result.[135] We can imagine that an exceptional aesthetical experience transforms our judgment. It radically transforms the way we see things. It is, of course, open for debate whether we should give aesthetical experiences such an authority. This is, however, not interesting here.

Not all cases of transformative madness that are of interest here produce stunning aesthetic value as Williams' *Gauguin* presumably did. Many instead restructure the moral or social values we endorse. The analogy between Williams' *Gauguin* and cases of transformative madness is, nevertheless, that both induce a change in our judgment that was unpredictable. In both cases, a behaviour for which nobody could see a justification succeeds in inducing a process of change after which we see things differently. Such a change, we can argue, is a change in subjectivity. It is not a previously unknown fact that is now discovered and that changes the picture. Instead, it is experiencing a new kind of pull. It is experiencing the normative force of a fact that we eventually

[135]See Wallace (2013) for academic philosophy as one of the best examples.

have often considered already, but not found to have this force before. Of course, there is a difference between momentous states of enthusiasm which lead us to make choices we later regret, and processes of profound and stable change. A stable and profound change is a change that we never seriously regret. It is a change that we would instead regret not to have happened.

Such changes, it seems, indicate a transformation of subjectivity. Transformations of subjectivity, it seems, belong to the history of humanity as well as to the history of personal lives. L.A. Paul, in her book *Transformative Experiences* (2014), even depicts basic life events, such as becoming a parent, as transformative experiences, changing the way we are, including the way we think and value things. If this were true, it would mean, according to the Collective Archive View, that we can never own a reason for most of our more important and unpredictable decisions. Changes in subjectivity – that means changes in the way of being a subject, which are more than mere changes of preferences held by the same subject – are changes in the normative landscape, as developed in subsection 3.4.3.

The frictionless self-determination of fully owned thought requires a closed normative landscape – a hermetic systematicity, which, it seems, we hardly ever find. We can argue that this puts the Collective Archive View under attack for promoting a picture of full-fledged thinking that is not more than a myth – hardly ever realizable and if so, only at the cost of the most basic human developments. This objection can be accepted in principle. The Collective Archive View does not refute this objection. By contrast, it acknowledges that the need for ownership of thought and self-determination may have consequences we tend to evaluate negatively, when we come to understand them. For one thing, the psychological need for owning our thoughts makes us vulnerable, because we might suffer from bad mental health if the need is not fulfilled. For another thing, the need for ownership of thought makes the adaption to new circumstances more difficult and might let us miss possibilities for development.

These aspects notwithstanding, I assume that the idea of self-determined thinking with which I have worked in this book is pervasive in human history and essential for personal existence and well-being so far. I think it also has an intense aesthetic appeal. The solitary mind moving in a collective archive and thereby achieving the full ownership of its thoughts is a both beautiful and empowering idea – although not without downsides. How to weigh the good of self-determination against the good of transformation is a question which the analysis of reasons cannot answer. My analysis suggests that this question is impossible to answer. Instead, an analysis of reasons can show the problem with weighing the psychology of reasoning against other goods – the

psychology of reasoning which forms the basis for our weighing and reflecting in terms of reasons. This problem, revealed by the Collective Archive View, reveals a tension in human self-understanding. We can, it seems, ask whether our way of evaluating is best, without being able to evaluate this question independently of our subjective way of evaluating.

Nevertheless, we cannot deny that the very possibility to ask this question introduces an openness into our normative reasoning. It makes our reasoning more fragile, but also more flexible and open for potentially beneficial changes. It exposes us to the existential risk of madness, but also enables us to adapt to a continuously changing world and find new ways of relating to it. The idea of retrospective justification offers a possibility of how we could relate to the dynamic and continuously changing nature of subjectivity. Which direction of change is right and whether change itself is good or bad is, nevertheless, beyond the scope of my work. Solving these questions in a definite way, it seems, would require reasons – reasons which only hold within a collective archive.

5.5. Conclusion

The unified and coherent domain of reasons that must be available in order to enable our self-understanding as free agents in a meaningful world, is a collective archive of mind – so the view defended in this book. Normative facts that hold independently of us and that we can discover by using our cognition correctly are socially constructed facts. A scientific understanding of cognition, as a variable way of modelling the world as unified, enables us to think beyond the collective archive of our mind, but not to evaluate it in terms of reasons. This chapter has considered the possibility to reject the collective archive of our mind in search for new ways of thinking that do not yet exist. The Collective Archive View with its comprehensive approach to reasons and its inclusion of an empirical perspective on what is first-personally given opens up this possibility.

After comparing my view to the previously discussed metaethical positions, and thereby highlighting the distinctive character of the view, I examined the question of what we gain when we comply to the reasons we have. Vice versa, I examined what we lose when we reject the collective archive of our mind: the ownership of our thought; a cognitive phenomenon with subjective and intersubjective components; components that define sanity of mind and that we associate with various aspects of mental health. Finally, I presented an

example of a person acting against her own reflective mind. While she risks ending up in sheer madness, there is a chance that giving up the ownership of her thought facilitates a transformative process after which the pioneer finds herself with new reasons. If this process succeeds is essentially dependent on others. Sometimes, only future generations can find themselves with the relevant reasons. The reasons for engaging in such a transformative process are by definition not reasons we can own – but they can justify the risky transformative endeavour retrospectively.

By completing the view developed in this book with that intriguing perspective, I have given an analysis of reasons that is more comprehensive and informative than the classical metaethical positions discussed in the beginning. I pointed out that we must accept reasons – at least the central reasons holding about what to do all-things-considered or unconditionally – as fundamental items of thought. However, we must only accept them as irreducible and not further analysable when we reason *on* them. When we study empirically what we are doing when we are reasoning on them, by contrast, we can analyse them as fundamental items in a collective archive of mind, rather than fundamental constituents of reality in an absolute sense. What we call "reality" in the end, the physical universe or the world of humans, is not more than a matter of terminology. The integration of two distinct perspectives into a comprehensive picture enables us to understand the dependence of our thought on contingent, socially constructed foundations. It enables us to make sense of the possibility that losing one's mind can contribute to the development of new potentials of mind – new forms of subjectivity. It is possible that the capacity to transform reason is an essential key capacity for adopting to a changing world. It might enable us to survive and rebuild meaningful worlds if need be. It enables us to gain new potentials. However, this capacity stands in a tension with our need for self-determination and ownership of thought. Survival and development, we might conclude, requires us to find a workable trade-off between preserving our reasoning and risking it for the discovery of new grounds.

Conclusion

This book revealed an astonishing fact about what we conceive as our capacity of self-determined thought. Self-determined thought, according to the conception employed here, is constituted by reason responses. When we understand our reasons, so the idea, we own our thoughts and decisions. Ownership in this sense has both a subjective and an intersubjective component. The *subjective* component is that understanding reasons provides us with a grasp of the world we are living in. We grasp ourselves and our actions as within a world – a causally, spatiotemporally and normatively unified system in which our actions can be evaluated. The *intersubjective* component, as opposed to that, is that when we have a good grasp of things, we can usually communicate and justify our actions to others. When we own our thoughts and have a clear and stable grasp of the world, we can be self-determined agents. We can base our actions on our own understanding. This might give us a sense of mental power and independence from others.

The *Collective Archive View* defended in this book brings an interesting twist into this self-conception: According to the Collective Archive View, we respond to an item created by communal tradition whenever we respond to a reason. The thought process constituted by reason responses is, according to this picture, the activity of moving within a collective archive. This means that we think the thoughts, or grasp the reasons, that people around us and before us already grasped. When a consideration is not in this historical cultural fund, we might eventually be able to consider it, but not in the way that is constitutive of *full-fledged self-determined thought*. As opposed to that, pioneers of thought are in danger of losing the ownership of their thoughts. They can continue by inflating their own judgment and discounting the judgment of others, or they will disintegrate as subjects, as subjects inhabiting an understandable world. Both ways of reacting to the loss of ownership come at the cost of healthy reasoning.

Consequently, we must acknowledge that the basis of our perceived intellectual independence, the basis of our self-determined thinking, is compliance with the structures and contents of a collective archive of mind. This makes self-determined thinking dependent on others in the most radical

sense. It makes, paradoxically enough, *thinking our own thoughts* dependent on *thinking the thoughts of others*. Another way of framing it would be to say that when we think our very own thoughts, we think the thoughts of others. The more we deviate from others, the less able are we to conceive of our thoughts as our own. The more accurate our insight and compliance with the collective archive, the stronger our sense of self-determination.

Thereby, the thesis presented in this book can account for the fictional scene portrayed at the very beginning: As described in the Introduction, Ingeborg Bachmann's fictional character Mara describes herself as an observer of her own thinking. She does not know what she thinks, or, even more interestingly, she feels that she cannot think what she wants. In light of the Collective Archive View, we can interpret this as the result of refusing, or being unable, to fully embrace the foundations of the available collective archive. Yet, while being unable to embrace the current foundations, Mara does not have proper access to new grounds.

In moments like the ones described here, we actually encounter the material limits of our thinking – the fact that our thinking moves in collective archives. Our thinking can get nowhere without being able to move on an archival ground. The less stable the ground, the less we can think what we want and experience our thoughts as ours. What we encounter here is the sheer fact that our capacity of reasoning, which gives rise to what we call our intellectual independence, is a socio-material phenomenon, itself radically dependent.

The Collective Archive View embraces metaethical Reasons Fundamentalism – a view according to which reasons are fundamental items of thought, irreducible and inexplicable, not further analysable from the immediate (first-personal) perspective of reasoners. As distinct from this perspective, which takes normativity at face value, there is another, more detached, scientifically interested perspective, critically scrutinizing normativity rather than taking it at face value. From this detached (third-personal) perspective, the fundamental items of thought can be analysed further, as items in a collective archive of mind. The claim that reasons are both fundamental (non-analysable) and analysable as items in a collective archive is not self-contradictory. The self-contradiction is only apparent. It springs from the fact that we can view our nature as reason-responsive creatures from two different perspectives. This amounts to a tension in our self-conception. However, the two perspectives are so different, that what they reveal does not amount to an outright contradiction. There is no easy way of answering the question if any of these perspectives is more fundamental than the other. This book does not take a clear stance here. It is at least permissible

to consider both perspectives as equally fundamental, in the sense of "equally constitutive for our access to the world" and, probably, "equally important."

What should we do with the knowledge provided by this book? What should we do with the knowledge that what we must take as fundamental in reasoning, and what constitutes our self-determined thought, is the product of a collective tradition?

Since clear normative reasons only hold for creatures within a collective tradition, there cannot be a clear reason-based answer to this question. The question is, as it were, asked "from nowhere". The question we ask is practically the question whether we should do what we have reason to do. The "should" here is a non-normative, a merely deliberative "should", as opposed to the normative concept "ought" which we might be able to replace by the concept of a reason. The deliberative "should" in question is radically free, radically non-determined by reasons.

The insights of the Collective Archive View, however, push us to ask this question, this radically open, radically non-determined existential question. The position from which we ask this question is a both intriguing and dangerous position. With regard to the normative landscape, it is indeed "nowhere". It is, to pick up again a formulation by Velleman that stood in the beginning of this book, a "moral black hole". If we encounter this place, or rather non-place, in a situation in which we must make a decision and function as a reasoner, we might conceive it as threatening. However, if we dare the step into this position for heuristic explorative purposes, in a safe situation, we might gain insights that could be beneficial – setting us into a new relation to the reason relations we take for granted, showing the openness and changeability of our realm of reasons and enabling us to embark for new grounds. This, however, comes with a risk. Whether we successfully arrive at new grounds, or whether we destroy ourselves and other values, is not entirely up to us. It essentially involves moral luck.

The Collective Archive View is a proposal for a comprehensive understanding of what reasons as a phenomenon in human life are. It goes beyond metaethics narrowly conceived. It shows both the limits of classical metaethical inquiry and a possibility to combine *conceptual* inquiry with *material* (substantial or factual) inquiry. Such a combination is urgently called for in the current debates, in which a very advanced cognitive science and psychology stands against a very fine-grained and detailed literature in metaethics. The state of the art in both disciplines offers many promising connections and overlaps. Yet, there is still not much systematic work translating concepts from the empirical sciences of mind into concepts of metaethics. The approaches, to be sure, are clearly distinct. However, they are

not talking about different worlds or different creatures – they are both talking about the same natural human being. The achievement of this book is a contribution to understanding what an entity analysed in metaethics *is* when it is analysed in empirically minded work.

As philosophers and normative creatures, we can relate very differently to the results of this systematic synthesis. We can embrace our reasons as the products of our ancestral communicative tradition. Likewise, we can reject the authority of reasons when we discover where they come from. Hopefully, we will find a balance between the two different reactions and use the knowledge in a beneficial and productive way, appreciating our traditions and embarking for a gradually improving future.

References

Abramson, K. (2014). "Turning up the lights on gaslighting." *Philosophical Perspectives* 28(1): 1-30.

Alm, D. (2011). "Defending Fundamental Requirements of Practical Reason: A Constitutivist Framework." *Journal of Philosophical Research* 36: 77-102.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders: DSM-5.* Arlington, VA: American Psychiatric Association.

Ameriks, K. (2003). *Interpreting Kant's Critiques.* Oxford: Clarendon Press/Oxford University Press.

Arpaly, N. (2002). *Unprincipled Virtue: An Inquiry Into Moral Agency*. New York: Oxford University Press.

Arroyo, C. (2011). "Freedom and the source of value: Korsgaard and Wood on Kant's formula of humanity." *Metaphilosophy* 42(4): 353-359.

Àsta (2018). *Categories We Live By*: *The Construction of Sex, Gender, Race, and Other Social Categories.* New York: Oxford University Press.

Bachmann, I. (1971). *Malina*. Frankfurt a. M.: Suhrkamp Verlag.

Bachmann, I. (1978). *Sämtliche Erzählungen*. München: Piper.

Bagnoli, C. (2017). "Constructivism in metaethics." *The Stanford Encyclopedia of Philosophy*.

Baiasu, S. (2016). "Constitutivism and Transcendental Practical Philosophy: How to Pull the Rabbit Out of the Hat." *Philosophia* 44(4): 1185-1208.

Baier, A. (1997a). "Doing things with others: The mental commons." L. Alanen, S. Heinämaa and T. Wallgren (eds.). *Commonality and Particularity in Ethic*s. New York: St. Martin's Press, pp. 15-44.

Baier, A. (1997b). *The Commons of the Mind*. Paul Carus Lectures (Book 19). Chicago and LaSalle: Open Court.

Bayne, T. (2010). *The Unity of Consciousness*. Oxford: Oxford University Press.

de Beauvoir, S. (1959). *Memoirs of a Dutiful Daugther.* Cleveland: World Publishing.

Bell, D. (2008). "Communitarianism." *The Stanford Encyclopedia of Philosophy*.

Bellah, R. N., et al. (1986). "Habits of the Heart: Individualism and Commitment in American Life." *Ethics* 96(2): 431-432.

Bennett, K. (2017). *Making Things Up*. New York: Oxford University Press.

Berdini, F. (2019). "Agency's Constitutive Normativity: An Elucidation*." Journal of Value Inquiry* 53(4): 487-512.

Bergström, L. (1990*). Grundbok i värdeteori*. Stockholm: Thales.

Berman, S. (2003). "Islamism, Revolution, and Civil Society." *Perspectives on Politics* 1(2): 257-272.

Bicchieri, C. (2005). *The Grammar of Society: The Nature and Dynamics of Social Norms*. New York: Cambridge University Press.

Binswanger, L. (1963). *Being-in-the-World. Selected papers of Ludwig Binswanger*. New York: Basic Books.

Block, N. (1995). "How many concepts of consciousness?" *Behavioral and Brain Sciences* 18(2): 272-287.

Block, N. (2011). "The higher order approach to consciousness is defunct." *Analysis* 71(3): 419 - 431.

Blomberg, O. (2013). "Review of Wolfgang Prinz's 'Open Minds: The Social Making of Agency and Intentionality'." *Metapsychology Online Reviews* 17(4).

Boyd, R. (2018). *A Different Kind of Animal: How Culture Transformed Our Species*, Princeton, NJ: Princeton University Press.

Brandom, R. B. (1986). "Frege's Technical Concepts: Some Recent Developments." L. Haaparanta and J. Hintikka (eds.). *Frege Synthesized*. Dordrecht: D. Reidel Publishing, pp. 253-295.

Brinck, I. (1997). *The Indexical 'I' the First Person in Thought and Language*. Dordrecht: Kluwer Academic Publishers.

Brinck, I. and R. Liljenfors (2013a). "The developmental origin of metacognition." *Infant and Child Development* 22: 85-101.

Brinck, I. and R. Liljenfors (2013b). "Reply to commentaries." *Infant and Child Development* 22: 111-117.

Broome, J. (2007). "Does Rationality Consist in Responding Correctly to Reasons?" *Journal of Moral Philosophy* 4(3): 349-374.

Broome, J. (2013). *Rationality Through Reasoning*. Malden, MA: Wiley-Blackwell.

Broome, J. (2018). "Reason Fundamentalism and What Is Wrong With It." D. Star (ed.). *The Oxford Handbook of Reasons and Normativity*. Oxford: Oxford University Press, pp. 297-318.

Brown, R., *et al*. (2019). "Understanding the Higher-Order Approach to Consciousness." *Trends in Cognitive Sciences* 23(9): 754-768.

Brown, R. and J. LeDoux (2017). "A Higher-Order Theory of Emotional Consciousness." *Proceedings of the National Academy of Sciences of the United States of America* 114 (10) E2016-E2025.

Bruno, D. (2020). "Error Theory, Unbelievability and the Normative Objection." *Journal of Ethics and Social Philosophy* 17(2).

Burge, T. (1979). "Individualism and the mental." *Midwest Studies in Philosophy* 4(1): 73-122.

Burge, T. (1992). "Frege on knowing the third realm." *Mind* 101(404): 633-650.

Buss, S. and L. Overton (2002). *Contours of Agency: Essays on Themes From Harry Frankfurt*. Cambridge, MA: MIT Press.

Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.

Cahill, C. (1996). "Psychotic experiences induced in deluded patients using distorted auditory feedback." *Cognitive Neuropsychiatry* 1(3): 201-211.

Calhoun, C. (1994). "Kant and Compliance With Conventionalized Injustice." *Southern Journal of Philosophy* 32(2): 135-159.

Calhoun, C. (2015). *Moral Aims: Essays on the Importance of Getting It Right and Practicing Morality with Others*. New York: Oxford University Press.

Campbell, J. (1999). "Schizophrenia, the space of reasons, and thinking as a motor process." *The Monist* 82(4): 609-625.

Campbell, J. (2001). "Rationality, meaning, and the analysis of delusion." *Philosophy, Psychiatry, and Psychology* 8(2-3): 89-100.

Carroll, N. (2001). *Beyond Aesthetics: Philosophical Essays*. Cambridge, UK: Cambridge University Press.

Cassam, Q. (1997). *Self and World*. Oxford: Oxford University Press.

Chang, R. (2009). "Voluntarist reasons and the sources of normativity." D. Sobel and S. Wall (eds.). *Reasons for Action*. New York: Cambridge University Press, pp. 243-271.

Chang, R. (2013). "Grounding practical normativity: going hybrid*." Philosophical Studies* 164(1): 163-187.

Chater, N. (2018). *The Mind is Flat: The Illusion of Mental Depth and The Improvised Mind*. London: Penguin Books.

Chemero, A. (2003). "An outline of a theory of affordances." *Ecological Psychology* 15(2): 181-195.

Chemero, A. (2009). *Radical Embodied Cognitive Science.* Cambridge, MA: MIT Press.

Ciaunica, A., & Fotopoulou, A. (2017). "The touched self: Psychological and philosophical perspectives on proximal intersubjectivity and the self." T. F. C. Durt, & C. Tewes (eds.) *Embodiment, enaction, and culture: Investigating the constitution of the shared world*. Cambridge, MA: MIT Press, pp.173–192.

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. New York: Oxford University Press.

Cohen, G. A. (2001). *If You're an Egalitarian, How Come You're So Rich?* Cambridge, MA: Harvard University Press.

Copp, D. (1997). "The ring of gyges: Overridingness and the unity of reason." *Social Philosophy and Policy* 14(1): 86-106.

Copp, D. (2008). "Darwinian skepticism about moral realism." *Philosophical Issues* 18(1): 186-*206.*

Cuneo, T. (2007). *The Normative Web: An Argument for Moral Realism*. New York: Oxford University Press.

Currie, G. (2006). "Narrative representation of causes." *Journal of Aesthetics and Art Criticism* 64(3): 309–316.

Currie, G. (2010*). Narratives and Narrators: A Philosophy of Stories*. New York: Oxford University Press.

Deem, M. J. (2016). "Dehorning the Darwinian Dilemma for Normative Realism." *Biology and Philosophy* 31(5): 727-746.

Dennett, D. C. (2017). *From Bacteria to Bach and Back: The Evolution of Minds*. New York: Norton and Company.

Dewey, J. and J. H. Tufts (1909). *Ethics*. New York: H. Holt and Company.

Diaz-Leon, E. (2015). "What Is Social Construction?" *European Journal of Philosophy* 23(4): 1137-1152.

Dreier, J. (2015). "Can reasons fundamentalism answer the normative question?" G. Björnsson, C. Strandberg, R. Francén Olinder, J. Eriksson and F. Björklund (eds.). *Motivational Internalism*. New York: Oxford University Press, pp. 167-181.

Dummett, M. (1991). *Frege and Other Philosophers*. Oxford: Clarendon Press.

Dunbar, R. (2000). "On the origin of the human mind." P. Carruthers and A. Chamberlain (eds.). *Evolution and the Human Mind: Modularity, Language and Meta-Cognition*. Cambridge, UK: Cambridge University Press, pp. 238-253.

Dunbar, R. (2016). *Human Evolution: Our Brains and Behavior*. New York: Oxford University Press.

Dunbar, R. (1996). *The Trouble with Science*. Cambridge, MA: Harvard University Press.

Dworkin, R. (2011). *Justice for Hedgehogs*. Cambridge, MA: Belknap Press of Harvard University Press.

Elkind, D. (1967). "Egocentrism in Adolescence." *Child Development* 38(4): 1025-1034.

Enoch, D. (2006). "Agency, shmagency: Why normativity won't come from what is constitutive of action." *Philosophical Review* 115(2): 169-198.

Enoch, D. (2010). "The epistemological challenge to metanormative realism: how best to understand it, and how to cope with it." *Philosophical Studies* 148(3): 413-438.

Enoch, D. (2011). "Shmagency revisited." M. Brady (ed.). *New Waves in Metaethics*. London: Palgrave-Macmillan.

Enoch, D. (2013). *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.

Epstein, B. (2015). *The Ant Trap: Rebuilding the Foundations of the Social Sciences*. New York: Oxford University Press.

Epstein, B. (2018). "Social Ontology." *The Stanford Encyclopedia of Philosophy*.

FitzPatrick, W. J. (2014). "Skepticism about Naturalizing Normativity: In Defense of Ethical Nonnaturalism." *Res Philosophica* 91(4): 559-588.

Fodor, J. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Frankfurt, H. (1987). “Identification and Wholeheartedness.” F. D. Schoeman (ed.). *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. New York: Cambridge University Press.

Frankfurt, H. G. (1971). "Freedom of the will and the concept of a person*." Journal of Philosophy* 68(1): 5-20.

Frankfurt, H. G. (1982). “The importance of what we care about.” *Synthese* 53 (2):257-272.

Frankfurt, H. G. (1998). *Necessity, Volition, and Love*. Cambridge, UK: Cambridge University Press.

Frankfurt, H. G. (2004). *The Reasons of Love*. Princeton, NJ: Princeton University Press.

Frege, G. (1892). "Über Sinn und Bedeutung". *Zeitschrift für Philosophie und philosophische Kritik* 100(1): 25–50.

Frege, G (1918). ‘Der Gedanke. Eine logische Untersuchung’. *Beiträge zur Philosophie des deutschen Idealismus,* I: 58–77.

Frege, G. (1956). "The thought: A logical inquiry." *Mind* 65(259): 289-311.

Frege, G. (1983). *Nachgelassene Schriften*. Hamburg: Felix Meiner Verlag.

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. New York: Oxford University Press.

Friston, K. (2009). "The free-energy principle: a rough guide to the brain?" *Trends in Cognitive Sciences* 13(7): 293-301.

Frith, C. (2007). *Making up the mind: how the brain creates our mental world.* Malden, MA: Blackwell Publishing.

Frith, C. D. (1992). *Essays in cognitive psychology. The cognitive neuropsychology of schizophrenia*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Gaertner, S. L., & Dovidio, J. F. (2000*). Reducing intergroup bias: The common ingroup identity model*. New York: Psychology Press.

Gibson, J. J. (1966). "The Senses Considered as Perceptual Systems." *Synthese* 17(2): 230-232.

Gibson, J. J. (1979*). The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.

Gilbert, M. (1989). *On Social Facts*. London: Routledge.

Graham, J., et al. (2009). "Liberals and Conservatives Rely on Different Sets of Moral Foundations." *Journal of personality and social psychology* 96: 1029-1046.

Hacking, I. (1995). “The looping effects of human kinds.” D. P. D. Sperber, & A. J. Premack (eds.). *Symposia of the Fyssen Foundation. Causal cognition: A multidisciplinary debate*. Oxford: Clarendon Press/Oxford University Press, pp. 351–394.

Hacking, I. (1999). *The Social Construction of What?* Cambridge, MA: Harvard University Press.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion.* New York: Pantheon/Random House.

Haidt, J. and C. Joseph (2004). "Intuitive ethics: how innately prepared intuitions generate culturally variable virtues." *Daedalus* 133(4): 55-66.

Hanisch, C. (2016). "Two Conceptions of Practical Reasons." *Les ateliers de l'éthique/The Ethics Forum* 11(2-3): 108-132.

Hanisch, C. and S. Baiasu (2016). "Constitutivism and Kantian Constructivism in Ethical Theory: Editorial Introduction." *Philosophia* 44(4): 1125-1128.

Haslanger, S. (2012). *Resisting Reality: Social Construction and Social Critique*. New York: Oxford University Press.

Heft, H. (1989). "Affordances and the body: An intentional analysis of Gibson's ecological approach to visual perception." *Journal for the Theory of Social Behaviour* 19(1): 1–30.

Heft, H. (2001). *Ecological psychology in context: James Gibson, Roger Barker, and the legacy of William James's radical empiricism*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Helmholtz, H. (1925). *Treatise on Physiological Optics*: Translated from the 3rd German Ed. Rochester, NY: Optical Society of America.

Heras-Escribano, M. (forthcoming) "Pragmatism, enactivism, and ecological psychology: towards a unified approach to post-cognitivism." *Synthese:* 1-27

Herman, B. (2002). "Bootstrapping." S. Buss a. L. Overton (eds.). *The Contours of Agency. Essays on Themes from Harry Frankfurt*. Cambridge, MA: MIT Press.

Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press.

Huntington, S. P. (1993). "The Clash of Civilizations?" *Foreign Affairs* 72(3): 22-49.

Hurlburt, R. and E. Schwitzgebel (2007*). Describing Inner Experience?: Proponent Meets Skeptic.* Cambridge, MA: MIT Press.

Hussain, N. J. Z. and N. Shah (2006). "Misunderstanding Metaethics: Korsgaard's Rejection of Realism." R. Shafer-Landau (ed.). *Oxford Studies in Metaethics 1*. Oxford: Clarendon Press, pp. 265-294.

Hussain, N. J. Z. and N. Shah (2013). "Metaethics and Its Discontents: A Case Study of Korsgaard." C. Bagnoli (ed.). *Moral Constructivism: For and Against*. Cambridge, UK: Cambridge University Press.

Inhelder, B. and J. Piaget (1958). *The Growth of Logical Thinking From Childhood to Adolescence an Essay on the Construction of Formal Operational Structures*. London: Routledge & Kegan Paul.

Jaspers, K. (1963). *Philosophy and the World: Selected Essays and Lectures*. Chicago: Regnery.

Johansson, P., Hall, L., Sikstrom, S., Olsson, A. (2005). "Failure to detect mismatches between intention and outcome in a simple decision task." *Science* 310(5745): 116-119.

Joyce, R. (2001). *The Myth of Morality*. Cambridge, UK: Cambridge University Press.

Joyce, R. (2006). "Metaethics and the empirical sciences." *Philosophical Explorations* 9(1): 133 – 148.

Kadi, U. (2012). „Der wahnsinnige König. Zu Wahn und Verstehen bei Jaspers und Lacan.“ G. Unterthurner & U. Kadi (eds.). *Wahn. Philosophische, psychoanalytische und kulturwissenschaftliche Perspektiven.* Wien/Berlin: Turia+Kant.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kant, I. (1999) [1787]. *Critique of Pure Reason* (The Cambridge Edition of the Works of Immanuel Kant). Translated and edited by Paul Guyer and Allen W. Wood. Cambridge, UK: Cambridge University Press

Kant, I. (2009) [1784]. *An Answer to the Question: “What is Enlightenment?”* London: Penguin Books.

Katsafanas, P. (2013). *Agency and the Foundations of Ethics: Nietzschean Constitutivism*. Oxford: Oxford University Press.

Katsafanas, P. (2018). “Constitutivism about Practical Reasons.” D. Star (ed.). *The Oxford Handbook of Reasons and Normativity*. Oxford: Oxford University Press, pp. 367-394.

Kidd, I. J., *et al*. (2017*). The Routledge Handbook of Epistemic Injustice*. New York: Routledge.

Kiesewetter, B. (2017). *The Normativity of Rationality*. Oxford: Oxford University Press.

Kitcher, P. (1993). "The Evolution of Human Altruism*." Journal of Philosophy* 90(10): 497.

Kiverstein, J., et al. (2019). "The feeling of grip: novelty, error dynamics, and the predictive brain." *Synthese* 196(7): 2847-2869.

Kleingeld, P. (1998). "Kant on the Unity of Theoretical and Practical Reason." *Review of Metaphysics* 52(2): 500-528.

Knausgaard, K. O. (2013). *My Struggle. Book Two*. New York: Penguin Random House.

Koffka, K. (1935). "Principles of Gestalt Psychology." *Philosophical Review* 45(4): 412-415.

Korsgaard, C. M. (1992). "Creating the kingdom of ends: Reciprocity and responsibility in personal relations." *Philosophical Perspectives* 6: 305-332.

Korsgaard, C. M. (1996). *The Sources of Normativity*. Cambridge, UK: Cambridge University Press.

Korsgaard, C. M. (2006). “Morality and the Logic of Caring.” H. G. Frankfurt. *Taking Ourselves Seriously and Getting it Righ*t. Stanford: Stanford University Press, pp. 55-76.

Korsgaard, C. M. (2008a). “Acting for a Reason”. C. M. Korsgaard. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press, pp. 207-230.

Korsgaard, C. M. (2008b). “Realism and Constructivism in Twentieth-Century Moral Philosophy.” C. M. Korsgaard. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press, pp. 302-326.

Korsgaard, C. (2009). *Self-Constitution. Agency, Identity, and Integrity*. Oxford: Oxford University Press.
Kusch, M. (2002). *Knowledge by Agreement: The Programme of Communitarian Epistemology*. New York: Oxford University Press.
Laland, K. N. (2017). *Darwin's Unfinished Symphony: How Culture Made the Human Mind*. Princeton, NJ: Princeton University Press.
LeDoux, J. (2015). *Anxious: Using the Brain to Understand and Treat Fear and Anxiety*. New York: Penguin Random House.
LeDoux, J. (2019). *The Deep History of Ourselves: The Four-Billion-Year Story of How We Got Conscious Brains.* New York: Viking.
Lenk, E. (1981). "Indiskretionen des Federviehs: Pariabewusstsein schreibender Frauen seit der Romantik." *Courage* 6(10): 24–34.
Lennox, S. (1992). "The Feminist Reception of Ingeborg Bachmann." *Women in German Yearbook* 8: 73-111.
Lennox, S. (2007). "Gender, the Cold War, and Ingeborg Bachmann." *Studies in 20th & 21st Century Literature* 31(1).
List, C. (2016). "Levels: Descriptive, Explanatory, and Ontological." *Noûs* (4): 852-883.
Locke, J. (1980) [1689]. *Second Treatise of Government.* Red. C. B. MacPherson. Indianapolis: Hackett Publishing.
Lo Presti, P. (2020). Persons and affordances. *Ecological Psychology* 32(1):25-40.
Lugones, M. (1987). "Playfulness, "World"-Travelling, and Loving Perception." *Hypatia* 2(2): 3 - 19.
MacIntyre, A. C. (1983). *After Virtue: A Study in Moral Theory*. Notre Dame, Indiana: University of Notre Dame Press.
Mackie, J. L. (1977). *Ethics, Inventing Right and Wrong*. New York: Penguin Books.
Mähringer, G. (2015). *Die strukturdynamische Konstitution personalen Willens durch ein soziales Gefüge. Analyse einer internalistischen Autonomiekonzeption*. Wien: Universität Wien (Master thesis).
Mähringer, G. (2020). *Reason Monolithism: A New Darwinian Dilemma* (manuscript).
Mallon, R. (2019). "Naturalistic Approaches to Social Construction." *The Stanford Encyclopedia of Philosophy*.
Mandler, G. (1967). "Organization and Memory." *Psychology of Learning and Motivation* 1: 327-372.
Mandler, J. M. (1983). "What a story is." *Behavioral and Brain Sciences* 6(4): 603.
Mandler, J. M. (1984). *Stories, Scripts, and Scenes: Aspects of Schema Theory*, Hillsdale, NJ: Lawrence Erlbaum.
Mantel, S. (2018). *Determined by Reasons: A Competence Account of Acting for a Normative Reason*. New York: Routledge.
Markovits, J. (2014). *Moral Reason*. New York: Oxford University Press.
Marti, G. (1998). "The significance of the distinction between concept mastery and concept possession." *Philosophical Issues* 9: 163-167.

Marti, G. (2007). "Weak and strong directness: Reference and thought." *Philosophy and Phenomenological Research* 74(3): 730-737.
McDowell, J. (1994). *Mind and World*. Cambridge, MA: Harvard University Press.
McDowell, J. (1998). *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press.
Mendonça, D. Curado., M.. Gouveia, S. (eds.). (2020). *The Philosophy and Science of Predictive Processing*. London: Bloomsbury Academic.
Mercier, H. and Sperber, D. (2017). *The Enigma of Reason.* London: Penguin Books.
Mercier, P. (2008). *Night Train to Lisbon*. New York: Grove Press.
Michal, M., et al. (2007). "Depersonalization, mindfulness, and childhood trauma." *The Journal of Nervous and Mental Disease* 195(8): 693-696.
Michal, M., et al. (2014). "Striking discrepancy of anomalous body experiences with normal interoceptive accuracy in depersonalization-derealization disorder." *PLoS One* 9(2): e89823.
Mink, L. O. (1972). "Interpretation and Narrative Understanding." *Journal of Philosophy* 69(20): 735.
Mink, L. O. (1987). *Historical Understanding*. Ithaca: Cornell University Press.
Moore, G. E. (1903). *Principia Ethica*. New York: Dover Publications.
Nagel, T. (1986). *The View From Nowhere*. New York: Oxford University Press.
Nagel, T. (1996). *The Last Word*. New York: Oxford University Press.
Nersessian, N. J. (2010). *Creating Scientific Concepts*. Cambridge, MA: MIT Press.
Nietzsche, F. W. (1903). *The Dawn of Day*. New York: Dover Publications.
Olson, J. (2014). *Moral Error Theory: History, Critique, Defence*. New York: Oxford University Press.
Olson, J. (2018). “The Metaphysics of Reasons.” D. Star (ed.). *The Oxford Handbook of Reasons and Normativity*. Oxford: Oxford University Press, pp. 255-274.
O'Neill, O. (1989*). Constructions of Reason: Explorations of Kant's Practical Philosophy*. New York: Cambridge University Press.
O'Sullivan, N., et al. (2018). "I Am There ... but Not Quite: An Unfaithful Mirror That Reduces Feelings of Ownership and Agency." *Perception* 47(2): 197-215.
Parfit, D. (1984). *Reasons and Persons*. New York: Oxford University Press.
Parfit, D. (2011a). *On What Matters. Volume I*. Oxford: Oxford University Press.
Parfit, D. (2011b). *On What Matters. Volume II*. Oxford: Oxford University Press.
Pauer-Studer, H. (2016). "“A Community of Rational Beings”. Kant’s Realm of Ends and the Dinstinction between Internal and External Freedom." *Kant-Studien* 107(1): 125-159.
Paul, L. A. (2014). *Transformative Experience*. New York: Oxford University Press.
Persson, I. (2005). *The Retreat of Reason: A Dilemma in the Philosophy of Life*. Oxford: Oxford University Press.
Persson, I. and J. Savulescu (2012). *Unfit for the Future: The Need for Moral Enhancement*. Oxford: Oxford University Press.

Pettit, P. (1993). *The Common Mind: An Essay on Psychology, Society, and Politics*. New York: Oxford University Press.

Piaget, J. (1954). *The Construction of Reality in the Child*. London: Routledge.

Prinz, W. (2013). *Open Minds. The Social Making of Agency and Intentionality*. Cambridge, MA: MIT Press.

Putnam, H. (2004*). Ethics Without Ontology*. Cambridge, MA: Harvard University Press.

Quine, W. V. (1948). "On What There Is." *Review of Metaphysics* 2(1): 21-38.

Rabinowicz, W. and T. Rønnow-Rasmussen (2005). "A Distinction in Value: Intrinsic and for its own sake." T. Rønnow-Rasmussen and M. J. Zimmerman. *Proceedings of the Aristotelian Society* 100: 115-129.

Rakoczy, H., Warneken, F., & Tomasello, M. (2008). "The sources of normativity: Young children's awareness of the normative structure of games." *Developmental Psychology* 44(3): 875–881.

Ratcliffe, M. (2017). *Real Hallucinations: psychiatric illness, intentionality, and the interpersonal world*. Cambridge, MA: MIT Press.

Rauscher, F. (2002). "Kant's moral anti-realism." *Journal of the History of Philosophy* 40 (4):477-499.

Rauscher, F. (2015). *Naturalism and Realism in Kant's Ethics*. Cambridge, UK: Cambridge University Press.

Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Raz, J. (1999a). *Engaging Reason: On the Theory of Value and Action*. Oxford: Oxford University Press.

Raz, J. (1999b). "Explaining normativity: On rationality and the justification of reason." *Ratio* 12(4): 354–379.

Read, C and A. Szokolszky (2020). "Ecological Psychology and Enactivism: Perceptually-Guided Action vs. Sensation-Based Enaction." *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2020.01270 (accessed 12th Aug 2020)

Reed, E. S. (1996). *Encountering the world: Toward an ecological psychology*. New York: Oxford University Press.

Rhodes, J. and R. G. T. Gipps (2008). "Delusions, Certainty, and the Background." *Philosophy, Psychiatry, and Psychology* 15(4): 295-310.

Rietveld, E. (2008). "Situated normativity: The normative aspect of embodied cognition in unreflective action." *Mind* 117(468): 973-1001.

Rietveld, E. and J. Kiverstein (2014). "A Rich Landscape of Affordances." *Ecological Psychology* 26(4): 325-352.

Risberg, O. (2020). *Guiding Concepts. Essays on Normative Concepts, Knowledge, and Deliberation.* Uppsala: Uppsala Universitet.

Rønnow-Rasmussen, T. (2013). “Motivation and Motivating Reason.” C. Svennerlind, J. Almäng and R. Ingthorsson (eds.). *Johanssonian Investigations. Essays in Honour of Ingvar Johansson on His Seventieth Birthday.* Frankfurt: Ontos Verlag, pp. 464-485.

Rosati, C. (2003). "Agency and the open question argument." *Ethics* 113(3): 490-527.

Rumelhart, D. E. (1980). "On Evaluating Story Grammars." *Cognitive Science* 4(3): 313-316.

Ryan, K. J. and S. Gallagher (2020). "Between Ecological Psychology and Enactivism: Is There Resonance?" *Frontiers in Psychology* 11.

Sandel, M. J. (1984). "The Procedural Republic and the Unencumbered Self." *Political Theory* 12(1): 81-96.

Sandel, M. J. (1998*). Liberalism and the Limits of Justice*. Cambridge, UK: Cambridge University Press.

Sapolsky, R. M. (2017). *Behave: The Biology of Humans at Our Best and Worst*. New York: Penguin Random House.

Sass, L. A. (1992). *Madness and Modernism: Insanity in the Light of Modern Art, Literature, and Thought.* New York: Basic Books.

Sass, L. A. (1994a). "Civilized madness: schizophrenia, self-consciousness and the modern mind." *History of the Human Sciences* 7(2): 83-120.

Sass, L. A. (1994b). ""My So-Called Delusions": Solipsism, Madness, and the Schreber Case." *Journal of Phenomenological Psychology* 25(1): 70-103.

Sayre McCord, G. (2012). "Metaethics." *The Stanford Encyclopedia of Philosophy*.

Scanlon, T. (1998). *What We Owe to Each Other*. Cambridge, MA: Belknap Press of Harvard University Press.

Scanlon, T. M. (2014). *Being Realistic about Reasons*. Oxford: Oxford University Press

Schlosser, G. & G. P. Wagner (eds.) (2004). *Modularity in Development and Evolution*. Chicago: University of Chicago Press.

Searle, J. (1995). *The Construction of Social Reality*. New York: The Free Press.

Searle, J. (2010). *Making the Social World: The Structure of Human Civilization*, Oxford: Oxford University Press.

Sen, A. (1999). *Commodities and Capabilities*. New Delhi: Oxford University Press.

Setiya, K. (2007*). Reasons Without Rationalism*. Princeton, NJ: Princeton University Press.

Shoemaker, D. W. (1996). "Theoretical Persons and Practical Agents*." Philosophy and Public Affairs* 25(4): 318-332.

Skorupski, J. (2010). *The Domain of Reasons*. Oxford: Oxford University Press.

Skorupski, J. (2012). “Précis of The Domain of Reasons” *Philosophy and Phenomenological Research* Vol. LXXXV (1): 174-184.

Skorupski, J. (2017). "Reply to Kurt Sylvan: Constructivism? Not Kant, not I." *Philosophical Quarterly* 67(268): 593-605.

Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell.

Smith, M. (2015). "The Magic of Constitutivism." *American Philosophical Quarterly* 52(2): 187-200.
Sober, E. (1994). *From a Biological Point of View: Essays in Evolutionary Philosophy*, Cambridge, UK: Cambridge University Press.
Sober, E. and D. S. Wilson (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.
Star, D. (2018). *The Oxford Handbook of Reasons and Normativity*. Oxford: Oxford University Press.
Strawson, P. (1962). “Freedom and Resentment.” *Proceedings of the British Academy*, Volume 48: 1962. 48: 1-25.
Street, S. (2006). "A Darwinian dilemma for realist theories of value." *Philosophical Studies* 127(1): 109-166.
Street, S. (2008a). "Constructivism about reasons." R. Shafer-Landau (ed.). *Oxford Studies in Metaethics* 3. Oxford: Clarendon Press, pp. 207-245.
Street, S. (2008b). "Reply to Copp: Naturalism, normativity, and the varieties of realism worth worrying about." *Philosophical Issues* 18(1): 207-228.
Street, S. (2010). "What is constructivism in ethics and metaethics?" *Philosophy Compass* 5(5): 363-384.
Street, S. (2011). "Mind-Independence Without the Mystery: Why Quasi-Realists Can't Have It Both Ways.″ R. Shafer-Landau (ed.). *Oxford Studies in Metaethics* 6. Oxford: Clarendon Press, pp. 1-32.
Streumer, B. (2017). *Unbelievable Errors: An Error Theory About All Normative Judgments*. Oxford: Oxford University Press.
Swanson, L. R. (2016). "The Predictive Processing Paradigm Has Roots in Kant." *Frontiers in Systems Neuroscience* 10(79)
Sylvan, K. (2016). "Skorupski on spontaneity, apriority and normative truth." *Philosophical Quarterly* 66(264): 617-628.
Tahko, T. (2018). "Fundamentality." *The Stanford Encyclopedia of Philosophy*.
Tait, W. (2005). *The Provenance of Pure Reason: Essays in the Philosophy of Mathematics and its History*. New York: Oxford University Press.
Taylor, C. (1989*). Sources of the Self: The Making of the Modern Identity.* Cambridge, MA: Harvard University Press.
Tiffany, E. (2007). "Deflationary normative pluralism." *Canadian Journal of Philosophy* 37(5): pp. 231-262.
Tomasello, M. (2001*). The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.
Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.
Turiel, E. (1983). *The Development of Social Knowledge. Morality and Convention*, New York: Cambridge University Press.

Turiel, E. (2006). “The development of morality.” D. N. Eisenberg, & R. M. Lerner (eds.). *Handbook of child psychology 3: Social, Emotional, and Personality Development.* Hoboken, NJ: Wiley, pp. 789–857.

Unterthurner, G. & U. Kadi. (eds.) (2012). *Wahn. Philosophische, psychoanalytische und kulturwissenschaftliche Perspektiven.* Wien/Berlin: Turia+Kant.

Varela, F. (1988). “Structural coupling of simple cellular automata: On the origin of meaning.” F. C. E.Secarz, N.A.Mitchinson, and T.Tada (eds.). *The Semiotics of Cellular Communication in the Immune System*, Vol. H23. New York: Springer, pp. 151-161.

Varela, F., *et al*. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Velleman, D. (1992). “The Guise of the Good.” *Noûs* 26(1): 3-26.

Velleman, D. (2000). *The Possibility of Practical Reason*. New York: Oxford University Press.

Velleman, J. D. (2003). "Narrative explanation." *Philosophical Review* 112(1): 1-25.

Velleman, J. D. (2013). *Foundations for Moral Relativism*. Cambridge, UK: OpenBook Publishers.

Veluwenkamp, H. (2017). "Parfit’s and Scanlon’s Non-Metaphysical Moral Realism as Alethic Pluralism." *Ethical Theory and Moral Practice* 20(4): 751-761.

Vygotsky, L. S. (1962). "Thought and Language." *Philosophical Quarterly* 14(55): 178-179.

Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Processes.* Cambridge, MA: Harvard University Press.

de Waal, F. (2006). “Morally evolved: Primate social instincts, human morality, and the rise and fall of 'Veneer Theory'.” S. Macedo and J. Ober (eds.). *Primates and Philosophers*. Princeton, NJ: Princeton University Press.

de Waal, F. B. (2008). "Putting the altruism back into altruism: the evolution of empathy." *Annual Review of Psychology* 59: 279-300.

de Waal, F. B. M. (1996). *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge, MA: Harvard University Press.

Wallace, R. J. (2013). *The View from Here: On Affirmation, Attachment, and the Limits of Regret*. New York: Oxford University Press.

Walsh K.S., McGovern D., Clark A., O'Connell R.G. (2020). " Evaluating the neurophysiological evidence for predictive processing as a model of perception." *Annals of the New York Academy of Science*. 1464(1): 242-268.

Weichold, M. (2015). *Zwischen Reflex Und Reflexion: Intelligenz Und Rationalität Im Unreflektierten Handeln*. Berlin: De Gruyter.

Weichold, M. (2018). "Situated agency: towards an affordance-based, sensorimotor theory of action." *Phenomenology and the Cognitive Sciences* 17(4): 761-785.

Weichold, M. and G. Thonhauser (2020). "Collective Affordances." *Ecological Psychology* 32(1): 1-24.

Wertsch, J. V. (1985). *Vygotsky and the Social Formation of Mind*. Cambridge, MA: Harvard University Press.

Williams, B. (1981). “Moral Luck”. B. Williams. *Moral Luck: Philosophical Papers 1973–1980*. Cambridge, UK: Cambridge University Press, pp. 20-39.

Wittgenstein, L. (1953). *Philosophical Investigations*. Hoboken, NJ: Wiley-Blackwell.

Wittgenstein, L. (1969). *On Certainty* (ed. Anscombe and von Wright). New York: Harper Torchbooks.

Wolf, S. (1982). "Moral saints." *Journal of Philosophy* 79(8): 419-439.

Wolf, S. (2002). “The True, the Good, and the Lovable: Frankfurt's Avoidance of Objectivity.” S. Buss & L. Overton (eds.). *The Contours of Agency. Essays on Themes from Harry Frankfurt*. Cambridge, MA: MIT Press.

Wollstonecraft, M. (2007). *A vindication of the rights of woman. Late Modern Philosophy: Essential Readings with Commentary*. E. S. Radcliffe, R. McCarty, F. Allhoff and A. Vaidya. Oxford: Blackwell.

Wood, A. W. (1999). *Kant’s Ethical Thought*. Cambridge, UK: Cambridge University Press.

Wood, A. W. (2000). “Kant's practical philosophy.” K. Ameriks (ed.). *The Cambridge Companion to German Idealism.* Cambridge, UK: Cambridge University Press, pp. 57-75.

Wood, A. W. (2007). *Kantian Ethics*. Cambridge, UK: Cambridge University Press.

Wringe, B. (2013). "Wolfgang Prinz, Open Minds: The Social Making of Agency and Intentionality. Review." *Philosophy in Review* 33(2): 138-141.