

Kantian constructivism: a restatement

Janis David Schaab

A thesis submitted for the degree of PhD
at the
University of St Andrews



2019

Full metadata for this thesis is available in
St Andrews Research Repository
at:

<http://research-repository.st-andrews.ac.uk/>

Identifier to use to cite or link to this thesis:

DOI: <https://doi.org/10.17630/10023-17374>

This item is protected by original copyright

This item is licensed under a
Creative Commons License

<https://creativecommons.org/licenses/by-nc-nd/4.0>

Candidate's declaration

I, Janis David Schaab, hereby certify that this thesis, submitted for the degree of PhD, which is approximately 80,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2015.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date

Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date

Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the

library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Janis David Schaab, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

Embargo on electronic copy for a period of 5 years on the following ground(s):

- Publication would preclude future publication

Supporting statement for electronic embargo request

Publication might preclude further publication as a monograph.

Title and Abstract

- I agree to the title and abstract being published.

Date

Signature of candidate

Date

Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Janis David Schaab, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date

Signature of candidate

Abstract

This thesis provides a restatement of Kantian constructivism with the aim of avoiding some of the objections and clearing up some of the ambiguities that have haunted previous versions of the view. I restate Kantian constructivism as the view that morality's normativity has its source in the form of second-personal reasoning, a mode of practical reasoning in which we engage when we address demands person-to-person.

By advancing a position about the source of moral normativity, Kantian constructivism addresses a metaethical question, albeit one that is distinct from the questions that many traditional metaethical positions, such as moral realism, focus on. Kantian constructivism has an advantage over competing views of the source of moral normativity when it comes to answering the so-called Normative Question, which I interpret as the question of why we are rationally required to do what we take to be our moral obligation. Kantian constructivism can answer this question because, unlike its competitors, it does not conceive of practical reason as a receptive faculty that is determined by external inputs. Instead, it regards the very form of second-personal reasoning as grounding the fact that morality is normative, thus explaining morality's rational authority.

Although second-personal reasoning is fundamentally distinct from the merely first-personal mode of reasoning that we must engage in insofar as we are agents, all those agents whom we would ordinarily consider bound by moral obligations seem to engage in it. Indeed, although it involves irreducibly second-personal notions, such as accountability and the authority to address legitimate demands, second-personal reasoning is not to be mistaken for a social practice. Instead, it can be applied to purely self-regarding contexts, such as that of committing oneself to a personal project and thereby holding oneself accountable for pursuing it, as well as to interactions with others.

CONTENTS

Acknowledgments	ix
References to Kant's Works	x
INTRODUCTION	1
PART I: KANTIAN CONSTRUCTIVISM AND ITS PLACE IN METAETHICS.....	7
Chapter 1: The Sources of Normativity	8
1. Introduction	8
2. Kantian Constructivism's Metaethical Ambitions	9
2.1 Metaethics vs. Normative Ethics	9
2.2 The Function of Normative Concepts.....	11
2.3 Stance-Independence	12
3. What are Sources of Normativity?	14
3.1 Sources as Grounds.....	14
3.2 The Sources vs. the Ontology of Normativity.....	16
3.3 Kantian Constructivism and Moral Realism	19
4. Autonomy as the Source of Morality's Normativity	21
4.1 Obligation and the Categorical Imperative.....	23
4.2 Autonomy and the Form of Law.....	27
4.3 Grounding Moral Normativity	33
4.4 Voluntarism, Internalism, Constructivism	35
5. Conclusion	37
Chapter 2: The Normative Question	38
1. Introduction	38
2. The Puzzle	38
3. What is the Normative Question?	41
3.1 Practical Reason and Moral Normativity.....	42
3.2 Motivating the Question.....	44
3.2.1 <i>The Critique of Reason</i>	45

3.2.2 <i>The Sceptical Threat from Debunking Explanations</i>	47
3.3 Formulating the Question	50
3.4 The Dialectic	53
4. Conclusion	56
Chapter 3: Answering the Normative Question	57
1. Introduction	57
2. Source Externalism	57
3. Source Internalism and Source Voluntarism	62
4. Kantian Constructivism	64
4.1 The Form of Practical Reason.....	64
4.2 Two Challenges.....	67
4.3 Philosophy as Defence.....	69
4.4 Avoiding the Shmagency Objection	74
5. Clarificatory Remarks	76
5.1 Philosophy as Defence and Kantian Constructivism	76
5.2 Non-moral Normativity.....	79
5.3 Humean Constructivism.....	80
5.4 The Normative Question and Epistemology	81
6. Conclusion	82
PART II: THE PRACTICAL STANDPOINT: FIRST OR SECOND-PERSONAL?	84
Chapter 4: Why the First-Personal Account Fails	85
1. Introduction	85
2. The First-Personal Account	85
2.1 Practical Identities and the Form of Law.....	86
2.2 Interlude: Korsgaard on the Formula of Humanity.....	89
2.3 Action as Self-Constitution.....	92
3. Against the First-Personal Account	93
3.1 The Naïve First-Personal Reasoner	94

3.2 Agency and the Critique of Reason	97
3.3 The Fact of Reason and the Inescapability of Agency	100
4. Conclusion	102
Chapter 5: The Second-Personal Vindication of the Kantian Analysis of Obligation	103
1. Introduction	103
2. The Second-Personal Account	104
2.1 Obligation and Accountability	104
2.2 Accountability and Autonomy	108
2.3 Authority and Equal Dignity	112
2.4 Dignity and the Fact of Reason	116
3. The Irreducibility of Second-Personal Reasoning	119
3.1 Korsgaard: The Second-Person Within	120
3.2 Schapiro: Desires as Demands	124
3.3 Hanisch: Second-Personality and Self-Constitution.....	126
4. Conclusion	129
Chapter 6: Kantian Constructivism and the Second-Person Standpoint	130
1. Introduction	130
2. Respect, Persons, and the Law	130
2.1 Second-Personal Authority as Fundamental	131
2.2 Rights and Duties	136
2.3 Respect for Persons vs. Respect for the Law	141
3. The Optionality of the Second-Person Standpoint	145
4. Second-Personal Address, the Fact of Reason, and Duties to the Self	150
4. Conclusion	153
Chapter 7: Commitment and the Second-Person Standpoint	154
1. Introduction	154
2. Change on Commitment	154

2.1 The Volitional Character of Commitment	155
2.2 Chang's Brand of Voluntarism	160
3. The Second-Personal Account of Commitment.....	164
4. Commitment and Morality	169
4.1 Commitments as Obligations (to Oneself).....	169
4.2 Constraints on Commitment.....	172
5. Second-Personal Reasoning and Intrapersonal Morality	175
6. Practical Identities Revisited	178
7. Conclusion	180
CONCLUSION.....	182
Bibliography.....	183

Acknowledgments

This thesis is the product of three years of philosophical research, in which I have received vital support, guidance, and feedback from various people and institutions.

First of all, I would like to express my gratitude my first supervisor, Jens Timmermann, for countless conversations on the contents of this thesis, thoughtful feedback on my writing, as well as his constant advice and support. Thanks to his expertise and guidance, I have gained a deeper appreciation for the fine details of Kant's philosophy—and a greater awareness of how much of it I do not understand—than I would have thought possible in three years.

Next, I would like to thank my second supervisor, Ben Sachs, for his exceptionally high level of commitment, his thoughtful and detailed feedback on my writing, as well as many long conversations on my work and academic development.

I would like to thank Steve Darwall for facilitating my visit to Yale University in the Spring of 2017, providing extremely helpful feedback on my work, and supporting my academic development. This thesis speaks volumes about the tremendous philosophical inspiration that I have drawn from Steve's work.

For facilitating my visit to Humboldt University of Berlin in the Summer of 2017, welcoming me into his circle of incredibly talented practical philosophers, and providing helpful feedback on my work, I would like to thank Thomas Schmidt.

For helpful, encouraging, and rightfully critical feedback on written work, I would also like to thank Lisa Bastian, Joseph Bowen, Sarah Broadie, Ruth Chang, Rowan Cruft, Miguel Egler, Olof Leffler, Stefano Lo Re, Theron Pummer, Tobias Rosefeldt, Joe Slater, Justin Snedegar, Leif Wenar, and the participants of Thomas Schmidt's colloquium for practical philosophy.

For feedback on presentations, I would like to thank audiences at the Friday Graduate Research Seminar, the Dublin Graduate Philosophy Conference 2016, the Gothenburg Responsibility Conference 2016, the St Andrews Kant Colloquium, the Kant's Scots Workshop in Spring 2018, and the Future of Normativity Conference.

For financial support, I would like to thank the Aristotelian Society, the Centre for Ethics, Philosophy, and Public Affairs, the Erasmus+ programme, the St Andrews Philosophy Department, the Russell Trust at the University of St Andrews, and the Santander Universities Scheme.

Finally, I want to express my deep gratitude to Élyse Fischer, without whose loving support over the last three years I would not be the person that I am today.

References to Kant's Works

In referring to Kant's works, I use the following abbreviations:

<i>CPR</i>	<i>Critique of Pure Reason</i>
<i>CPrR</i>	<i>Critique of Practical Reason</i>
<i>G</i>	<i>Groundwork of the Metaphysics of Morals</i>
<i>MM</i>	<i>The Metaphysics of Morals</i>
<i>Rel</i>	<i>Religion Within the Bounds of Mere Reason</i>
<i>WE</i>	<i>An Answer to the Question: What is Enlightenment?</i>
<i>WOT</i>	<i>What Does it Mean to Orient Oneself in Thinking?</i>

Most references use the volume and page numbers of the Preußische Akademie der Wissenschaften edition of Kant's works. The only exceptions are references to the *Critique of Pure Reason*, which use the page numbers from the first and second editions, indicated by "A" and "B" respectively.

Most translations are taken from the *Cambridge Edition of the Works of Immanuel Kant*. The only exceptions are translations from the *Groundwork of the Metaphysics of Morals*, which are taken from Jens Timmermann's revised version of Mary Gregor's translation.

INTRODUCTION

In *A Theory of Justice*, originally published in 1971, John Rawls proposes a conception of justice for the “basic structure of society”, i.e., for “the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation” (1999: 6). Although his efforts are therefore aimed primarily at the domain of political philosophy, Rawls explicitly points out their relevance to moral theory more generally. In particular, he notes that he follows a distinct approach “which generalizes and carries to a higher level of abstraction the familiar theory of the social contract as found, say, in Locke, Rousseau, and Kant” (1999: 10).

In his later work—most notably in his *Dewey Lectures*, delivered at Columbia University in 1980—Rawls explains that his distinctive approach to moral theory is a “Kantian form of constructivism” (1980: 515). “The leading idea” of this form of constructivism “is to establish a suitable connection between a particular conception of the person and first principles of justice, by means of a procedure of construction” (Rawls 1980: 516).

Rawls argues that Kantian constructivism is uniquely suited to a central task of modern political philosophy: to find a workable conception of justice for the basic structure of a liberal democratic society (1980: 517-518). The challenge, according to Rawls, is that the conception of the person implicit in the political culture of such a society depicts its citizens as free and equal moral persons. This conception entails that, while persons are capable of social cooperation under a public set of principles of justice, they also have a capacity for—and an interest in—freely choosing, revising, and pursuing their own conception of the good. This means that a practicable conception of justice for a liberal democracy cannot be based on any particular conception of the good. For example, it would not be practicable to base a conception of justice for a liberal democracy on a particular religious view of the good life. The citizens of such a society do not regard themselves as beholden to any such view, even if they happen to agree with it, and will thus view any conception of justice based on it as arbitrary.

Rawls contends that his Kantian constructivism is uniquely suited to overcome this challenge because it suggests that the conception of justice for a liberal democracy is to be *constructed* on the basis of the liberal democratic conception of the person itself. Rawls’s Kantian constructivism supposes that a conception of justice that gives expression to this conception of the person will be practicable in a liberal democratic society because it spells out principles by which the social life of free and equal moral persons as

such ought to be regulated. In order to arrive at such a conception of justice, according to Rawls, we need to devise a procedure for the construction of principles of justice, and the constraints of this procedure are to express the different elements of the liberal democratic conception of the person. The constructive procedure that Rawls suggests is a hypothetical contract situation, the “original position”, in which the parties to the contract are envisaged as freely advancing their interests under a constraint on information that guarantees their impartiality (1980: 522).

It is important to note that, in the Kantian constructivist framework, the constructive procedure does not play the role of a device that *tracks* the correct principles of justice. Indeed, prior to the construction of principles through the appropriate procedure, there are no such principles. Rawls writes (1980: 565):

The parties in the original position do not agree on what the moral facts are, as if there already were such facts. It is not that, being situated impartially, they have a clear and undistorted view of a prior and independent moral order. Rather (for constructivism), there is no such order, and therefore no such facts apart from the procedure of construction as a whole; the facts are identified by the principles that result.

Rawls contrasts this aspect of Kantian constructivism with the position he labels “rational intuitionism” and which holds that “first principles of morals [...] are self-evident propositions [...] fixed by a moral order that is prior to and independent of our conception of the person and the social role of morality” (1980: 557).

Rawls points out that Kantian constructivism and rational intuitionism can be read as providing different interpretations of moral objectivity (1980: 554-572). While rational intuitionism regards moral principles as objective insofar as they are self-evident propositions that hold independently of our conception of the person, Rawls’s Kantian constructivism regards moral principles as objective insofar as they are the outcome of a constructive procedure that is based on a conception of the person that is *shared* among the citizens of Western liberal democracies. This does not mean that citizens *choose* the principles of justice that govern the basic structure of their society (1980: 567-568). The conception of justice that is constructed via the original position is objectively valid for citizens in a Western liberal democracy, i.e., it’s principles *bind* them independently of whether they choose to live by them, because it is the best expression of *their* conception of the person (cf. Korsgaard 2008e: 320-321).

Although Rawls’s remarks are intended to characterise his approach to the principles of social justice for a Western liberal democracy, they have inspired several

philosophers to develop theories that extend the basic tenets of his Kantian constructivism to morality as a whole—or, in some cases, all of normativity (e.g., Korsgaard 1996). Indeed, Rawls himself proposes such an extension of his Kantian constructivist framework as an interpretation of Kant’s view (Rawls 2000: 235-252).

Nowadays, the label ‘Kantian constructivism’ denotes a family of different views that are inspired by Rawls’s position to varying degrees (cf. Bagnoli 2013a, 2017, Lenman and Shemmer 2012). Yet, there is a tendency—at least among the most prominent contemporary versions of Kantian constructivism—to converge on some key notions and to share what we might call a general philosophical outlook. According to these versions of Kantian constructivism, moral principles can be derived (or ‘constructed’) from a formal characterisation of our practical standpoint (e.g., Bagnoli 2013b: 2, 2016, Korsgaard 1996, 2008e, Street 2010). Our practical standpoint is the point of view from which we deliberate about what to do. This is also sometimes referred to as our *practical reason* or our *will* (e.g., Korsgaard 2007, 2009b).

Mirroring Rawls’s distinction between Kantian constructivism and rational intuitionism, contemporary authors frequently contrast Kantian constructivism with *moral realism* (e.g., Bagnoli 2016, Korsgaard 2008e, O’Neill 1989a, Shafer-Landau 2003, Sensen 2013, Stern 2013, Street 2010, Wood 2008). Like Rawls, contemporary Kantian constructivists tend to argue that their view offers an approach to moral objectivity which renders it more attractive than both moral realism and various forms of anti-realism (e.g., Korsgaard 2008e, O’Neill 1989e). By depicting moral principles as derived from a formal characterisation of our practical standpoint, Kantian constructivism supposedly avoids moral realism’s reliance on an independent order of moral facts, which some theorists deem problematic (e.g., O’Neill 1989e, 2015b, 2015d). At the same time, by regarding moral principles as derived from a *formal* characterisation of our practical standpoint, as opposed to its *contents* (such as our actual judgments, preferences, or cultural traditions), Kantian constructivism supposedly avoids the anti-realist mistake of rendering morality subjective, conventional, or otherwise contingent in some objectionable sense (e.g., Bagnoli 2013b: 2, Korsgaard 2008e: 309).

Since moral realism and its competitors are metaethical views, this characterisation implies that Kantian constructivism provides a position in metaethics. It is important to note, however, that Kantian constructivism is usually taken to have implications for first-order normative theory as well. After all, to say that normative facts in some domain are derived from a formal characterisation of some practical standpoint is to say something about how *correct* normative judgments within that domain are to be reached. Indeed, some Kantian constructivists do not take their view to make a

contribution to metaethics at all and, instead, intend to provide a first-order normative theory only (e.g., Scanlon 1989, 2012). In this thesis, however, I shall focus on versions of Kantian constructivism that aim to provide a metaethical position.

As the reader might have noticed already, the above characterisation of Kantian constructivism is potentially vulnerable to a host of objections or, in any case, is so vague as to require further specification. And, indeed, Kantian constructivism has been challenged or attacked on various grounds since Rawls's introduction of the view into the contemporary debate. However, my suspicion is that many of the objections levelled against Kantian constructivism are based on misconceptions about the view and its ambitions. For example, it is sometimes supposed that Kantian constructivism must contradict moral realism if it is to have anything interesting to say within the domain of metaethics. This leads some theorists to conclude, from the lack of contradictions between Kantian constructivism and moral realism, that the former does not have anything of interest to say about metaethical issues. As I shall argue in Chapter 1, this is a misconception because Kantian constructivism is a view about the source of moral normativity—a metaethical issue that moral realism, on at least some characterisations of that view, does not address.

My overall objective in this thesis is to provide a *restatement* of Kantian constructivism which clears up some of these misconceptions and thus avoids some of the objections and problems that have haunted the view. The term 'restatement' is intentionally broad since it is supposed to cover both *departures* from earlier versions of Kantian constructivism and mere *clarifications* of the way in which these earlier versions were stated. I believe that this thesis contains both, but since earlier statements of Kantian constructivism often leave room for interpretation, I am not always sure which is which. I want to avoid getting caught up in such interpretative issues as much as I can and, instead, focus on my attempt to provide a clear statement of the version of Kantian constructivism that I favour.

For example, one of the most influential and developed versions of Kantian constructivism in the contemporary literature is that defended by Christine Korsgaard (e.g., 1996, 2009a). Thus, Korsgaard's work serves as a natural starting point to the discussion in many of my chapters. Whether the position that I develop over the course of these chapters qualifies as a clarification, a defence, a revision, or even a rejection of her views will depend on the correct interpretation of her writings. Although I try to be conscientious in reading her work, the question of whether the positions that I attribute to her are really hers is secondary for the purposes of this thesis. Again, my efforts are focused on providing a clear statement of my version of Kantian constructivism.

I should also note that this thesis does not aim to provide an interpretation of Kant's moral theory. In this context, I would like to second the following remark by Rawls (1980: 517):

[T]he adjective 'Kantian' expresses analogy and not identity; it means roughly that a doctrine sufficiently resembles Kant's in enough fundamental respects so that it is far closer to his view than to the other traditional moral conceptions that are appropriate for use as benchmarks of comparison.

Thus, that a certain claim is endorsed by Kant does not, *by itself*, count as a reason to endorse it, as far as my restatement of Kantian constructivism is concerned. Of course, I shall take into account Kant's views and arguments where they seem relevant to the discussion for philosophical reasons.

One might wonder in how far my position deserves the label 'Kantian constructivism' if it is not an interpretation or defence of Kant's, Korsgaard's, or, for that matter, Rawls's position. One might expect me to answer that my position is true to the *spirit* of the Kantian constructivist position. But while I think that this is true, I do not argue for this claim directly. Instead, I leave it to my readers to judge for themselves how well the position I advance resonates with the spirit of Kantian constructivism.

Before I provide an overview of the thesis, two more qualifications are in order. First, I should note that, while I think that Kantian constructivism entails a position in normative ethics as well as in metaethics, I shall mostly bracket the question of how its normative claims are to be spelled out. In particular, I shall not address the important and difficult question of how the Categorical Imperative, which tests maxims by their *form*, can fully determine the *content* of morality. Second, as we shall see in Chapter 1, the version of Kantian constructivism that I propose in this thesis depends on a specific, Kantian analysis of obligation. Kantian constructivism, as I restate it, therefore turns out to be a view about the source of the normativity of *moral obligation*. Whether it thereby qualifies as a view about all of *morality* depends on whether morality is exhausted by obligations, a question I shall not attempt to settle in this thesis (although I shall remark on it occasionally).

The thesis is divided into two parts. Part I is concerned with Kantian constructivism's place in metaethics. In Chapter 1, I argue that Kantian constructivism's contribution to metaethics is the notion that the form of our practical standpoint is the *source* of morality's normativity. The argument for this notion relies upon a Kantian analysis of obligation.

In Chapters 2 and 3, I argue that Kantian constructivism's position on the source of moral normativity has an advantage over competing positions when it comes to answering the so-called Normative Question. In Chapter 2, I suggest that this question should be interpreted as asking why we are rationally required to do what we believe we are morally obligated to do. In Chapter 3, I argue that Kantian constructivism is uniquely suited to answer question because it locates the source of moral normativity in the very form of practical reason. But since this answer to the Normative Question also appeals to the Kantian analysis of obligation, it does not provide an independent argument for Kantian constructivism.

Since the case for Kantian constructivism thus crucially depends on the Kantian analysis of obligation, Part II focuses on providing a motivation for that analysis. I discuss two different accounts that purport to elicit this analysis from the practical standpoint of ordinary agents, the first-personal and the second-personal account.

The first-personal account purports to elicit the Kantian analysis from the standpoint that we must take up, *qua* agents. In Chapter 4, I argue that this account fails.

The second-personal account purports to elicit the Kantian analysis from the second-person standpoint, i.e., the standpoint that we take up when we address demands person-to-person. I motivate this account in Chapter 5. I reject several arguments to the effect that the second-personal account supports the first-personal account because the constraints on agency as such require us to take up the second-person standpoint.

In Chapter 6, I reply to several objections that might be levelled against the second-personal account from a Kantian constructivist point of view. Most of these objections falsely presuppose that the second-personal account abandons the notion that moral obligations have their source in the form of our practical reason.

In Chapter 7, I address the criticism that the second-personal account depicts morality as dependent upon social interactions, and thereby releases isolated individuals from moral obligations. I argue that we do not only reason second-personally when interacting with others but in certain self-regarding contexts as well, such as when we commit ourselves to a project. Hence, the second-person standpoint is not a social practice but the locus of a distinct mode of practical reasoning which governs interpersonal and intrapersonal relations alike.

PART I: KANTIAN CONSTRUCTIVISM AND ITS PLACE
IN METAETHICS

Chapter 1: The Sources of Normativity

1. Introduction

It is relatively uncontroversial that Kantian constructivism has implications within first-order normative theory. But some theorists suggest that Kantian constructivism also includes second-order claims about morality, or indeed all of normativity. In other words, they take Kantian constructivism to entail a metaethical position. Frequently, these theorists regard Kantian constructivism as offering an alternative to moral realism as well as to “relativism, skepticism, subjectivism, and all of the various ways that ethics might seem hopeless” (Korsgaard 2008e: 309).

Yet, whether Kantian constructivism provides a metaethical position of its own is highly controversial. Critics argue that Kantian constructivism, by itself, does not have the resources to suggest any novel and distinct answers to metaethical questions. Indeed, it seems to be compatible with all the more traditional views in metaethics, including moral realism. Accordingly, some authors suggest that Kantian constructivism’s claims are confined to normative ethics.

In this chapter, I defend Kantian constructivism against the charge that it does not provide any distinct answers to metaethical questions. I do so by clarifying and motivating the—initially elusive—notion that Kantian constructivism takes up a distinct position on the *source* of morality’s normativity. I argue that the question of the source of something’s normativity is a metaethical question, albeit one that is distinct from the questions addressed by the views with which Kantian constructivism is often contrasted, such as moral realism. In particular, I contend that views about the sources of normativity belong to the *metaphysics* but not to (what I shall call) the *ontology* of normativity. Following Ruth Chang, I suggest that the source of something’s normativity is that which *grounds* its having normative force. Subsequently, I suggest that Kantian constructivism holds that morality’s being normative is grounded in our having autonomous wills.

I should note that my aim in this chapter is not to come up with definitions of Kantian constructivism, constructivism, or moral realism that capture how these positions generally are or should be understood. Instead, my goal is to outline, and to some extent defend, Kantian constructivism as I understand it.

The chapter has the following structure. Section 2 sketches some of the challenges to Kantian constructivism’s ambition to provide a novel and distinct metaethical position. Section 3 introduces the question of the source of something’s normativity as a metanormative inquiry that must be distinguished from questions about the ontology

normativity. Section 4 spells out Kantian constructivism's claims about the source of the normativity of morality in greater detail and provides an argument for understanding them as claims about the grounding of moral normativity.

2. Kantian Constructivism's Metaethical Ambitions

Following some of the most influential advocates of Kantian constructivism, I characterise the metaethical component of Kantian constructivism as providing an account of the *source* of moral normativity (e.g., Bagnoli 2016, Korsgaard 1996, Reath 2006b). However, as I show in this section, it is surprisingly difficult to pinpoint the metaethical question that such an account might answer. This difficulty manifests itself in the fact that, upon closer examination, it is not clear that Kantian constructivism's claims about the sources of normativity conflict with traditional positions in metaethics, such as moral realism and non-cognitivism. Note that this section does not give a complete survey of past attempts to formulate Kantian constructivism as a metaethical position. It only serves to illustrate the pitfalls that must be avoided if such an attempt is to succeed.

2.1 Metaethics vs. Normative Ethics

Korsgaard provides the following sketch of the supposed Kantian constructivist position in metaethics (1996: 19):

Kantian constructivists [...] believe that the source of the normativity of moral claims must be found in the agent's own will, in particular in the fact that the laws of morality are the laws of the agent's own will and that its claims are ones she is prepared to make on herself.

Now, if Kantian constructivism provides an alternative to traditional positions in metaethics by making claims about the *source* of morality's normativity, one would expect these positions to make conflicting claims about the same subject. However, as Nadeem Hussain and Nishi Shah point out, metaethics is traditionally characterised as the field that deals with semantic, epistemological, and metaphysical aspects of morality (2006, 2013). They write (2006: 268):

A metaethical account offers an interpretation of the normative claims that are offered as answers to [first-order normative] inquiries (for example, that it is morally right to maximize utility or that one has reason to do those actions that are morally right), aiming to tell us what these claims mean, whether they involve

metaphysical commitments, if so, what these commitments are, and whether and how we acquire knowledge of these normative claims.

By contrast, they argue, an inquiry into the source of something's normativity might be concerned with one of two questions, both of which belong to first-order normative theory.

First, an inquiry into the sources of normativity could be interpreted as asking which features of an act *make* it have a certain normative status. In this sense, the fact that brushing one's teeth reduces plaque might be said to be the source of the goodness of brushing one's teeth. However, this is not a metaethical claim; for it does not address the questions "of what it means to say that reducing plaque is good, what metaphysical commitments such a judgment involves, or how we come to know that brushing one's teeth is good" (2006: 271). Instead, it addresses the questions of which things are good, and why, where the 'why' question is answered by a further normative claim (that reducing plaque is good).

Second, a claim about the source of the normativity of morality might attempt "to *place* morality within practical reason, explaining whether we have reason to do what morality demands and, if so, whether these reasons are derived from another branch of practical reason" (2006: 267, italics added). However, according to Hussain and Shah, this is not what metaethics is about; again, its subject matter are the metaphysics, epistemology, and semantics of moral thought. They write (2006: 268-69):

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.

Now, one might disagree with Hussain and Shah's definition of metaethics. In particular one might want to expand metaethics to include theorising about the placement of morality within practical reason. One might then argue that Kantian constructivism engages in metaethical theorising precisely because it attempts to place morality within practical reason. On such a reading, Kantian constructivists might be understood as arguing that our reasons to be moral derive from another set of reasons, e.g., our reasons to sustain agential integrity (cf. Korsgaard 1996, 2009a). Opponents of such a position might argue that moral reasons are foundational to practical reasoning (cf. Prichard 1912). However, Kantian constructivism would thereby not be taking a metaethical position in the sense of 'metaethical' in which moral realism and its competitors are metaethical positions. This can be seen from the fact that the foundational role of morality in practical reason

is not only endorsed by many moral realists but also by some non-cognitivists (e.g., Blackburn 1998: 237). In any case, I do not think that it is promising to interpret Kantian as attempting to place morality within practical reason.

2.2 *The Function of Normative Concepts*

Now, it might be argued that Kantian constructivism's source claims suggest a theory of "what it is to think a normative thought" after all (*pace* Hussain and Shah 2006: 268-69). In particular, Korsgaard criticizes both realists and non-cognitivists for presupposing that "it is the function of all of our concepts, or anyway all of our authentically cognitive concepts, to describe reality" (2008e: 306). By Korsgaard's lights, this presupposition leads moral realists to conclude that normative concepts must have the role of describing reality, whereas it leads non-cognitivists to conclude that normative concepts are not 'authentically cognitive'. However, according to Korsgaard, there is an account of normative concepts on which they are neither assigned the role of describing reality, nor robbed of their status as 'authentically cognitive'. In particular, she writes (2008e: 322):

[A]ccording to constructivism, normative concepts are not (in the first instance—a caveat I will explain below) the names of objects or of facts or of the components of facts that we encounter in the world. They are the names of the solutions of problems, problems to which we give names to mark them out as objects for practical thought.

As we shall see in the next chapter, Korsgaard's remarks on the function of normative concepts provide an important clue when it comes to interpreting the so-called Normative Question. For now, however, the question is whether these remarks suffice to show that Kantian constructivism makes *metaethical* claims. And the answer to that question is no.

The problem is the very 'caveat' that Korsgaard hints at in the quote above. She is referring to the fact that, even if her account of concepts is true, "normative concepts may after all be taken to refer to certain complex facts about the solutions to practical problems faced by self-conscious rational beings" (2008e: 325, cf. Copp 2013: 123). Accordingly, she concedes that "considered in one way, constructivism and realism are perfectly compatible" (*ibid.*). She immediately adds that viewing normative concepts in this way is uninteresting because it takes up a third-personal point of view and "it is only viewed from the perspective of those who actually *face* those problems in question that these truths will appear normative" (*ibid.*). However, since moral realism is a metaethical account, its aim is not to ask what normative truths there are, but to provide an account

of what it is for a truth to be normative in the first place. And Korsgaard's lack of interest in this aim does not suffice to show that her Kantian constructivist account provides an alternative to moral realism.

Indeed, the same goes for non-cognitivism. Korsgaard concedes that “[v]iewed from outside of that [practical] perspective, those who utter these [normative] truths will appear to be simply expressing their values” (2008e: 325). Again, she adds that this is not what is most *interesting* about normative concepts. However, as Hussain and Shah observe, this is precisely what non-cognitivists are most interested in (2013: 103). And the fact that Korsgaard has different interests than non-cognitivists, by itself, does not show that her account offers an alternative to their views.

2.3 *Stance-Independence*

Thus, it can be hard to see how Kantian constructivism conflicts with established metaethical views such as moral realism and non-cognitivism. Indeed, as David Copp points out, many Kantian constructivists happen to affirm central tenets of moral realism, such as the claim that there are moral properties to which moral predicates refer (2013: 120-21).

In a similar vein, on Russ Shafer-Landau's characterization, Kantian constructivism partly agrees with moral realism by endorsing the “reality” of the moral domain (2003: 14). Nevertheless, Shafer-Landau proposes what he takes to be a sharp distinction between moral realism and Kantian constructivism (*pace* Copp, Hussain and Shah). Specifically, he argues that the former, but not the latter, endorses the “*stance-independence* of moral reality” (2003: 15). The thesis of stance-independence holds that “the moral standards that fix the moral facts are not made true by virtue of their ratification from within any given actual or hypothetical perspective” (*ibid.*, italics deleted).

This construal seems to resonate with the way Kantian constructivists themselves understand their opposition to realism. Korsgaard, using “procedural moral realism” as a term for Kantian constructivism, writes (1996: 36-37, italics added):

The procedural moral realist thinks that there are answers to moral questions *because there are correct procedures* for arriving at them. But the substantive moral realist thinks that there are correct procedures for answering moral questions because there are moral truths or facts.

Similarly, Carla Bagnoli writes that Kantian constructivism's main concern “is not whether there is a moral ontology, but whether and how it is related to practical reason” (2013c: 159, cf. 2014, 2016). Thus, it seems that Kantian constructivism's claims about

the source of moral normativity are not about the question of whether there is a moral reality or of what ontological, epistemological, and semantic status such a reality might have. Rather, they seem to be concerned with the question of what *order* (of priority or dependence) holds between moral normativity on the one hand and the procedure provided by the form of our practical standpoint on the other hand.

It will turn out that this is on the right lines (see Section 3 below). However, it is still not clear that considerations of the source of moral normativity are to be classified as metaethical. Shafer-Landau's suggestion that these considerations concern the thesis of stance-independence is somewhat ambiguous in this respect. As we have seen, he interprets the question of stance-independence as the question of whether moral standards are "made true by virtue of their ratification from within [some] perspective" (2003: 15). Yet, one might argue that views of what makes moral standards *true* are best understood as part of first-order normative theory, not metaethics. Following Hussain and Shah, one might interpret these considerations as being concerned with the *features* (e.g., reducing plaque) that make an act (e.g., brushing one's teeth) have a certain normative status (2006: 270-71). On this construal, a Kantian constructivist might disagree with, say, a utilitarian over whether it is an act's maximization of happiness or its endorsement from the practical standpoint that 'makes' it good.

Or, again, one might interpret considerations of stance-independence as attempting to 'place' morality within practical reason, by stating whether our reasons to be moral derive from another, more fundamental part of practical reason (Hussain and Shah 2006: 267). But, as we saw in Section 2.1, if Kantian constructivism was characterised as placing morality within practical reason, this would not make it a metaethical view, at least not in the sense in which moral realism and its competitors are metaethical views.

More generally, the nature of the relation that Kantian constructivism takes to hold between moral normativity and our practical standpoint is not entirely clear. So far, we have not been given reason to think that it is not a sort of *justificatory* relation that might hold between one normative standard and another, more basic one. The notion of stance-independence does not help, at least without further clarification, because it might plausibly be classified as belonging to first-order normative theory (cf. Blackburn 1998: 311). In the following section, I therefore spell out what I take claims about the sources of normativity to amount to without relying on the notion of stance-independence (although I return to it in Section 3.3). I argue that the question of the source of

something's normativity is a metaphysical—although not, as I shall explain, an ontological—question about normativity.

3. What are Sources of Normativity?

We have seen reasons to be sceptical about Kantian constructivism's metaethical ambitions. However, I do not think that we should conclude that Kantian constructivism's claims are confined to normative ethics (*pace* Darwall et al. 1993, Horgan and Timmons 2006, Hussain and Shah 2006, 2013, Scanlon 2012, Timmons 2003). Instead, I am going to suggest that claims about the sources of normativity address a respectable metanormative question, albeit one that carefully needs to be distinguished from other questions in this area.

Before I start, I should briefly explain what I mean by 'normativity' or 'normative force' in this context. There are at least two different things that we might mean by saying that morality is normative. First, we might mean that morality involves requirements that can be expressed by sentences containing words like 'should' or 'ought'. In this sense of 'normative', etiquette and chess are just as normative as morality. In particular, we might say that etiquette and chess, as well as morality, are *normative* systems because they are systems *of norms*.

Second, by calling morality normative we might intend to accord it "a special dignity and necessity", as Philippa Foot puts it (1972: 308). It is this second sense of 'normative' that I have in mind when I use 'morality's normative force', 'moral normativity', and related expressions. Admittedly, this sense is somewhat elusive. It is often expressed by saying that moral requirements give us *reasons for action* (e.g., Darwall 1990: 258, Foot 1972: 309, Wallace 2004: 454). But this way of expressing it has its limitations. For one thing, it is difficult to specify the relevant sense of 'reason' without relying on the very sense of normativity that is to be expressed (cf. Parfit 2011a: 31, Scanlon 1998: 17). For another, it is questionable whether the terminology of reasons is well-suited to capture the obligatory force of morality (see Section 4 below). Nevertheless, I shall leave this issue here and turn to the question of what I mean by 'sources of normativity'.

3.1 Sources as Grounds

Following Chang, by a 'source of normativity' I will understand "*that in virtue of which* something has normative force" (2009: 243). Chang also refers to the source of a consideration's normativity as its metaphysical "fount" or "ground" (2013a: 165). However, Chang's account does not depend on the defensibility of any of the specific accounts that have been advanced in the metaphysical debate over "grounding" (e.g., Fine 2001, Rosen

2010). For Chang's notion of grounding to be tenable, it only needs to be plausible that there is an *explanatory* relation that cannot be reduced to certain other forms of explanation (Chang 2013a: 164-165).¹

First of all, the source of something's normativity provides an explanation of its having normative force, but not by being the *cause* (if there is such a thing) of its having normative force (Chang 2013a: 164).

Second, the relevant explanatory relation is not that of *modal covariance*, which is taken to hold between a normative property and its subvening base. The explanation provided by the source of something's normativity is different and, in a sense, deeper than that provided by a subvening base (Chang 2013a: 164). To see this, imagine a version of divine command theory according to which God's authority is the source of morality's normativity but the content of morality is captured by classical utilitarianism.² In this view, facts about pleasure and pain are the subvening base of moral facts, but they are not the source of their normativity. God's authority *explains* the normativity of morality in a way in which hedonic facts do not. In fact, it even explains why there is a supervenience relation between moral facts and hedonic facts in the first place.

Finally, the relevant explanatory relation is not that of one normative consideration's *subsumption* under a more general, or more fundamental, normative principle. Instead, when we inquire about the source of a consideration's normativity we are asking about its *metaphysical* ground (Chang 2013a: 164-165).

Now, one way in which something might be the source of a consideration's normativity is by *constituting* that consideration's having normative force. Put another way, a consideration's having normative force might *consist* in this source's obtaining. Analogously, "[t]he fact that it's H₂O grounds the fact that it's water in that its being H₂O constitutes its being water, and its being water consists in its being H₂O" (Chang 2013a: 165).

Chang suggests that there might be another way for the source relation to obtain. In particular, something might be the source of a consideration's normativity by way of *creation* or *construction* (2013a: 166). To illustrate this suggestion, she notes that the fact that the laws of causality hold might be explained by God's commanding that they do.³ Now, the relation that holds between God's commands and the laws of causality

¹ For another account that links the notion of sources of normativity to grounding, see Väyrynen (2013).

² This is a slightly altered version of Chang's example (2013a: 164).

³ This is just one of three examples that Chang uses (cf. Chang 2013a: 166).

cannot be that of *causation*. To begin with, this relation is supposed to be outside of, and prior to, the realm of causality. More importantly, God seems to give rise to the laws of causality in a more *direct* or *immediate* fashion than the notion of causation suggests. Hence, it is plausible to construe the relation between God and the laws of causality as one of grounding. Nevertheless, to conceive of the fact that the laws of causality hold as constituted by, and thus consisting in, the fact that God commands them seems to tie the relata together *too closely*. It therefore seems best to say that God's commands ground the laws of causality without constituting them, namely by metaphysically creating or constructing them.

Now that we have seen what claims about the sources of normativity amount to, one might wonder how such claims might be established. Fortunately, the idea that claims about the sources of normativity are claims about what grounds the fact that something is normative helpfully suggests some desiderata. To begin with, if sources are grounds, then a plausible source claim should have *explanatory* power. This suggests that the presence of the source of something's normativity is a sufficient condition for its being normative. In other words, the source should not obtain in cases in which the normativity does not.⁴ Furthermore, by linking sourcehood to explanation, the above account suggests that the source relation is *asymmetric* (Väyrynen 2013: 163). That is, if something grounds a consideration's being normative, it cannot in turn be grounded in the fact that the consideration is normative. Finally, if source claims are claims about grounding, they should advance an explanation that proceeds neither (a) in causal terms, nor (b) in terms of the subvenience base of a given normative fact, nor (c) by subsuming a normative fact under a more general normative fact or principle.

3.2 *The Sources vs. the Ontology of Normativity*

We have seen that the question of the sources of normativity is concerned with the *metaphysics* of normativity. However, it is important to distinguish this question from (what I shall call) questions about the *ontology* of normativity.

⁴ If there are "enablers" for explanations, i.e., conditions that need to be in place for a candidate explanans to explain an explanandum, then it is more accurate to say that the obtaining of the candidate explanans, in conjunction with the obtaining of the enablers, is a sufficient condition for the explanandum to obtain (cf. Dancy 2006: 45-49). However, in the main text, I shall continue to speak as if the obtaining of the explanans was sufficient by itself. I therefore ask readers to insert the relevant qualification if they deem it appropriate.

Following Chang, I distinguish between two separate questions about the ontology of normativity. Both are clearly to be distinguished from the question of the source of something's normativity. The first ontological question asks what kinds of *things*, fundamentally, have normative force. In Chang's words, it asks about the ultimate "bearers of normativity" (2009: 243). By contrast, the source question asks what grounds the fact that something bears normativity, irrespective of the bearer's ontology. A second ontological question about normativity is concerned with what Chang refers to as "the nature of normativity" or "normativity's essential features" (ibid.). This question is concerned with the ontological status of *normativity itself*. That is, it inquires what kind of *force* normativity is. The source question, by contrast, asks wherein the fact that something *has* normative force is grounded, whatever the ontological status of that force may be.

These two ontological questions are sometimes conflated with one another and with the source question because, presumably for historical reasons, certain combinations of answers to these three questions are more common than others.

Consider the position that normativity is an irreducibly normative force, which we might call *non-reductivism*. Next, consider the position that the bearers of normativity are normative properties instantiated in the real world, which we might call *realism*.⁵ It is quite common to combine non-reductivism with realism, and thus endorse the position we might call *non-reductive realism*, although non-reductivism and realism do not logically entail one another. One could be a realist and reduce normativity to some non-normative force, endorsing *reductive realism*, or one could be a non-reductivist while rejecting that there are any in-the-world normative properties, e.g., by endorsing non-reductive "projectivism" (cf. Blackburn 1998). The prevalence of non-reductive realism threatens to conceal these possibilities.

For similar reasons, the distinction between the sources and the ontology of normativity sometimes goes unnoticed. When it comes to explaining *why* something has normative force, some non-reductive realists deny that there is any meaningful question that is not already answered by their view of the bearers and ontological status of normativity (Parfit 2006, 2011b: 415-420, Scanlon 2014: 14). Yet, by denying the need for an explanation of the normativity of irreducibly normative properties, these theorists effectively argue that these properties themselves already provide all the explanation of their normativity that we could possibly hope for. By adopting this view, however, these theorists go beyond non-reductive realism and, additionally, endorse a position we might

⁵ My use of "bearers" of normativity might diverge from Chang's.

call *source externalism*, which “locate[s] the source of normativity in a realm of external, irreducibly normative facts” (Chang 2009: 244).

To see this, consider that non-reductive realism could also be combined with *source internalism*, i.e., the view that considerations have normative force in virtue of their relation to our motivational states.⁶ It is often assumed that source internalists are committed to the view that normativity is reducible to some kind of motivational force and that its bearers are desires and other mental states (Chang 2009: 243-44). However, some source internalists explicitly reject this position (e.g., Darwall 1990, 1995, 2017, Street 2017).⁷ Sharon Street, for example, endorses a version of non-reductivism (2017: 131, italics added):

Internalists [...] agree that there are reasons in the *standard normative sense* [i.e., these reasons are irreducibly normative], but deny that there are any normative reasons of a certain sort—namely ones that exist *independently* of whether there is a sound deliberative route to the conclusion that they exist from the subjective motivational set of the agent whose normative reasons are in question.

Admittedly, source internalists like Street sometimes describe the disagreement between their position and source externalism as a matter of *substantive* normative theory (e.g., Darwall 2017: 261, Street 2017: 130). However, I do not think that this is the best way of capturing the debate. After all, substantive normative theory is concerned with the question of what we ought (or have reason) to do. And we would usually not regard, ‘Only those things that stand in a certain relation to our subjective motivational set’, as a satisfactory answer to that question. Importantly, the issue with this answer is not just that it is uninformative; rather, it entirely misses the point of the question. Even if, say, Street and Parfit completely agreed on what we ought (or have reason) to do, they could still be at odds about whether internalism is true. This is because their disagreement is not about substantive normative questions but about the *sources* of normativity.

The distinction between the ontology and the sources of normativity can also be illustrated by considering the position of *source voluntarism*, i.e., the view that considerations have normative force in virtue of their relation to our volitional acts or states. It

⁶ Source internalism is to be distinguished from “judgment internalism”, the view that “a *judgment* that a reason exists to do something entails some motivation in the judge to act accordingly” (Darwall 1990: 262).

⁷ Such a position might be attributable to Nagel (1970) and Williams (1981).

is sometimes assumed that source voluntarists are committed to a *volitional conception* of normativity, according to which normativity is a volitional force and its bearers are volitional acts or states. However, Chang’s own account of the sources of normativity can serve as a counter-example here (2009, 2013a, 2013b). Chang proposes a *hybrid voluntarism* about the sources of normativity. In particular, she suggests that, while most of our normative reasons are simply “given” in the sense that they exist independently of what we will, we can sometimes give ourselves reasons by exercising the “normative powers” of our will (Chang 2013b: 101-103, also see Chapter 7 below). Chang thus proposes *source externalism* about some reasons, the ones she calls “given reasons”, and *source voluntarism* about others, which she calls “voluntarist reasons” (ibid.). While she is committed to non-reductive realism regarding the ontology of given reasons (because source externalism implies that there are irreducibly normative properties), she remains neutral about the ontology of voluntarist reasons (2009: 270-271). That these reasons have their source in our own will does not rule out that their normativity is a matter of irreducibly normative properties—if it is, then Chang’s account implies that our will has the capacity to “create” such properties (ibid.). Neither is the notion of voluntarist reasons incompatible with motivational or volitional conceptions of the ontology of normativity.⁸

3.3 Kantian Constructivism and Moral Realism

I hope to have shown that questions about the sources of normativity are respectable metanormative questions that can and should be distinguished from questions about the ontological status and bearers of normativity. Therefore, by taking a stance on the source of moral normativity, Kantian constructivism is not committed to a particular view of the ontological status or bearers of moral normativity. Does this mean that Kantian constructivism is compatible with the position called *moral realism*?

Above, I have used the label ‘realism’ to refer to a view of the bearers of normativity. If moral realism is just the application of that view to the domain of morality, then Kantian constructivism is compatible with it. However, one might argue that moral realism encompasses more than that.

⁸ Chang’s account also serves as a further illustration of the difference between the source question and *substantive* normative questions. Her hybrid account of the sources of normativity allows her to say that there can be two distinct reasons that have the same content; for we can exercise our normative powers to will a fact to be a reason to φ even if this very fact is already a (given) reason to φ (2009: 257-258).

To begin with, moral realism is sometimes taken to involve claims about the epistemology and semantics of normativity (e.g., Copp 2013: 120-121). Now, by advancing a position on the source of moral normativity, Kantian constructivism does not commit to a view of the semantics of normativity. Hence, Kantian constructivism does not conflict with moral realism on this count. Whether Kantian constructivism engages in epistemological theorising about normativity is a more difficult question, which I shall take up again in Chapter 3. At any rate, it is arguable that claims about epistemology and semantics are not essential to moral realism. With respect to semantics, William FitzPatrick remarks (2009: 746-47, cf. Miller 2009):

If one holds that there are objective facts about the wrongness of slavery, for example, then one ought to count as an ethical realist regardless of whatever views one might hold about the semantics of the contingent forms of discourse that have evolved among language users.

I take it that one might make a similar argument against including epistemological claims in one's definition of moral realism. After all, it also seems inessential to one's commitment to 'objective facts about the wrongness of slavery' that one thinks that our epistemic faculties trace such facts.

We might say, then, that moral realism is exhausted by claims about the *metaphysics* of moral normativity. But does it include a position about the *sources* of that normativity or is it confined to claims about ontology? Proposals according to which moral realism endorses the thesis of stance-independence might be taken to imply that it does include such a position, namely source externalism (cf. Shafer-Landau 2003: 14-15, see Section 2.3 above). We might say that source externalism regards normative properties as stance-independent in the sense that these properties provide the source of their own normativity, independently of any practical standpoint. However, the problem with including this view in one's definition of moral realism is that several theories that are labelled 'moral realist' in the literature do not endorse stance-independence (Copp 2013: 121). These include sentimentalist theories as well as certain naturalist versions of moral realism (e.g., Copp 2005, Driver 2013).

Now, one could accommodate the latter group of views under the broad heading of 'moral realism' while reserving the label 'robust moral realism' for those theories that endorse stance-independence (Fitzpatrick 2008: 166, 2009: 749). However, I doubt that dividing views into more or less robust versions of realism, depending on their acceptance of stance-independence, is helpful. It might be taken to suggest that stance-independent realists are *more* realist than others, which is misleading since the issue that

divides realists from anti-realists (the bearers of normativity) is distinct from the issue that divides stance-independent realism from stance-dependent realism (the sources of normativity). Indeed, I would speculate that it is a failure to see this distinction that leads theorists to characterize Kantian constructivism as a view opposed to moral realism.

In any case, from now on, by ‘moral realism’ I shall only mean the *ontological* claim that the bearers of moral normativity are in-the-world normative properties. This terminology allows me to say that Kantian constructivism is compatible with moral realism while maintaining that it takes a position on a metaethical issue, namely on the source of moral normativity. And while this position is compatible with moral realism, it contradicts a view held by many *advocates* of moral realism, namely source externalism (e.g., Enoch 2006, Parfit 2006, Scanlon 2014, Shafer-Landau 2003).

4. Autonomy as the Source of Morality’s Normativity

Now that we know what claims about the sources of normativity amount to, we are in a position to spell out what Kantian constructivism says about the source of moral normativity. As I shall explain shortly, Kantian constructivism claims that the source of morality’s normativity is located in the form of our will or practical standpoint.

The first thing to note is that, on my construal, Kantian constructivism is a view about the normativity of *morality* rather than a view about *all* normativity (*pace* Korsgaard 1996, 2008e, O’Neill 2015a). In Yonatan Shemmer’s terminology, I put forward a “local” constructivism, according to which “only part of the normative domain is constructed”, rather than a “global” one, according to which “all our norms are constructed” (2012: 161). My reason for restricting Kantian constructivism’s source-claim to morality is that morality purports to possess a special kind of normative force, namely that of *obligation*.

As Stephen Darwall observes, “[w]hen we are morally obligated, we are not morally free to act otherwise” (2006: 27). In other words, when we are obligated to act in a certain way, it would be *wrong* not to.⁹ Echoing a distinction introduced by Hobbes, Darwall expresses this by saying that obligations take the form of “commands” rather

⁹ It is tempting to express this by saying that moral obligations provide *conclusive* reasons for action. However, having an obligation to φ does not necessarily conclude deliberation. Obligations can be *pro tanto* and can be outweighed by other obligations. What distinguishes them from recommending reasons is that they *purport* to conclude deliberation by *demanding* some course of action (Darwall 2006: 26). We might say that obligations already “enter the deliberative field in a distinctive normative key” (Wallace 2013: 164).

than “counsels” (Darwall 2006: 12, Hobbes 1651: Chapter 25). Now, one might object that not all of morality takes this form. Instead, it might be pointed out, many moral considerations take the form of reasons that *recommend* or *favour* an action, rendering it good or choice-worthy, without implying that it would be wrong not to perform it. Indeed, the existence of such reasons seems necessary to account for the notion of supererogation, i.e., actions that are morally good but not required (Darwall 2016: 265-266). Now, I think that the concept of obligation has wider application than it is sometimes thought to have. In fact, I do not believe that there are acts of supererogation (cf. Timmermann 2005). However, within this thesis, I cannot adequately defend these claims.¹⁰

I therefore qualify my formulation of Kantian constructivism by restricting its claims about the sources of normativity to *that part of morality* which takes the form of obligations, leaving open whether this encompasses *all* of morality. For the sake of simplicity, however, I will continue to speak of the normativity of morality or moral normativity, hoping that the reader will bear in mind that my claims are restricted to moral obligation. For the same reason, and in keeping with the literature, I will sometimes speak of the normativity of moral obligations as involving reasons. Here, too, it is important to keep in mind that, insofar as moral obligations involve reasons at all, these are not reasons that merely *recommend* courses of actions (see Chapters 4 and 6, and Section 6 in this Chapter). Finally, I should note that I am using the expressions ‘is obligated to’, ‘is morally required to’, and ‘morally ought to’ interchangeably.

Kantian constructivism’s claim about the source of moral normativity rests on an analysis of the concept of obligation according to which obligations purport to bind an agent merely by virtue of the *form* of her will. As we shall see, the way in which this analysis of obligation supports Kantian constructivism’s source claim is strongly connected to Kantian constructivism’s claims about the *content* of our moral obligations. In particular, the Kantian analysis of obligation implies that the Categorical Imperative is the supreme principle of morality, i.e., that the content of our moral obligations is determined by that principle. And this implication, in turn, implies that all and only agents who have an autonomous will are bound by moral obligations. In what follows, I elaborate on both implications and highlight their relation to Kantian constructivism’s source claim.

¹⁰ But see Chapter 7 below for some reasons to think that the concept of obligation covers more phenomena than is usually thought.

4.1 Obligation and the Categorical Imperative

The Kantian analysis of obligation starts from the notion that an agent is obligated to φ only if her judgment *that* she is obligated to φ is sufficient to motivate her to φ (cf. Korsgaard 1989a: 314). Note that this is not the neo-Humean claim that an agent ought to φ only if φ -ing stands in the right kind of relation to her *antecedent* motivational states, i.e., the motivational states that the agent has independently of whether she is obligated to φ (cf. Street 2017, Williams 1981). By contrast, the Kantian analysis of obligation holds that an agent is obligated to φ only if she can be motivated to φ *because* she is obligated to do so, and thus independently of her antecedent motivational states.

In Kant's terminology, the agent must be capable of acting "from duty" rather than merely "in conformity with duty" (*G* 4:398). For moral obligation to be possible, then, morality must provide its own, *sui generis* motive. We might say that obligations require "uptake" by the agent in the sense that their obligatory nature must, by itself, be capable of motivating the agent (Korsgaard 2008d: 211). As we shall see, Kant eventually identifies this motive—sometimes referred to as 'the motive of duty'—with *reverence* or *respect* for the law (*G* 4:400, cf. Timmermann 2007: 39).

Now, I said that to φ from duty is to φ , motivated by one's judgment that one is obligated to φ . But I also said that one is obligated to φ only if one can φ from duty. This seems to create a puzzle: how can there be any obligations if, to be obligated, one must have a certain motive which in turn *presupposes* that one is obligated? In short, the Kantian analysis seems to put the cart before the horse by taking the motive to φ from duty to be prior to the obligation to φ (cf. Ross 2002: 5-6).

The apparent puzzle is resolved by the Kantian notion that what duty requires of us is not, primarily, a function of the external act in question, i.e., φ , but of the quality of the *volition* involved in performing that act. All and only those acts that could in principle spring from a certain *type* of volition are morally permissible. We can identify the relevant type of volition with the help of the above observations about the motive of duty: to φ from duty is to act on the judgment that one is obligated to φ , where what one is *obligated* to do is in turn a function of what one *can* do from duty. This suggests that the type of volition involved in acting from duty has a certain self-referential or *reflexive structure*: to φ from duty is to φ , motivated by the judgment that one *can* φ from duty (cf. Engstrom 2013: 145-146, Korsgaard 2008d: 213-225).

But this insight might not seem to address the worry mentioned above. While it tells us that the volition involved in acting from duty exemplifies a certain structure, it still does not seem to tell us which actions duty requires of us. Indeed, the observation

that acting from duty involves a volition that has a reflexive structure seems to *diagnose* the problem rather than to solve it. After all, one might ask, isn't the reflexive structure of the volition involved in acting from duty what gave rise to the worry that the Kantian analysis is putting the cart before the horse in the first place? As a solution to this problem, Kant proposes that, in order to determine the moral status of a particular action, we have to attend to the "form" of the "principle" underlying our decision to perform that action (G 4:416).

Kant refers to the underlying principle of an action as its "maxim" (G 4:400). For our purposes, we can define a maxim as the principle that expresses an agent's proposed course of action in the following form: I will perform a certain act, φ , toward a certain end, E (cf. Korsgaard 2009a: 10-11). The first thing to note is that our maxim captures the *volition* that underlies our action. As such, it is the right kind of thing for duty to latch onto, as it were. After all, we saw that duty is concerned, in the first instance, with our volition. However, we also saw that duty is concerned with the structure of our volition, rather than its content. So, the moral status of our actions cannot depend on the *substance* of our maxim, viz., its particular φ and E . Fortunately, a maxim does not only have a substance, it also has a *form*. By a maxim's form we may understand its internal arrangement, the structural relations between its φ and E (Korsgaard 2009a: 16). If an action is performed for the sake of the form of its underlying maxim, then, it is in effect performed for the sake of the *structure of the volition* going into the action.

Whether our action is in line with duty can thus be determined by attending to the form of our maxim. But what must the form of our maxim be like if it is to be in line with duty? Above, I noted that to φ from duty is to act on the judgment that one is obligated to φ , where what one is *obligated* to do is in turn a function of what one *can* do from duty. All that this tells us is that the volition going into our action it is to be guided by the notion of *necessity* or *requirement*. From this, Kant concludes that the maxim which underlies a morally permissible action must take the form of *lawfulness* (G 4:402, italics altered):

Since I have robbed the will of all impulses that could arise for it from following some *particular law*, nothing remains but *as such the universal conformity of actions with law*, which alone is to serve the will as its principle, i.e. I ought never

to proceed except in such a way *that I could also will that my maxim should become a universal law*.¹¹

This means that the maxim underlying an action must pass the test of the Categorical Imperative, which in its first variation, the so-called Formula of Universal Law, states: “*act only according to that maxim through which you can at the same time will that it become a universal law*” (G 4:421).

We can thus see that one implication of the Kantian analysis of obligation is that the content of moral principles is determined by the Categorical Imperative, which we may therefore call “the supreme principle of morality” or “the moral law” (G 4:392 and 4:449, respectively).¹² Before I go on to show that this analysis also implies that all and only autonomous wills are under moral obligations, some clarificatory remarks are in order.

The first remark concerns the relation between morally permissible actions and actions done from duty. The argument just given does *not* imply that an action is morally permissible if and only if it is done from duty. Morally permissible actions are those whose underlying maxims pass the Categorical Imperative test. That is, they are performed based on maxims the adoption of which is compatible with their own universalisation. For this to be the case, it is not necessary that the maxims in question be adopted *because* they are universalisable. Where an action’s moral permissibility is concerned, the underlying maxim may just *happen* to pass the Categorical Imperative. The link between morally permissible actions and actions from duty is that morally permissible

¹¹ Some writers regard Kant’s argument in this passage as problematic (e.g., Wood 1999: 78-82). They question whether the Categorical Imperative is really the only candidate for a law that we can follow from duty. However, as Timmermann points out, the law in question cannot be any specific, substantive law because whether an action can be performed from duty depends on its underlying principle; in particular, this principle must “conform to universality as such” (2007: 75).

¹² Strictly speaking, Kant thinks that, while the moral law governs all autonomous wills, only beings with a sensuous nature (who do not necessarily act in accordance with that law) experience this law as an *imperative* (G 4:413, cf. Timmermann 2007: 61). This might suggest a Kantian theory of what I have called the ontology of normativity: what we experience as an *ought* is really the *descriptive* law of our non-sensuous, purely rational counterparts (G 4:453, cf. Sensen 2011: 115-116, Timmermann 2007: 62-63). However, I see no reason why Kantian constructivism should be committed to this aspect of Kant’s view. Accordingly, I shall not distinguish the moral law from the Categorical Imperative.

actions are ones that *could* be done from duty. Here, we must proceed with care to avoid misunderstandings.

First, it is important to note that, for an action to be classified as done from duty, the universalisability of the underlying maxim need not be the *sole* reason for which it is selected. Kant famously distinguishes an agent who helps others from sympathy for their plight from an agent who does so from duty (*G* 4: 398-399). However, Korsgaard remarks (2008c: 179):

The problem with [the sympathetic agent] is not that he wants to help others only because it pleases him to do so. The trouble is that he *chooses* to help others *only* because he *wants* to.

The agent's *end*, by itself, is insufficient to determine whether his action is done from duty. In fact, Korsgaard notes, it is plausible that the agent who helps from sympathy and the agent who helps from duty have the *same* end: to help someone (2008c: 178). For one's action to be done from duty, then, it is not necessary that one's *end* be to do one's duty. Korsgaard writes that "[t]he person who acts from immediate inclination and the person who acts from duty in a sense act in accordance with the same material principle, which Kant specifies as 'to be beneficent where one can'" (Korsgaard 2008c: 180, *G* 4:398). The difference between the two is that the person who acts from duty *chooses* to act on this maxim because it is his *duty* whereas the other chooses to act on it because he *wants* (i.e., has some antecedent motivational state) to do so. Importantly, this does not rule out that the person who acts from duty *also* wants to act on this maxim. Indeed, his wanting to act on it might explain why the maxim of being beneficent where he can occurred to him in the first place. In this sense, it might even be *part* of what motivates him to act on this maxim. The important point is that it is not his *entire* motivation for doing so.

Second, matters are different in the case of duties that do not command but *forbid* maxims. Here, acting from duty cannot be a matter of one's motivation for choosing the maxim on which one ends up acting. After all, one can avoid acting on a forbidden maxim (e.g., "when I believe myself to be in need of money I shall borrow money, and promise to repay it, even though I know that it will never happen", *G* 4:422) by choosing to act on any *other* maxim whatsoever (e.g., 'when I am bored, I shall play a game to entertain myself'). And it seems implausible to say that, in order for my compliance with the duty prohibiting some maxim to qualify as compliance from duty, in choosing *my* maxim (whichever it turns out to be) I must at least partly be motivated by the fact that that *other* maxim is prohibited. Instead, in the case of duties prohibiting maxims, the

label ‘from duty’ applies, strictly speaking, not to the choice of the maxim we act on, but to our *avoidance* of the prohibited maxim. I comply with such a duty from duty only if my choice *not* to act on the prohibited maxim is at least partly motivated by my recognition that it is prohibited by duty.

The second clarificatory remark concerns the sense in which the Categorical Imperative determines moral *principles*. Strictly speaking, it does not yield any substantive moral principles on its own. First, the agent must, explicitly or implicitly, formulate a maxim (cf. Sensen 2013: 80-81). It is important to note that an agent’s maxim is furnished, not by some prior and independent moral faculty, but by the agent’s *empirical* psychology, by which I mean any aspect of the agent’s deliberative faculties that is independent from the mere form of her will. The Categorical Imperative then provides a test for the *permissibility* of the agent’s maxim. Accordingly, its verdicts are, first and foremost, prohibitions or permissions regarding the adoption of certain maxims (although, indirectly, the Categorical Imperative can also command the adoption of a certain maxim, by prohibiting all maxims that contradict it). The primary task of the Categorical Imperative is therefore not that of providing us with a body of moral principles, informing us which acts we may, may not, or must perform. However, for simplicity, I will continue to occasionally refer to its verdicts as ‘principles’ or ‘requirements’. After all, every prohibition or permission of a maxim can, in principle, be reformulated as a substantive principle along the lines of ‘Don’t (do) act on such-and-such a maxim’.

The third clarificatory remark pertains to why a maxim’s lawfulness requires that it can be willed as a *universal* law. To see this, recall that whether my maxim is lawful depends on its *form* rather than its substance. Now, the substance of my maxim is furnished by my empirical psychology at a particular point in time. Accordingly, when I want to determine whether my maxim has the form of lawfulness, I need to abstract from my individual empirical psychology. I need to ask if I, *qua* agent, can will that I act on the maxim in question. And this, in turn, is to ask if I can will that not only *I*, on *this* occasion, but *all* agents, at *all* times, act on this maxim.

4.2 *Autonomy and the Form of Law*

We have seen that the Kantian analysis of obligation implies that the content of our moral obligations is determined by the Categorical Imperative. Now, the Categorical Imperative is a merely formal principle. It does not, or not directly, issue any substantive commands. Instead, it merely demands that our maxims display the *form* of lawfulness. Accordingly, Kant characterizes the motive of duty as “respect for *the law*” rather than respect for the *moral* law, or for that matter *any* particular law (*G* 4:400, italics added).

For the remainder of this thesis, by ‘the motive of duty’, I shall specifically mean this Kantian idea of respect for the mere form of lawfulness.

Now, from the above remarks it follows that action from the motive of duty requires a rather peculiar agential capacity. In particular, it requires that one be able to shape one’s actions by “[n]othing other than *the representation of the law in itself*” (G 4:401). In other words, if we are capable of acting from duty, we are capable of being motivated by the mere idea of an unconditionally binding requirement, and to adopt particular maxims in light of this merely formal idea. This capacity, the capacity to make concrete *laws* by applying the *form of law* as such, is what Kant calls “autonomy of the will”, and which he defines as “the characteristic of the will by which it is a law to itself (independently of any characteristic of the objects of willing)” (G 4:440).

It follows that only agents who possess autonomous wills are bound by moral obligations. As I now want to argue, having an autonomous will is not only necessary but also sufficient for being under moral obligations. What is more, our having autonomous wills is the source of the normativity of our moral obligations.

To see this, note that autonomy is a capacity that has normative implications (Reath 2006b: 127-29). After all, an autonomous will is capable of being ‘a law to itself’, meaning that it is bound by a certain type of normative requirement. Accordingly, to possess an autonomous will, an agent must already *accept* (implicitly or explicitly) the normative force of the law that is her own will. Now, as I stated above, the law that the will is to itself is independent ‘of any characteristic of the objects of willing’. That is, the law that the will is to itself does not depend on the *substance* of its volitions. It therefore cannot command the adoption of specific maxims on account of their matter. Instead, it must govern volition by a purely *formal* requirement. And this can only be the requirement that the will’s maxims be compatible with the idea of the will as being a law to itself (cf. Reath 2013).

But what does it mean for the will to be a law to itself? At the very least, it means that there is a requirement that applies to the will *qua* will, i.e., independently of any contingent facts about the individual whose will it is. This must be a requirement that applies to *all* wills as such (Reath 2006b: 133-34). The law that the will is to itself is thus a *universal* law, in this specific sense. It follows that if the adoption of a maxim is to be compatible with the will’s being a law to itself, it must be compatible with the will’s being a universal law. As Kant says about the Formula of Autonomy: “According to this principle, all maxims are rejected that are not consistent with the will’s own universal legislation” (G 4:431). Therefore, the *form* that maxims must have in order to be compatible with the will’s being a law to itself is *universality*. But this just means that it must pass

the test of the Formula of Universal Law: “[A]ct only according to that maxim through which you can at the same time will that it become a universal law” (*G* 4:421, italics deleted).

This, then, is how Kantian constructivism motivates the claim that possessing an autonomous will is sufficient for being bound by the Categorical Imperative. However, there might seem to be a problem with the argument I just gave. This argument started from the premise that an agent who has an autonomous will already *accepts* the normativity of the law that her will is to itself. And this might seem to suggest that the normativity of the Categorical Imperative is *prior* to the notion of an autonomous will, thus threatening to undermine the claim that the form of the will is the *source* of moral normativity. To address this worry, we need to understand that the Categorical Imperative is the *constitutive* principle of autonomous volition.

To say that a principle is constitutive of an activity is to say that “unless you are guided by the principle in question, you are not performing that activity at all” (Korsgaard 2009a: 29). To borrow two examples from Korsgaard, the constitutive principle of walking requires that you set one foot in front of the other and the constitutive principle of building a house requires that you erect a physical structure that provides shelter from the weather (*ibid.*). If you do not try to conform to these principles, you are not engaged in the corresponding activity at all. But constitutive principles not only enable us to determine whether someone is performing an activity at all, they also serve as normative standards for the activity in question (Korsgaard 2009a: 29-32). For example, building a house that has random holes in the walls and ceiling will count as building a house *badly*. After all, it will only imperfectly conform to the constitutive principle of that activity. However, whereas normative standards that are not constitutive of house-building (e.g., standards of beauty) depend on the presence of certain external factors (e.g., whether the house-builder cares about beauty), the constitutive principle of an activity seems to provide a normative standard which is entirely *internal* to that activity.

Now, to see why the Categorical Imperative is the constitutive principle of autonomous willing, let us return to what this type of volition involves. As Andrews Reath points out, there is a sense in which a “valid” maxim, i.e., one that is consistent with the will’s being a law to itself, must “*contain the ground of its own validity within itself*” (2006b: 133, cf. Korsgaard 2008d: 213-225). After all, as noted above, a valid maxim cannot draw its validity from its particular contents, which are external to the will’s own workings. Instead, the validity of a maxim must come from its form. In particular, it is only when our maxim has the form of lawfulness that our volition can be said to have the normative status of a law *by itself*.

Reath raises an important point about this: valid maxims exhibit a certain *self-referential* character (2006b: 134).¹³ The idea is that valid maxims must be able to contain their own adoption as a universal law as their object, alongside their empirical contents.¹⁴ Reath points out that this sheds light on formulations of the Categorical Imperative like the following (2006b: 134):

[A]ct according to maxims that can at the same time have as their object themselves as universal laws of nature. (*G* 4: 437, italics deleted)

[A]ct on a maxim that at the same time contains in itself its own universal validity for every rational being. (*G* 4: 437-438)

We can see, then, that the Categorical Imperative is the constitutive principle of autonomous willing: unless our maxims can be willed in conjunction with their own universal validity—and thus pass the test of the Categorical Imperative—they do not count as an instance of autonomous volition at all. Connectedly, the requirement that our maxims contain the ground of their own validity within themselves can be read as a command that our will live up to its *own* ambition of being a law to itself. This is why Kant emphasizes that, at bottom, the Categorical Imperative “commands neither more nor less than just this autonomy” (*G* 4:440).¹⁵

Reath puts the thought of the Categorical Imperative as the constitutive principle of willing as follows (2006b: 134, italics added):

[T]he deliberative procedure provided by the [Formula of Universal Law] is plausibly understood as concerned with *internal features of the maxim*, because it focuses on the reasoning going into the maxim, rather than assessing a maxim in

¹³ This is closely connected to the observation that the volition involved in acting from duty exhibits a reflexive structure (see previous subsection).

¹⁴ By the maxim’s ‘object’, we may understand its end, i.e., something in the world at which it is *aimed* (cf. Timmermann 2006a: 72).

¹⁵ The notion that the law is the constitutive principle of an autonomous will means that an autonomous will already accepts the law as normative and, in this sense, aspires to comply with it. Thus, the law can be said to command that we follow our own aspiration *qua* autonomous wills. Perhaps it would be more accurate to say that the law commands that we *realise* our autonomy (cf. Sensen 2011: 168-169). But the quote from Kant captures the important point that this command is internal to the constitution of an autonomous will. As Reath puts it, the autonomous will is “self-consciously guided by [...] the formal end of satisfying its own internal norm (the conditions of universality)” (2012: 223).

terms of its conformity to a further substantive principle, or conduciveness to an end. [...] [T]he focus is on the normative claims *implicit in the maxim*.

This highlights that the Categorical Imperative is not really a principle, in the sense in which most moral theories advance principles, at all. It does not play the role of *informing* our deliberation, by providing an external standard against which our maxims are assessed. Autonomous willing does not proceed in the form of a practical syllogism, with the proposed maxim as its minor and the Categorical Imperative as its major premise (cf. Korsgaard 2008e: 321, Reath 2013: 577). Instead, the Categorical Imperative is a mere expression—a distilled, purified statement, as it were—of the form that our valid maxims *already embody*. When we abstract from the concrete substantive content of our valid maxims, only their form remains. By expressing this form, the Categorical Imperative *spells out* the feature by which maxims gain validity. But it does not play the role of an external foundation, a *punctum Archimedis*, from which the correct moral principles are deduced (Bagnoli 2014: 318, cf. Forst 2011: 46, Rawls 1980: 572, Reath 2013: 577-578, Sensen 2011: 113-118).

Indeed, the Categorical Imperative itself *exemplifies* the form of valid maxims. After all, it is itself a principle that can, and indeed *must*, be adopted by all wills, *qua* wills (Reath 2010: 37-38). Put another way, the Categorical Imperative, considered as a maxim, passes its own test. We might say that it is the *paradigm*, the purest exemplar, of a maxim that is valid merely by virtue of its form. This is implicit in Kant's statement, quoted above, that the Categorical Imperative commands nothing but autonomy. The Categorical Imperative ends the autonomous will's search for a principle that is valid merely by virtue of its form, not by giving a substantive answer, but by providing what is essentially a re-formulation of the question: the principle that is valid by virtue of its mere form *is* the requirement to act only on maxims that are valid merely by their form (cf. Korsgaard 1996: 98, Sensen 2011: 109-111).¹⁶ This is, perhaps, the ultimate sense in which the workings of an autonomous will are self-referential.

We are now in a position to see why Kantian constructivism's source claim is not undermined by the notion that an autonomous will acknowledges the normativity of the form of law. It is true that the Categorical Imperative is implicitly presupposed in the workings of an autonomous will and that it is therefore, in a sense, *more fundamental* than the maxims whose validity it tests. As Oliver Sensen observes, while we construct

¹⁶ In a similar vein, Darwall observes that the very line of questioning that leads us to adopt a principle of autonomy, asking for a standard of "unqualified" normativity, is itself an exercise of autonomy (1990: 263-264).

moral principles from our antecedently given maxims under the guidance of the Categorical Imperative, the Categorical Imperative itself is “a direct command of reason” (2013: 73, 80-81, cf. Kain 2004: 301). That is, it is not ‘up to us’, as it were, to formulate the moral law.¹⁷

However, it does not follow that the Categorical Imperative is part of a ‘moral reality’ that has normative force independently of the existence of autonomous wills (Bagnoli 2013b: 172). The Categorical Imperative does not play the role of a substantive input to our practical deliberation, i.e., it does not confront us as a principle external to our will that commands the performance of actions or the adoption of maxims of a specific content. Instead, the Categorical Imperative is the constitutive principle of autonomous willing, whose command—‘be autonomous!’—is *internal* to that type of volition.¹⁸ Accordingly, in Sensen’s words, if we did not engage in autonomous willing, “there would be no principle of reason” (2013: 77).¹⁹ But once we do engage in autonomous willing, “reason creates the law spontaneously and of its own accord” (2013: 77, cf. 2011: 115).²⁰ I take it that this is one aspect of what Kant means when he says that the law “forces itself upon us of itself” (*CPrR* 5:31).

¹⁷ I take it that this is what Kantians have in mind when they emphasise that the moral law has no *author* (*MM* 6:227, cf. Kain 2004, Timmermann 2007: 107, Wood 2008: 113). The idea is that although our will is the source of the normativity of the moral law—and, in this sense, we are its *legislators* or the authors of our *obligation* to comply with it—what the law *is* does not depend on us, nor on anyone else.

¹⁸ Thomas Hill Jr. writes (2007: 496): “We cannot coherently think of the rational legislators as ‘making’ and ‘giving themselves’ the basic rational standards that constitute their rationality. Nevertheless, their autonomy implies that they identify themselves with these constitutive standards rather than seeing them as externally imposed and alien.”

¹⁹ Sensen labels this position “transcendental constitutivism” rather than “constructivism” (2013: 65). Like many Kantians, Sensen assumes that the label “constructivist” applies only to theories that regard the moral law as “a result of the conscious deliberation of an individual or group” (*ibid.*, cf. Kain 2004: 260, Wood 1999: 374-375, 2008: 107). However, this use of terminology is misleading since most Kantian constructivists explicitly state that they do *not* subscribe to such a position (cf. Bagnoli 2016, Korsgaard 1996: 234-236, Rawls 1980: 567-568, 2000: 239-240, Reath 2006c: 176-180, 2013: 577-578). Accordingly, one common objection to Kantian constructivism—that “[c]onstructivism could not explain how morality could be *unconditionally* binding” because it depicts moral normativity as contingent upon some *act* of construction—misses its target (Sensen 2013: 80, cf. Kain 2004: 262-265, Wood 1999: 156, 2008: 106-116).

²⁰ I regard the use of “create” as metaphorical in this context. I say more about the grounding relation that holds between our will and moral normativity in Section 4.4 below.

4.3 Grounding Moral Normativity

We have seen that having an autonomous will is sufficient for being bound by the Categorical Imperative. But does this mean that autonomy of the will is the source of moral normativity?

Now, one desideratum of a plausible source claim is that it has explanatory power (see Section 3.1). That is, the proposed source of something's normativity should explain its having normative force. I think that the Kantian constructivist source claim meets this desideratum. Morality is normative because we possess an autonomous will, where our having an autonomous will implies that the law that our will is to itself, i.e., the moral law, is normative for us. Thus, our autonomy explains why we are under moral obligations. It does so because the Categorical Imperative, the supreme principle of morality, is the constitutive principle of autonomous willing.

What of the desideratum of asymmetry? One might argue that the Kantian constructivist account outlined here does not meet this criterion because it says that having an autonomous will is not only sufficient but also necessary for being under moral obligations. However, note that two propositions can be logically equivalent while, metaphysically, standing in an asymmetric relation. Consider the following example by Matthew Kramer (2000: 39):

If a person X very much desires a slope in his yard so that he can enjoy the downward perspective from the top, he can and should recognize that his having the slope will also provide the yard with an upward perspective from the bottom. [...] That relationship of strict simultaneity between the upwardness and the downwardness is scarcely undone by a justificatory concentration on one or the other of those two simultaneous aspects.

Just as the 'strict simultaneity' of the slope's downwardness and upwardness does not preclude a 'justificatory concentration' on its downwardness, so the biconditional that holds between our having autonomous wills and morality's being normative does not preclude that the former is the metaphysical ground of the latter.

And, indeed, the metaphysical relation between autonomy and moral normativity is *not* symmetric. The Categorical Imperative's having normative force does not ground our having autonomous wills. Instead, our *acceptance* of the Categorical Imperative is constitutive of our having autonomous wills. Hence, while the *concept* of the law that the will is to itself, i.e., the moral law, is prior to the workings of an autonomous will, its *instantiation* is not (cf. Bagnoli 2016: 1233). To borrow Sensen's words again, "if one were never prompted to deliberate morally, there would be no principle of reason" (2013:

77). As I will elaborate further in Chapter 3, I adopt Kant's view that, although moral normativity and the autonomy of our wills imply one another, moral normativity is merely the "*ratio cognoscendi*" of autonomy while autonomy is the "*ratio essendi*" of moral normativity (CPrR 5:4, footnote). Roughly, while the normativity of morality indicates that we have autonomous wills, our having such wills is what grounds morality's having normative force.

Now, a plausible claim about sources of normativity should not only possess explanatory power. In addition, the explanation it gives should not be a causal one, nor one in terms of supervenience or subsumption under a more general normative consideration (see Section 3). Clearly, the Kantian constructivist claim cannot be read as suggesting that our having autonomous wills *causes* morality's being normative. Neither does it seem plausible to say that the autonomous will is not the ground but the supervenience basis of morality's normativity.

Whether the Kantian constructivist claim can be read as subsuming morality under a more general normative consideration is somewhat more difficult to settle. Kantian constructivism might seem to depict the relation between having an autonomous will and being subject to moral obligations as akin to that between occupying a certain social role and being subject to the obligations attached to it. In the latter case, it is not implausible to suppose that there is a prior and independent moral fact that makes it the case that by occupying a certain social role, e.g., the role of parent, one is subject to certain obligations, e.g., parental obligations. Thus, if the Kantian constructivist claim followed this model, it would not amount to the claim that autonomy is the source of moral normativity at all. Rather, it would state that our having autonomous wills triggers a prior and independent normative fact or principle (e.g., 'all and only those who possess autonomous wills are under moral obligation') that makes it the case that we are under moral obligations.

However, there is good reason not to read the Kantian constructivist claim in this way. First, Kantian constructivism goes beyond saying that having an autonomous will is a sufficient condition for having moral obligations; it claims that the Categorical Imperative is the *constitutive* principle of autonomous willing. This suggests that, if we did not have autonomous wills, moral obligations would not exist, i.e., morality would not have any normative force. Our having autonomous wills not only explains why *we have* moral obligations, it also explains why morality has normative force *at all*. This is unlike the case of parental obligations, in which an agent's *having* parental obligations might depend on her occupying the role of parent, but the normativity of parental obligations *in general* does not. Finally, the Kantian constructivist source claim explains the

normativity of morality in terms of what it *is* to have an autonomous will, and hence in terms of the *metaphysics* of such a will. In the case of parental obligations, by contrast, it is arguably not the metaphysics of being a parent itself but the fact that it features in some normative fact or principle that makes it the case that being a parent comes with parental obligations.

4.4 *Voluntarism, Internalism, Constructivism*

We have seen that the Kantian constructivist position spelled out in this section amounts to the claim that the form of our will, the form of universal lawgiving, is the source of moral normativity. One might wonder how this claim is to be interpreted in light of the taxonomy I introduced earlier (see Section 3).

To begin with, one might wonder whether our having autonomous wills grounds the fact that morality is normative by constituting it or by constructing or creating it. Here, it is important to remind ourselves of the distinction between the normativity of the Categorical Imperative as the supreme principle of morality on the one hand, and the normativity of substantive moral principles that prohibit or command the adoption of certain maxims on the other hand. I suggest that the former is best understood as *constituted* by our having autonomous wills whereas the latter is best understood as being *constructed* by our applying the Categorical Imperative test to particular maxims.

The distinction between constitution and construction is admittedly quite blurry. All I said was that the former is a ‘closer’ relation between a normative consideration and its source than the latter. But this blurriness is not a serious problem. After all, the distinction between the Categorical Imperative and the substantive moral principles it yields is quite blurry itself. This is one upshot of the observation that the Categorical Imperative is the constitutive principle of autonomous willing. As I noted above, this observation implies that the Categorical Imperative spells out the form of valid maxims. Indeed, the Categorical Imperative itself can be understood as a valid maxim.

The most important difference that remains is that, for there to be any substantive moral principles, our empirical psychology needs to provide us with maxims (see Section 4.1 above). Accordingly, the ‘gap’ between the form of our will and substantive moral principles seems to be larger than that between the form of our will and the Categorical Imperative. In addition, the maxims that feed into the Categorical Imperative test and thereby facilitate its prescription of substantive moral principles are, in a sense, ‘up to us’—although, of course, the outcome of the test is not (Kain 2004: 294-296). These are my reasons for thinking that the notion of construction is appropriate to capture the grounding relation that holds between our having autonomous wills and substantive

moral principles' having normative force (cf. Sensen 2013: 80-81). Since the relation between the form of our will and the Categorical Imperative, by contrast, is much 'closer' and does not involve any element that is 'up to us', I think it is more appropriately captured by the notion of constitution.

Insofar as it claims that the form of our will is the source of moral normativity, Kantian constructivism is clearly *not* a version of source externalism. As I pointed out above, autonomy can be understood as the capacity to act from the motive of duty (see Sections 5.1 and 5.2). This suggests that Kantian constructivism can be understood as a form of source internalism. However, it must be distinguished from those versions of source internalism that locate the source of moral normativity in our antecedent, empirical motives (e.g., Street 2017, Williams 1981). After all, the motive of duty consists in respect for the law and thus depends on our autonomy, i.e., our will's being a *law* to itself. In Kant's words, it requires that "pure reason" can be "practical of itself", i.e., that we possess "pure practical reason" (*CPrR* 5:31 and 5:33, resp., cf. Korsgaard 1986).

Kantian constructivism's view of the source of moral normativity might be likened to the version of internalism that Darwall labels "autonomist internalism" (1990: 263, 1995: 16). According to this view, an agent's being under an obligation to φ is grounded in her having "a conclusive motive for doing so as a result of the practical thinking that realizes autonomy" (Darwall 1992: 168).²¹ Darwall emphasises that "autonomist internalism retains a normative element in its ideal of autonomy" (1990: 264). This, I take it, is the thought that I expressed by saying that autonomy has normative implications insofar as it presupposes, logically, the normativity of the law that the will is to itself.

Since Kantian constructivism emphasizes our will's ability to be a law to itself and to give rise to its own motive, one might also be tempted to interpret its claim about the source of moral normativity as a version of source voluntarism. However, it is important to distinguish Kantian constructivism from versions of source voluntarism, such as Chang's, according to which we can employ the normative powers of our will to create normativity without being bound by any formal, internal constraints on willing (Chang 2013a: 184, 2013b: 109). According to the latter type of view, we can will almost anything to be normative for us as long as we do not overstep certain *substantive* constraints whose normativity has its source outside of our will. Kantian constructivism, by contrast,

²¹ Darwall does not formulate this position in the terminology of grounding. However, it is not hard to see how it could be restated in that way.

holds that the will is subject to the merely *formal* constraint of being a law to itself, which springs from its very constitution (see Section 4.2).

The preceding thoughts also suggest a way of understanding the distinction between Kantian constructivism and other, non-Kantian versions of constructivism (e.g. Street 2010). We might say that all versions of constructivism, insofar as they are metaethically ambitious, hold that the source of morality's normativity resides in our own will or motivational setup. The distinctively Kantian claim is that it resides in the will's capacity to be a law to itself by its mere form or, what amounts to the same thing, our capacity to act from the *sui generis* motive of respect for the mere form of law.

5. Conclusion

In this chapter, I argued that Kantian constructivism entails a metaethical position as well as a position concerning the content of our moral obligations. More specifically, it asserts that the fact that we have autonomous wills grounds the fact that morality has normative force. While Kantian constructivism is not opposed to moral realism, understood as an ontological position, it does conflict with source externalism, a view that is held by many moral realists.

Chapter 2: The Normative Question

1. Introduction

We have seen that Kantian constructivism takes a stance not only on the content of our moral obligations but also on the source of moral normativity. Kantian constructivism's source claim is primarily motivated by the Kantian analysis of obligation outlined in the previous chapter. I provide an argument for that analysis in Part II of the thesis, by showing that the relevant concept of obligation is implicit in the practical standpoint of ordinary agents.

However, as I want to argue in this chapter and the next, there is another line of argument in favour of Kantian constructivism. While this line of argument is not independent of the Kantian analysis of obligation, and thus also relies on my argument in favour of that analysis, it highlights a major benefit of accepting the Kantian constructivist view of the source of moral normativity. In particular, I want to argue that Kantian constructivism has an advantage over other views, including source externalism as well as non-Kantian forms of voluntarism and internalism, when it comes to answering the so-called Normative Question.

The claim that Kantian constructivism has this advantage features prominently in the work of several Kantian constructivists, especially in Korsgaard's (Korsgaard 1996, 2008e, cf. Bagnoli 2016). However, there is a puzzle about what exactly the Normative Question amounts to. In this chapter, I propose an interpretation of the Normative Question as asking for an explanation why we are rationally required to do what we believe we morally ought to do. I argue that this question is one that arises from within the practical standpoint of agents who already accord rational authority to morality but wonder whether this authority can be vindicated.

The chapter is structured as follows. In Section 2, I show that there is a puzzle about what exactly the Normative Question asks for. In Section 3, I introduce the interpretation of the Normative Question that I favour. I show that the question, as I interpret it, is motivated by the Kantian idea of a critique of reason.

2. The Puzzle

Since Kantian constructivism is often portrayed as an alternative to moral realism, the Normative Question is typically formulated as a challenge to that view. However, as I argued in the previous chapter, Kantian constructivism does not conflict with moral realism. Accordingly, I will advance the Normative Question as a challenge, not to moral

realism, but to source externalism as well as non-Kantian forms of source internalism and source voluntarism.

Now, Korsgaard initially formulates the Normative Question as the question of “what *justifies* the claims that morality makes on us” (1996: 10). She emphasizes that to raise the Normative Question about a specific moral truth is not to ask whether it really is a truth, by the standards of morality (1996: 13-14). In other words, it is not the question of whether moral concepts have been applied correctly to the situation at hand. Rather, according to Korsgaard, it is the question of what justifies the normative import that such truths purport to have: “Is it really true that this is what you *must* do?” (1996: 38).

Korsgaard’s objection to moral realism, recast as an objection to source externalism, amounts to the accusation that source externalism *refuses* to answer the Normative Question (1996: 39). When faced with the Normative Question about a given moral truth, source externalists might appeal to some more basic moral truth, pointing out how the initial moral truth can be derived from it. However, this does not answer the Normative Question, as Korsgaard understands it. Instead, it addresses the question of whether our moral concepts have been applied correctly, which, as we have seen, Korsgaard treats as a separate question (1996: 39):

[The source externalist’s response] addresses someone who has fallen into doubt about whether the action is really required by morality, not someone who has fallen into doubt about whether moral requirements are really normative.

This can be seen from the fact that, once it is asserted that our initial moral truth derives from a more basic moral truth, the Normative Question will simply change, and ask what justifies the alleged normative import of this more basic moral truth.

This dialectic will continue until the source externalist invokes the normativity of the normative properties that she takes to be the ultimate bearers as well as the source of moral normativity. At this point, she will no longer be able to appeal to yet a more basic layer of moral or normative truths. And since the source externalist believes that the normativity of these properties has no source beyond these properties themselves, all she can do is *assert* that their claim on us is justified. But this, according to Korsgaard, is a way of saying that the Normative Question cannot and need not be answered: the normative properties that are the ultimate bearers of normative force simply *are* normative, independently of any facts about our will or motivational states (cf. 1996: 39). However, Korsgaard objects, this is nothing but an expression of confidence in the normativity of morality, which is of little use when we raise the Normative Question; for we raise that question precisely when we lack this confidence (1996: 40).

Now, here's the puzzle: if the Normative Question is the question of whether morality's normative claim on us is justified, then source externalism seems to have a straightforward answer to it. Source externalism holds that morality's normative claims are justified because of its relation to stance-independent irreducibly normative properties.

Note that Korsgaard's objection is not that source externalism is false but rather that, even if it is true, it cannot answer the Normative Question. Accordingly, Korsgaard seems to think that, when the source externalist appeals to truths about stance-independent normative properties, the Normative Question simply remains open. It now becomes the question of whether we should choose our actions in light of truths about such properties (2008: 317). However, as Derek Parfit points out, it is puzzling that this should be an open question (2006: 358-360). After all, the stance-independent properties invoked by source externalism are *normative* properties. So, on the assumption that source externalism is true—and, accordingly, that such properties exist—how can there be an open question of whether we should choose our actions in light of these properties?

Korsgaard's assertion that source externalism cannot answer the Normative Question seems to be connected to her rejection of a certain account of normative concepts, which we briefly encountered in Chapter 1. According to Parfit and other source externalists, when we ask whether some requirement is truly normative, "there is something that [...] we are trying to find out" (Korsgaard 1996: 44, Parfit 2006: 380). That is, source externalists view a question about what we ought to do as part of a theoretical inquiry. In Parfit's words, "[t]hough it is a question about practical reasons, it is not a practical question" (2006: 379, cf. Scanlon 1998: 23). According to Korsgaard, by contrast, truths and properties are not the right *kind* of things to answer the Normative Question (1996: 44, 2008: 317). In her view, the function of normative concepts is not, first and foremost, to describe reality, but to solve practical problems. Connectedly, Korsgaard regards the Normative Question as "practical all the way down" (2008: 318).

Yet, Korsgaard's remarks about the practical function of normative concepts, by themselves, do not suffice to clarify her claim that source externalism cannot answer the Normative Question. In particular, it is not clear in what sense source externalism's approach to the Normative Question fails to be practical, and in how far this failure is objectionable. After all, as pointed out above, source externalism's invocation of truths about stance-independent properties is practical at least insofar as these truths and

properties are *normative*. In what further sense might they be practical? One might take the following passage to provide a clue (Korsgaard 1996: 38):

If someone finds that the bare fact that something is his duty does not move him to action, and asks what possible motive he has for doing it, it does not help to tell him that the fact that it is his duty just is the motive. That fact isn't motivating him just now, and therein lies his problem.

Here, it might seem as if Korsgaard is suggesting that a satisfactory answer to the Normative Question is practical in that it will *move* us to do the right thing. However, as Parfit notes, if this is how Korsgaard conceives of the Normative Question, then source externalists need not worry about not being able to answer it (2006: 371-372). After all, on this reading, the Normative Question is not really a question about normativity, considered as an irreducibly normative force, at all. Instead, it seems to presuppose a reductive conception according to which “[n]ormativity [...] is a kind of motivating force” (Parfit 2006: 371). And source externalists do not share this reductive conception of normativity.

Now, it is quite clear that Korsgaard does not intend to advance a reductive account of normativity by posing the Normative Question (e.g. Korsgaard 1996: 43, 2008: 32-45, manuscript: 8). Nor, indeed, could the Normative Question serve as a Kantian constructivist challenge to source externalism if it presupposed such an account. Not only is source externalism committed to a *non-reductive* realist conception of the ontology of normativity, Kantian constructivism is also supposed to remain *neutral* with respect to that ontology (see Chapter 1). But if the Normative Question does not presuppose a motivational conception of the ontology of normativity, the sense in which source externalism’s approach to that question fails to be sufficiently practical remains mysterious. I shall now advance an interpretation of the Normative Question that resolves this mystery.

3. What is the Normative Question?

In this section, I advance an interpretation of the Normative Question that constitutes a genuine challenge to source externalism and other views about the source of moral normativity. I argue that this question is motivated by the Kantian idea of a critique of reason. I also briefly comment on the dialectic of the challenge posed by the Normative Question.

3.1 Practical Reason and Moral Normativity

The Normative Question cannot be the question of whether morality is genuinely normative. As we have seen, source externalism can answer that question (cf. Dreier 2015: 172). But what, then, is the Normative Question? On the one hand, Korsgaard suggests that source externalism cannot answer the Normative Question because its approach is not sufficiently ‘practical’. On the other hand, the Normative Question must be a question of ‘normativity’ in a sense of that term which is shared by source externalists if it is to figure in an argument against their position. Accordingly, it cannot presuppose a reductive conception of normativity.

Now, Korsgaard’s remark that the Normative Question is practical seems to be connected to her insistence that it arises from the *first-person* point of view of the agent who is confronted with a moral obligation and deliberates about what to do. Consider the following passages from Korsgaard (1996: 13, 16, and 38, respectively):

When we do moral philosophy, we also want to know whether we are *justified* in according this kind of importance to morality. [...] We want to know what, if anything, *we* really ought to do.

The normative question is a first-person question that arises for the moral agent who must actually do what morality says.

[Y]ou are being asked to face death rather than do a certain action. You ask the normative question: you want to know whether this terrible claim on you is justified. Is it really true that this is what you *must* do?

In all these passages, Korsgaard seems to suggest that the Normative Question arises once we shift from a third-personal point of view to the first-personal point of view of the agent who must make a decision in light of, among other factors, what morality demands of her. This suggests that, while the Normative Question is in some sense a question about normativity, it is different from the question of whether morality is normative, and this difference is somehow connected to the fact that the Normative Question arises from the first-person perspective of the moral agent.

Now, one might take this to mean that the Normative Question asks whether morality is normative *for the agent*, as opposed to normative *simpliciter* (cf. Korsgaard manuscript: 6). Yet, I do not think that this is the right way to read the Normative Question. Neither should we interpret it as asking whether the agent *has* reason, as opposed to there *being* reason, to act morally. To interpret the Normative Question in one of these ways is to make it dependent upon philosophical issues that seem orthogonal to the

dispute between Kantian constructivism and other views about the sources of moral normativity. Besides, source externalists will insist that they can explain why morality is normative for us (or why we have reason to be moral) just as well as they can explain why morality is normative simpliciter (or why there is reason to be moral).

Instead, I want to suggest that the Normative Question is concerned with the relation between moral normativity and rationality. As James Dreier points out, by interpreting the Normative Question along these lines we can capture a sense in which it is *practical* (2015: 172-174). In particular, whichever way we answer the question of what we *ought* to do, we would expect that answer also to be relevant to the question of what we *shall* do, i.e., what we are going to do (2015: 172-174). Of course, our answer to the ‘ought’ question need not actually *determine* our answer to the ‘shall’ question. As Parfit correctly notes, “[m]oral philosophy cannot make our decisions” (2011b: 415). So, it would be wrong to expect our answer to the ‘ought’ question to be followed, by *natural* necessity, by a certain answer to the ‘shall’ question. Nevertheless, as Dreier goes on to argue, one would expect a connection of *rational* necessity to hold between the answers to these two questions (2015: 176-178). We might say that rationality bridges the gap between moral normativity and the practical question of what we shall do.

The Normative Question presupposes, then, that there is a close link between moral normativity and rationality. We might say that it presupposes that morality has *rational authority*. When we are confronted with a view of the sources of moral normativity, by raising the Normative Question we ask: if *this* is how we should think about moral normativity, how are we to make sense of the rational authority that we commonly attribute to morality?

I shall turn to the question of how we might unpack the notion of the rational authority of morality (and how, consequently, the Normative Question is to be formulated) in Section 3.3. For now, note that my approach to the Normative Question is congenial to Korsgaard’s overall argument against source externalism. In particular, Korsgaard repeatedly characterises her disagreement with source externalism as a disagreement about the relation between our *faculty* of practical reason, or rationality, and the particular substantive considerations that we might call *reasons* (2008e, 2009a, 2009b, manuscript). This suggests that the problem with source externalism, for Korsgaard, is not with its account of moral normativity *per se* but with the way it depicts

the *interface*, as it were, of moral normativity and practical reason. In particular, she writes (2009a: 6):

[Source externalists] believe that norms exist outside of human reason—they arise from Objective Values or Moral Facts or some sort of rational structure that exists “out there” in the universe. But if reflection on that fact prompts us to ask why human reason finds it necessary to conform to these standards, there is no real answer. The [source externalist] can only reply: “well, that's just what it means to be rational, to have a mind (or a will) that conforms to the standards that we call ‘rational’.” In fact in theories of this kind “human reason” is really nothing more than the name of that faculty within us, whatever it might be, that conforms to rational standards. It is not identified in any other way.

Korsgaard’s complaint seems to be that source externalism depicts practical reason as a *receptive* faculty (2009b: 24, manuscript: 11, cf. Korsgaard 1989b). In the source externalist picture, the operations of practical reason are determined by substantive inputs which are fixed by stance-independent normative properties and lie outside the faculty of practical reason itself. When we ask *why* our practical reason singles out these particular substantive considerations as inputs to its operations, and not others, source externalists will respond that they form the proper input to practical reasoning because they are *reasons*, and practical reason *just is* the faculty that responds to practical reasons (cf. Chang 2013a: 170, Parfit 2011a: 112).

The interpretation of the Normative Question suggested by these remarks also makes sense of Korsgaard’s emphasis on the first-personal character of the question. The Normative Question is a first-personal question, on this interpretation, because it arises from the point of view of our faculty of practical reason. It is concerned with that faculty’s interface with moral normativity. In other words, it arises when *we* are employing *our* faculty of practical reason in order to determine what we are rationally required to do, and it concerns the role that we should assign to moral requirements in that process.

3.2 *Motivating the Question*

I have suggested that Korsgaard’s complaint against source externalism seems to be that this view characterises practical reason as a receptive faculty. Now, source externalists might agree that this is the way they characterise the relation between moral normativity and practical reason but deny that this is a problem. So, one might wonder how the Normative Question is motivated. I want to suggest that the primary motivation for the Normative Question, as I understand it, is the Kantian idea of a critique of reason. While this

means that source externalists need not worry about the Normative Question if they are not attracted to that idea, I suspect that many people find this idea appealing. I argue that one indicator of its appeal is the extent to which we are troubled by the possibility of debunking explanations of morality's rational authority.

3.2.1 *The Critique of Reason*

The Normative Question is motivated primarily by the idea of a *critique of reason*, which plays a role in the work of several Kantians, especially in that of Onora O'Neill (1989a, 2015a, cf. Bagnoli 2013b: 10, Forst 2011: 32, Rawls 2000: 262). This idea entails that reason reflexively *criticizes itself*. Unfortunately, I can only sketch this idea in terms that are somewhat vague and metaphorical. Nevertheless, I hope that the following remarks suffice to convey the rough outlines of this idea and bring out some of its attractions.

The notion of a critique of reason implies that the claims of reason require *vindication*. That is, they do not play the role of self-evident axioms that provide the foundation for all justification without being in need of justification themselves (e.g., 1989c: 14, 27, 2015c: 14-15). The following passage from Kant—regardless of whether Kant intended it to be read that way²²—can serve as an expression of this idea (*CPR* A738/766):

Reason must in all its undertakings subject itself to criticism; should it limit freedom of criticism by any prohibitions, it must harm itself, drawing upon itself a damaging suspicion. Nothing is so important through its usefulness, nothing so sacred, that it may be exempted from this searching examination, which knows no respect for persons. Reason depends on this freedom for its very existence. For reason has no dictatorial authority; its verdict is always simply the agreement of free citizens, of whom each one must be permitted to express, without let or hindrance, his objections or even his veto.

According to O'Neill, Kant's use of political metaphors to describe the role of reason is not accidental. Just as a political authority that suppresses criticism and refuses to justify

²² In these remarks, Kant is exclusively concerned with pure theoretical reason. In fact, in the *Critique of Practical Reason*, Kant explicitly denies the need for a critique of pure reason in the practical sphere—although he had acknowledged this need in the *Groundwork of the Metaphysics of Morals* (*CPrR* 5:3, 15-16, *G* 4:445, cf. Stern 2010). In the *Critique*, Kant states that it suffices to show that pure practical reason *exists*, which he then attempts to demonstrate via the argument from the “fact of reason” (*CPrR* 5:30-31). I say more on this argument in Chapters 3, 4, and 5. This argument cannot spare us the need for a critique of pure practical reason because, I argue, its conclusion is fallible.

itself in the public forum must eventually lose its credibility, so reason loses its characteristic authority if it refuses to justify its claims (O'Neill 1989d: 37, cf. *WOT* 8:145-146).

But why does reason's authority depend on a critical vindication of its claims? The answer is connected to reason's function or role. Reason is the faculty which is supposed to facilitate *enlightenment* (e.g., O'Neill 1989d: 37). And, as Kant notes, the guiding idea of enlightenment is to emerge from "self-incurred minority", to think critically and judge for oneself instead of accepting the verdicts of unchecked authorities, such as church or state (*WE* 8:35). Thus, if reason is to facilitate enlightenment, it must be the faculty of *criticism*. But this means that if reason left its *own* workings unexamined, it would undermine the project of enlightenment. After all, this would be tantamount to reason's propping up *itself* as an unchecked authority. We might say that reason could not play the role of reason—and its claims would not have the special *authority* we ascribe to reason—if it refused to criticise itself (cf. O'Neill 1989d: 34). The verdicts of reason would be like theorems deduced within an axiomatic system, which do not convince anyone who does not already regard its axioms as self-evident. Instead of transcending the pre-enlightenment discourse and casting off the "yoke of minority" (*WE* 8: 35), the defenders of the various authorities would continue to talk past each other—like the infamous builders of the tower of Babel (O'Neill 2015b: 3-4, cf. *CPR* A707/B735, *WOT* 8:145).

But if reason is the faculty of criticism, then the critique of reason must be carried out by reason *itself*. That is, we must use the principles of rational justification to investigate the rational justifiability of these very principles themselves. I say more on this in the next chapter. For now, note that the idea of a critique of reason suggests that when a claim presents itself to us as endowed with the authority of reason, we must question its justification to see if it can be sustained under the scrutiny of reason.

How does the idea of a critique of reason support Korsgaard's complaint against source externalism? The problem with source externalism is that, by depicting practical reason as a receptive faculty, it renders a critical vindication of the verdicts of practical reason *impossible*. After all, according to source externalism, the substantive considerations that serve as inputs to practical reasoning and determine what rationality requires of us, lie outside the faculty of reason itself. Consequently, it remains opaque to us how and why our practical reason singles out these *particular* substantive considerations as requiring a response from a rational agent. The fundamental standards of practical reason lie outside the critical scrutiny of reason. They remain hidden from view, like the inner workings of a *black box*, and can therefore be neither criticised nor justified. We might say that, by depicting practical reason as determined by substantive inputs fixed

by stance-independent normative properties, source externalism likens the verdicts of practical reason to theorems in a deductive system based on axioms.

3.2.2 *The Sceptical Threat from Debunking Explanations*

I have suggested that the Normative Question is motivated by the Kantian idea of a critique of reason. However, this idea is quite abstract and it might, therefore, be difficult to see why we should find it appealing. I now want to suggest that one indicator of their appeal is the extent to which we are alarmed by the possibility of debunking explanations of morality's rational authority.

Robert Stern independently identifies the possibility of debunking explanations as a motivation for Korsgaard's rejection of source externalism (2010, 2013). He observes that the Normative Question is evidently motivated by some kind of moral scepticism. Yet, he notes, we need to be careful not to conflate this with certain familiar types of moral scepticism. To begin with, the Normative Question is not motivated by the amoralist's question of 'Why be moral?', echoing Glaucon's challenge to Socrates in the *Republic* (357b). That is, it is not motivated by the kind of moral scepticism that asks for a non-moral reason to be moral, from the point of view of someone who does not already recognise morality as normative and rationally authoritative (Stern 2010, 2013: 24). Second, the Normative Question also needs to be distinguished from the kind of moral scepticism that doubts the validity of moral requirements because of their supposedly "queer" ontological or epistemological status (Stern 2013: 24, cf. Mackie 1977).

Instead, Stern suggests, the Normative Question is motivated by the prospect that an explanation of our attitudes towards morality might not *vindicate* these attitudes (2013: 24). In her own discussion of the Normative Question, Korsgaard writes (1996: 14):

A theory that could explain why someone does the right thing—in a way that is adequate from a third-person perspective—could nevertheless fail to justify the action from the agent's own, first-person perspective, and so fail to support its normative claims.

As an example of such a theory, Korsgaard asks us to imagine a theory that explains our attitudes towards morality by reference to their tendency to advance the reproduction of

our genes (1996: 15). She argues that, faced with such a theory, we might wonder whether following moral requirements is still “worth it” (1996: 15). Stern remarks (2013: 24):

What we have here, then, is a *debunking explanation* of morality, of the sort that makes it pressing to question the point of favouring it over our interests, or accepting what it tells us about how we should act.

Imagine an agent who believes that morality determines what she is rationally required to do, despite the supposed ‘queerness’ of moral requirements, and even if it means sacrificing her own self-interest. Now, suppose that this agent is told that her propensity to accord this kind of importance to morality is a clever trick by which her genes secure their own reproduction. The agent might now begin to toy with a certain kind of moral scepticism after all.

Now, what form does this moral scepticism take? Stern seems to suggest that the possibility of debunking explanations prompts the agent to take seriously the two forms of scepticism that she had previously cast aside—the scepticism that asks for a reason to be moral and the scepticism that emphasises the supposed queerness of moral requirements (ibid.). However, I think that the possibility of debunking explanations pose a *special* sceptical threat which is distinct and independent from these two forms of scepticism. And I want to suggest that we need to appeal to the idea of a critique of reason in order to account for this special sceptical threat.

The sceptical threat that I have in mind is one that persists even if source externalism is true. It is therefore to be distinguished from a different objection to source externalism, which also appeals to debunking explanations. In particular, it is sometimes argued that, in light of the evolutionary development of our practical reason, it is not to be expected that this faculty would be responsive to stance-independent irreducibly normative properties at all (Street 2006). Therefore, this line of argument concludes, such properties cannot furnish the source of the normativity of moral requirements. By contrast, the form of scepticism that I have in mind takes it for granted that our faculty of practical reason is responsive to stance-independent normative properties. However, supposing that our responsiveness to such properties is due to, say, our evolutionary history, this form of scepticism questions in how far moral requirements whose normativity has its source in these properties can possess *rational* authority.

If we were firmly in the grasp of the source externalist conception of reason as a receptive faculty, we would not be able to register this form of scepticism. According to this conception, the only credential that our practical reason has—the feature that marks it out as a faculty of *reason* in the first place—is its responsiveness to the right kind of

inputs: the substantive considerations that are *reasons* by virtue of their relation to stance-independent normative properties—and thus independently of *our* faculty of practical reason. In other words, source externalism suggests that the referent of ‘the faculty of practical reason’ or ‘rationality’ is to be fixed by a *stipulation* (Korsgaard manuscript: 4). As Korsgaard puts it (2009a: 6):

[I]n theories of this kind “human reason” is really nothing more than the name of that faculty within us, whatever it might be, that conforms to rational standards. It is not identified in any other way.

According to source externalism, moral requirements possess rational authority because the faculty by which we are responsive to these requirements is a faculty *of reason*. And it is a faculty of reason because being responsive to the right kind of substantive inputs is *all it means* for something to be a faculty of reason—regardless of how this faculty *came to be* responsive with those inputs.

I suspect that many people will not find this approach satisfactory. Many of us will be alarmed by the notion that our rational faculties are only accidentally (e.g., by virtue of evolutionary mechanisms) responsive to normative properties and, just by virtue of this accidental responsiveness, are properly called *reason*. I submit that the explanation why we are alarmed by this notion is that we would hope that something more could be said to vindicate the standards of reason as authoritative. If what makes the verdicts of our faculty of reason authoritative is the mere fact that this faculty happens to pick out those substantive considerations that are *reasons*, then the fundamental standards of reason—the standards by which it picks out those substantive considerations in the first place—do not have any justification that is *transparent* to us. Either our faculty of practical reason tracks the correct substantive inputs (in which case it is properly called a faculty of reason and its verdicts possess rational authority) or it does not (in which case its verdicts do not possess rational authority). There is nothing that we, as reasoners, can do, merely by investigating the workings of our practical reason, to ascertain that its verdicts possess rational authority. In other words, there is no possibility of a critique of reason.

It seems, then, that the unease that we feel at the prospect of debunking explanations of our propensity to accord rational authority to morality is explained by the appeal of the Kantian idea of a critique of reason. What we need, in order to overcome this unease, is a critical vindication of our faculty of practical reason itself. In particular, we need to be reassured that the verdicts of our practical reason possess the authority of reason by virtue of the way in which reason *produces* them. To provide such reassurance,

reason's workings must be transparent to itself. If we have such reassurance, the possibility of debunking explanations will not be able to shake our confidence in the rational authority of moral requirements. After all, if the verdicts of practical reason possess rational authority by virtue of the inner workings of that faculty—and practical reason commands that we follow moral requirements—then the rational authority of moral requirements will not depend on how we came to possess and exercise the faculty of practical reason in the first place.

One way to express this thought is to say that if the source externalist conception of reason is true, the explanation of why we possess the faculty of practical reason and the explanation of why that faculty is responsive to stance-independent irreducibly normative properties are the *one and the same*: we happen to have developed a faculty which is responsive to these properties (and therefore called 'practical reason'). By contrast, if a critical vindication of reason is possible, then the verdicts of our faculty of practical reason will be explained by the internal standards of that faculty, and that explanation holds independently of how that faculty came about.

3.3 Formulating the Question

I have argued that the Normative Question is concerned, not with moral normativity *per se*, but with the relation between moral normativity and rationality. In particular, when faced with a particular account of the sources of moral normativity, the Normative Question asks how, on that account, the relation that we commonly take to hold between moral normativity and our faculty of practical reason is to be explained.

James Dreier, who also interprets the Normative Question as a question about the link between rationality and normativity, notes that our formulation of the Normative Question will depend on how exactly we characterise the link between moral normativity and rationality (2015: 174-176). Dreier himself holds that we are irrational if we recognise that we ought to φ but fail to be *motivated* to φ . In his view, then, our failure to φ is not irrational if, knowing all the relevant facts, we do not recognise that we ought to φ . He argues that there is a difference between the failure to recognise what we ought to do, a kind of normative "blindness", and the failure of irrationality (Dreier 2015: 176,

cf. Scanlon 1998: 25, *pace* Parfit 1997: 99, 2011a: 119-120). Accordingly, Dreier formulates the Normative Question as follows (2015: 177):

(NQ1) Why is it irrational to fail to be motivated to do what one believes one ought to do?²³

From the point of view of Kantian constructivism, Dreier's use of the label 'irrational' is still not restrictive enough. On the Kantian analysis of obligation, as laid out in Chapter 1, an agent who is obligated to φ is already motivated to φ , by her respect for the law. An agent who does not have this motive does not possess pure practical reason, and can therefore not be judged to be irrational by virtue of her failure to be *motivated* to φ . In order to formulate the Normative Question in such a way that it can figure in an argument for Kantian constructivism, then, we should reserve the label 'irrational' for the failure to *act* on one's motive to do what one believes one ought to do:

(NQ2) Why is it irrational to fail to act on one's motivation to do what one believes one ought to do?

Note that this formulation of the Normative Question does justice to the Kantian analysis of obligation while retaining the spirit of Dreier's formulation. After all, on Dreier's interpretation, the question is ultimately concerned with how moral normativity interacts with practical reason in such a way as to yield a motivation to do what we are morally required to do. In other words, the guiding assumption of the Normative Question is that the motivation to do what one ought to do is *produced* by our faculty of rationality (cf. Dreier 2015: 176-179).

Keeping in mind that the Normative Question arises even—or, if the Kantian analysis of obligation is correct, only—for agents who are already motivated to do what they believe they ought to do, we can reformulate (NQ2) as follows:

(NQ3) Why is it irrational to fail to do what one believes one ought to do?²⁴

²³ For simplicity, I shall speak as if every violation of a requirement of rationality is an instance of irrationality. Yet, none of my arguments depend on that being the case.

²⁴ One might object that the most that rationality can require of us is that we *intend* or *try* to perform the act, φ , that morality requires of us. Now, I would argue that since 'ought' implies 'can', morality will not require any acts which we cannot be rationally criticised for failing to perform. However, I do not want to put much weight on this argument. Readers who are not convinced by it should therefore feel free to read my interpretation of the Normative Question as the question of why we are rationally required to *intend* or *try* to do what we believe we morally ought to do.

From now on, I shall primarily use (NQ3) as my formulation of the Normative Question.²⁵

Note that the ‘ought’ in (NQ3) is not explicitly identified as a moral ‘ought’. And, indeed, I see no reason why we should not be able to raise the Normative Question about non-moral normativity, too. Yet, my discussion focuses on moral normativity because Kantian constructivism is a view about moral obligation and, accordingly, cannot be expected to answer the Normative Question for other parts of the normative domain. I say a few more words on this in Section 5.2 of the next chapter.

One might suspect that, by tailoring the Normative Question to a use of ‘irrational’ that is compatible with the Kantian analysis of obligation, I have given Kantian constructivism an unfair advantage. However, note that *any* of the proposals regarding the link between ‘ought’ and ‘irrational’ discussed above would deem it irrational to fail to act on one’s motivation to do what one believes one ought to do. If anything, by using the label more restrictively than Dreier and Parfit, I am making it *easier* to answer the Normative Question. After all, if any of the failures discussed here—the failure to recognise what one ought to do, the failure to be motivated to do what one believes one ought to do, and the failure to act on this motivation—qualifies as a practical irrationality, it is the latter.

But now one might object that I am making it *too* easy for Kantian constructivism to answer that question. That is, one might argue that even if Kantian constructivism answers the Normative Question as I have formulated it, it does not follow that it has a dialectical advantage vis-à-vis competing views because it has not answered the Normative Question for the less restrictive use of ‘irrational’ that these views favour. However, Kantian constructivism cannot be expected to answer the Normative Question for a less restrictive use of ‘irrational’. After all, why should Kantian constructivists explain the irrationality of something they do not deem to be irrational? Nevertheless, Kantian constructivism retains an advantage over competing views by explaining the link between rationality and moral normativity for a class of actions that *all* participants in the debate deem rationally required.

²⁵ Note that (NQ3) seems to presuppose that we are irrational if we fail to do what we believe we ought to do, *whatever* we believe we ought to do. One might object that it is irrational to fail to do what one believes one ought to do only if one’s relevant beliefs are *rational*. I ask readers who hold this view simply to assume, for the discussion in this chapter and the next, that we are concerned exclusively with cases in which the agent’s belief about what she ought to do are rational. Nothing I say is incompatible with this assumption.

3.4 *The Dialectic*

Before I discuss whether and how source externalism and other views answer (or fail to answer) the Normative Question in the next chapter, I would like to make some dialectical remarks on what counts as an answer to the Normative Question, and what cost a view incurs by failing to provide such an answer.

It is important to note that the Normative Question is *not* the question of the source of morality's normativity, which I discussed in Chapter 1. Any view that is under consideration here provides an answer to *that* question. There is nevertheless an important connection between the Normative Question and the question of the source of moral normativity. In particular, one would expect a satisfactory answer to the source question to answer the Normative Question also. An account that explains why morality is normative without explaining its link to our rational faculties explains normativity in a way that renders puzzling one of the features that we most naturally ascribe to it. In short, such an account creates a puzzle where there should be none. It leaves us with an explanatory gap—between morality's being normative and morality's having rational authority—where we would not expect to find one.

This highlights that the Normative Question is raised from the point of view of an agent who already believes that she is rationally required to do what she believes she morally ought to do. By asking the Normative Question with respect to some view of the source of moral normativity, the agent challenges that view's ability to *vindicate* the rational authority that she attributes to morality. This has an important implication for the dialectic of the Normative Question: in answering the Normative Question one cannot *rely* on the plausibility of the notion that there is a close link between moral normativity and practical reason. That is, it is dialectically inadmissible to respond to the agent's question, 'Why am I rationally required to do what (I believe) I morally ought to do', by stating that doing what you believe you ought to do is just what it means to be rational. After all, the Normative Question is precisely the question of *why* this is the case, *given* a certain view of moral normativity (cf. Dreier 2015: 173-174, 176-179).

More generally, my interpretation of the Normative Question suggests that it should not be understood as an independent, stand-alone challenge to source externalism. Instead, it is motivated by a complex of philosophical considerations that provide a rather deep explanation for Korsgaard's rejection of source externalism. The question of why one is rationally required to do what one believes one is morally required to do, viewed in isolation, might appear innocuous. To give an answer that does justice to the philosophical considerations that motivate this question, however, is far from trivial. In

order to do justice to these considerations, an answer to the Normative Question must be able to explain, in a manner that is *transparent to the rational agent*, why she is rationally required to do what she believes she morally ought to do; for only then will this answer do justice to the idea of a critique of reason that motivates the Normative Question.

How high is the cost of failing to answer the Normative Question? Thomas Scanlon seems to think that an account's inability to answer the Normative Question does not constitute a significant objection to that account (2014: 10-14). He distinguishes between two types of reasoning, labelled "external" and "internal" reasoning (2014: 12-13). Note that this distinction is not related to the distinction between internalist and externalist theories of the sources of normativity, which I introduced in Chapter 1. Instead, by "external reasoning", Scanlon refers to the type of reasoning that is employed in "a context in which two people are arguing about what reasons for action one of them has" (2014: 12). According to Scanlon, this type of reasoning is to be contrasted with "internal reasoning", "which is reasoning about what reasons one has oneself" (2014: 13, cf. Harman 2011).

About *external* reasoning, Scanlon says (2014: 13):

It is an obvious dialectical advantage to be able to "get a grip" on your opponent by saying: "But you accept that ... and it follows from this that the fact that *p* counts in favor of *phi*-ing."

By contrast, about *internal* reasoning, he writes (2014: 14):

But such claims are not as relevant in internal reasoning as in the external variety, and it is the point of view of internal reasoning that is primary in an investigation of reasons and normativity. From this point of view the question of how reasons "get a grip on one" properly disappears. There is only the question: what reasons do I have?

Scanlon thus seems to think that being able to answer the Normative Question is helpful for external reasoning. When it comes to internal reasoning, however, our questioning naturally comes to an end once we encounter normative claims that are "clearly correct" on account of their "substantive merits" (ibid.). For example, Scanlon holds that it is "obviously true" that "[f]or a person in control of a fast moving automobile, the fact that the car will injure and perhaps kill a pedestrian if the wheel is not turned is a reason to turn the wheel" (2014: 2-3). Scanlon seems to suggest that once we encounter a normative

truth like this we do not wonder *why* we are rationally required to act in accordance with it.

There are several things to note about Scanlon's remarks. First, it is unclear why the point of view of internal reasoning, as opposed to that of external reasoning, should be "primary in an investigation of reasons and normativity" (2014: 14). To begin with, the question of 'what reasons one has oneself' belongs to first-order normative theory, which hardly exhausts our 'investigations of reasons and normativity'. In fact, even first-order normative theory does not seem to be confined to the point of view of internal reasoning. Many first-order normative theorists seem to be concerned with convincing each other of normative claims that are not 'obviously true' (cf. Kagan 1991: 11-15).

Second, and more importantly, the distinction between internal and external reasoning is a red herring. Both kinds of reasoning are ultimately concerned with what reasons we have, albeit from different points of views. Yet, as Scanlon himself seems to realise, the Normative Question is not the question of what we morally ought to do, but of why we are rationally required to act in accordance with our answer to this question (cf. Scanlon 2014: 10).²⁶ Therefore, by focusing on the question of whether internal or external reasoning should be our primary concern in normative theory, the Normative Question is ruled out as a legitimate concern for normative theory from the start.

Thus, the real question is not whether internal or external reasoning is primary but whether the question of what reasons we have—as opposed to the Normative Question—is primary. Perhaps the point that Scanlon *really* means to drive home by prioritizing the perspective of internal reasoning is that *those* normative investigations that commence *from within the agent's point of view* are limited to the question of what reasons we have, and thus do not give rise to the Normative Question. He might mean to suggest that an agent who is already convinced that she is rationally required to do what she believes she morally ought to do will only be concerned with the question of *what it is* that she ought to do (cf. Scanlon 2007: 85-86). Against this claim, I can only point to the motivation for the Normative Question which I outlined above: the appeal of the Kantian idea of a critique of reason, which expresses itself through the unease we feel in

²⁶ Admittedly, Scanlon does not state the Normative Question in those exact terms. By Scanlon's lights, in a situation where *p* is a reason for an agent, the Normative Question asks for "some further explanation [...] of why the agent in this situation *must* treat *p* as a reason" (2014: 10, cf. Scanlon 2007). Yet, although Scanlon does not specify what it means to *treat* a reason as a reason, I believe that we can assume, for the purposes of this discussion, that his reading of the Normative Question is equivalent to mine.

the face of a certain type of debunking explanation. The cost of failing to answer the Normative Question is determined by how much weight and plausibility we accord this motivation. I hope that its weight and plausibility suffice to show that the Normative Question does indeed arise from the point of view of agents who are motivated to do what they believe they ought to do, and that the cost of failing to answer the Normative Question is significant.

4. Conclusion

In this chapter, I have argued that the Normative Question should be interpreted as the question of why it is irrational not to do what one believes one morally ought to do. Answering this question is far from trivial. Since the Normative Question is motivated by the Kantian idea of a critique of reason, an answer to this question will have to explain the connection between our faculty of practical reason and moral normativity in a way that makes this connection *transparent* to us as reasoners. In the next chapter, I argue that Kantian constructivism provides such an explanation while source externalism, as well as non-Kantian versions of source internalism and source voluntarism, do not.

Chapter 3: Answering the Normative Question

1. Introduction

In the previous chapter, I interpreted the Normative Question as the question of why we are rationally required to do what we believe we morally ought to do. In this chapter, I discuss potential answers to this question. I begin by arguing, in Section 2, that source externalism does not give an answer to the Normative Question, because it identifies normative properties as the sources of normativity without explaining how such properties relate to our rational faculties. In Section 3, I argue that non-Kantian versions of source internalism and source voluntarism also fail to answer the Normative Question, and for similar reasons. In Section 4, I turn to Kantian constructivism and argue that the notion that the source of moral normativity resides in the form of our will gives an answer to the Normative Question. I argue that, by relying on Rawls's notion of 'philosophy as defence', this answer avoids an important objection to constitutivist explanations of normativity, David Enoch's Shmagency Objection. Section 5 offers some clarificatory remarks on Kantian constructivism and the Normative Question.

2. Source Externalism

In interpreting and formulating the Normative Question, I have partly been guided by the desideratum that source externalism is not supposed to be able to answer this question. Yet, the accusation that source externalism does not provide an answer to this question may still seem puzzling. In particular, it may seem that Parfit's response to the Normative Question, as outlined in Section 2 of the previous chapter, still stands. Adapted to my formulation of the question, this response states: if you judge that you *ought* to φ , there is no open question of *why* rationality requires that you act on your motivation to φ (cf. Parfit 2006: 377-380, Scanlon 2014: 10). After all, as Dreier puts it on behalf of Parfit, rationality just *is* the faculty that, among other things, "produces [...] motivations to φ from beliefs that we ought to φ " (2015: 177).

Chang sums up this response to the Normative Question on behalf of source externalism (2013a: 170):

[I]f being rational involves recognizing one's reasons and responding appropriately to them, then it makes little sense to ask how reasons can be authoritative for rational persons—that is, why a rational person should do what she has reasons to do. So the way in which reasons whose normative source is given by normative facts 'get a grip' on us is through our capacity for rationality.

As persuasive as this response may appear initially, it runs afoul of the dialectical remark that I made in the previous chapter (Section 3.4). That is, it cites the crucial presupposition of the Normative Question—that we are rationally required to do what we believe we morally ought to do—in favour of the source externalist conception of moral normativity. According to this response, ‘it makes little sense to ask how reasons can be authoritative for rational persons’ because this is what rationality *is*. The problem is that this response *relies* on the very link between rationality and moral normativity that it is supposed to *explain*.

As I pointed out above, an answer to the Normative Question must explain this link in a manner that is transparent to the rational agent. But source externalism cannot deliver such an explanation. The problem is that, on the receptive conception of practical reason suggested by source externalism, stance-independent normative properties mark out certain substantive considerations (which are therefore called reasons) as the proper determinants of practical reasoning. We are rational insofar as we respond to these considerations because, according to the source externalist conception, this is what it means to be rational.

What is missing, in this picture, is an explanation of why practical reason tracks these substantive considerations. The source externalist will respond that it tracks these considerations because they stand in the right relation to stance-independent normative properties. However, this response does not do justice to the idea of a critique of reason. In order to do justice to that idea, we have to explain why stance-independent normative properties are relevant to practical reasoning in the first place. Source externalism seems to imply that, from the point of view of the rational agent, it is a *brute fact* that certain substantive considerations constitute proper inputs for practical reasons while others do not (Korsgaard 1996: 12, cf. Shafer-Landau 2003: 48).

Parfit disagrees (2006: 377):

Justifications can end with some irreducibly normative truth. And such truths are *not* brute facts. The most important normative truths could not have been false. If we ask why these truths are true, we can sometimes give no further answer. But, since these truths are not brute facts, they can provide full, or complete, justifications.

Parfit seems to think that truths about irreducibly normative properties are not brute facts, at least when it comes to answering questions about what we *should* do, precisely because they are *normative* truths. Adapted to my interpretation of the Normative Question, Parfit’s claim seems to be that, although there is no explanation of the fact that

stance-independent normative properties furnish the proper inputs to practical reason, this fact is not therefore to be regarded as a brute fact. In this picture, the normative character of these properties, by itself, provides a sufficient explanation for their link to rationality. Indeed, we might read Parfit as suggesting that the belief that we morally ought to φ entails that we are rationally required to φ .

To see why this picture is unsatisfactory, in light of the philosophical considerations that motivate the Normative Question, consider the following metaphor by Korsgaard (2008: 315):

If to have knowledge is to have a map of the world, then to be able to act well is to be able to decide where to go and to follow the map in going there. The ability to act is something like the ability to *use* the map, and that ability cannot be given by *another map*. (Nor can it be given by having little normative flags added to the map of nature that mark out certain spots or certain routes as good. You still have to know how to use the map before the little normative flags can be of any use to you.)

To paraphrase: our grasp of truths, by itself, is insufficient to determine what we are rationally required to do. We also need to be able to *apply* these truths in accordance with their relevance to rational decision-making. It does not help to designate some of the truths as *normative*—or, for that matter, as entailing that we are rationally required to act in a certain way; for we would have to be able to apply such truths in practical reasoning *prior* to learning that they have this feature. In applying a certain kind of truth, then, we cannot be guided by the normative character (or intrinsic relevance to what we are rationally required to do) of this kind of truth (cf. Korsgaard 2008: 53-54, manuscript). After all, we already need to know how, rationally, to apply or respond to these truths if we are to make anything of the fact that they have this character.

Hussain and Shah suggest a helpful reconstruction of Korsgaard's remarks (2006: 280-281). Adapted to my interpretation of the Normative Question, it runs as follows. Let us take for granted that our belief that we morally ought to φ , because moral normativity has its source in normative truths, entails that we are rationally required to φ . Korsgaard's complaint is that we cannot infer from this belief that we are rationally required to φ unless we already follow a principle according to which we are rationally required to do what we believe we are rationally required to do. Loosely following Hussain and Shah, we can formalise this principle as follows (cf. 2006: 280):

- (1) $RR(BRR(\varphi) \rightarrow \varphi)$

This principle says that, if you believe that you are rationally required to φ , you are rationally required to φ . As Hussain and Shah note, Korsgaard’s point is that the *content* of our belief is not sufficient to explain why we are rationally required to respond to it in a certain way; to explain this, we need to appeal to a principle that regulates the application of our normative beliefs in practice. But we cannot come to be guided by this principle by forming the following belief (Hussain and Shah 2006: 280):

$$(2) \quad \text{BRR}(\text{BRR}(\varphi) \rightarrow \varphi)$$

After all, the belief (2) is itself a belief of the form $\text{BRR}(\varphi)$, and thus cannot be applied in practical reasoning unless we are *already* guided by principle (1). This is why it does not help to emphasize that the stance-independent normative truths which, according to source externalists, determine what we are rationally required to do are *normative* truths: even if the content of my belief is that I am rationally required to φ (or, for that matter, that I am rationally required to φ if I believe that I am rationally required to φ), I cannot arrive at the conclusion that I am rationally required to φ unless I am already guided by the principle that I am rationally required to φ if I believe that I am rationally required to φ .²⁷

²⁷ Source externalists might agree that, for a full explanation why we are rationally required to φ if we believe that we morally ought to φ , we need to appeal to a principle of rationality of this form. Yet, they might deny that this is a problem for their view. After all, they might argue, it is commonly held that the reasons provided by stance-independent normative properties (which make it the case that we morally ought to φ) are not the only things that have genuine normativity; in addition, there are normative requirements of rationality that demand that our attitudes exhibit certain patterns of coherence (e.g., Broome 1999, Darwall 1983: 14-16, Parfit 2011a: 113, Scanlon 2007, Wallace 2001). And the principle stated in (1) might be taken to express just such a requirement. There are two problems with this argument.

First, if an independent normative requirement of rationality provided the needed answer to the Normative Question, then this confirms my claim that source externalism, by itself, cannot answer this question. Second, and more importantly, a requirement of *coherence* cannot provide an answer to the Normative Question. To see this, consider that if my failure to φ is to be *incoherent* with my belief that I am rationally required to φ , then I must already know how to *apply* such a belief in my rational decision-making. One might respond that it is obvious how to apply such a belief in rational decision-making; after all, it is a belief *about* what I am rationally required to do. Yet, as I have pointed out, I cannot, rationally, be guided by the *content* of a belief in determining whether and how that belief is to be used in rational decision-making. Now, one might propose the relevant normative requirement as a *substantive* requirement rather than as a

Hussain and Shah argue that, although the reasoning just presented is sound, it does not follow that source externalism is false (2006: 281). After all, even if we cannot *come to follow* principle (1) by grasping stance-independent normative truths, such truths can still be what makes it *correct* to follow this principle. In their words, “what makes something correct to do and what one’s grounds are for doing it need not coincide” (ibid.). Fortunately, I can happily grant this. The aim of my argument is not to show that source externalism is false but that it cannot answer the Normative Question. And to show this, it is sufficient to point out that an agent cannot, rationally, come to be guided by substantive considerations merely on the basis that they stand in a certain relation to stance-independent normative properties. Practical reason must already be following standards that render such properties relevant to its operations if their normative nature is to make a difference to these operations.²⁸ The fact that, on the receptive conception of practical reason suggested by source externalism, these standards are opaque to reason’s critical scrutiny renders source externalism incapable of providing an answer to the Normative Question. This sheds light on Korsgaard’s otherwise obscure statement that source externalism “seems to leave us in need of a reason to be rational” (2008: 53). The thought is that, on the source externalist picture of reason, the authority of reason is not fully *explicable* to reason itself.

requirement of coherence. But then the Normative Question arises all over again; for then the question becomes how this substantive requirement can be rationally vindicated.

²⁸ Perhaps, Parfit’s claim is not that the *normative* character of stance-independent normative properties explains their link to rationality. Instead, his argument might be that truths about such properties are *necessary* truths, and *that’s* why they provide the proper inputs to practical reasoning. He writes (2011a: 129):

[Normative] beliefs are about truths that are not empirical and contingent, but necessary. Undeserved suffering, for example, could not have failed to be in itself bad. For such normative beliefs to be rational, we do not need to have evidence that they match the actual world, since these beliefs would be true in any possible world.

However, this argument fails for similar reasons as the one I discussed in the main text. The problem is that the necessity of the truths in question is not ascertained by a standard that is transparent to us as reasoners. It can only be part of the *content* of our beliefs about these truths, and therefore cannot explain how and why our practical reason should be determined by such beliefs in the first place (cf. Korsgaard manuscript: 6-7).

3. Source Internalism and Source Voluntarism

Although I have interpreted the Normative Question as a challenge primarily to source externalism, I now want to argue that non-Kantian versions of source internalism and source voluntarism cannot provide an answer to this question either.

Recall that source internalism and source voluntarism hold that morality's normativity has its source in our motivational states or volitional states, respectively. It is important to note that, according to non-Kantian versions of these views, the motivational or volitional states that explain why morality is normative are *antecedent* states in the sense that they are independent of morality's normativity (see Chapter 1, Section 4.4 above). My discussion in this section is restricted to such non-Kantian versions.

Initially, it may appear as if source internalism and voluntarism are in a better position to answer the Normative Question than source externalism. Where source externalism asserts that the normative properties which it regards as the ultimate bearers of moral normativity are simultaneously the source of their own normativity, source internalism and voluntarism explain the normativity of morality by appeal to something that lies within our own will or motivational states, and thus, it may seem, closer to our faculty of practical reason. However, the Normative Question remains: why, if moral normativity has its source in our motivational or volitional states, are we rationally required to do what we believe we are morally required to do? After all, why should practical reason attribute such importance to our motivational and volitional states?

The problem is that, while our antecedent volitional and motivational states are in some sense internal to us, they are still external to *reason*. Accordingly, moral requirements whose normativity has its source in such states can only function as substantive inputs to reasoning and are therefore beyond the critical scrutiny of reason. Of course, we can critically reflect on these inputs to some extent, e.g., by asking whether φ -ing really stands in such a relation to my antecedent volitional and motivational states as to make it the case that I ought, or have reason, to φ (cf. Williams 1981). But once it is ascertained that φ -ing stands in the relevant relation, I cannot expect source internalism or voluntarism to provide a further explanation of why I am rationally required to φ . The assurance that, given the relation between φ -ing and my antecedent motivational or volitional states, I really ought, or have reason, to φ is all that these views can provide.

Source internalists and voluntarists might contend that complying with requirements whose normativity is explained by our motivational or volitional states is part of what it is to be rational. Sharon Street suggests a reply along these lines (2012: 51-52, 58). According to this reply, while it is sensible to ask for an explanation of why morality

is normative—the answer being that moral normativity is explained by our motivational or volitional states—it is not sensible to ask for an explanation of why it is *rationally required* to act in accordance with moral normativity, so construed.

However, this reply does not answer the Normative Question; instead, it claims that this question cannot and need not be answered. After all, this reply suggests that what we are rationally required to do *just is* determined by requirements grounded in our motivational or volitional states, and that there is no further explanation of *why* such requirements are accorded rational authority. As Street acknowledges, this reply effectively denies Kant’s remark that “one cannot possibly think of a reason that would self-consciously receive guidance from any other quarter with regard to its judgments” (*G* 4:448, Street 2012: 58).²⁹ But to deny this is to deny the possibility of a critique of reason. After all, if reason must accept that its verdicts are determined by requirements that are grounded in our motivational or volitional states, i.e., by the requirements of an external authority, then the rational authority of reason’s own workings is not fully explicable to reason.

Of course, one might argue that this is not a major problem since the idea of a critique of reason is ill-founded to begin with. Indeed, Street argues that, if her version of source internalism is true, it does not make sense to ask why we should accord authority to the (set of) motivational states that are the source of morality’s normativity because to step back from these motivational states is to abandon the standpoint of practical reason (2012: 51-52). Connectedly, Street seems to suggest that, if her view is correct, the availability of debunking explanations of the kind I mentioned in the previous chapter is not a threat to the rational authority of moral requirements: what matters is that moral normativity is explained by our motivational states, not whether these motivational states can in turn be rationally vindicated (2012: 52-54).³⁰

But these arguments do not support the conclusion that source internalism or voluntarism can answer the Normative Question. To the contrary, they suggest that *either* source internalism and voluntarism are false, *or* the Normative Question is ill-founded. Accordingly, these arguments count in favour of rejecting the Normative Question only if one already has independent reason to believe that source internalism or voluntarism are true. By contrast, if one thinks that the Normative Question is not ill-

²⁹ Note the close resemblance to Hume’s *dictum* that “reason is, and ought only to be the slave of the passions” (1896: 217).

³⁰ Perhaps it is more accurate to say that Street argues that, if her view is true, no explanations of our propensity to accord rational authority to morality could possibly ‘debunk’ this authority.

founded, these arguments count in favour of rejecting source internalism and voluntarism.

In any case, this section was meant to establish only that non-Kantian versions of source internalism and source voluntarism fail to give an answer to the Normative Question. We have seen that this is the case because, like source externalism, these views locate the source of moral normativity outside the scope of reason's critical scrutiny. In other words, they depict reason as a receptive faculty whose workings are determined by inputs which cannot be rationally vindicated and must therefore function as axioms.

4. Kantian Constructivism

I now want to argue that Kantian constructivism, as opposed to source externalism and non-Kantian versions of source internalism and source voluntarism, provides an answer to the Normative Question. This answer involves two major steps. First, Kantian constructivism provides an explanation of the link between practical reason and moral normativity by depicting this normativity as having its source in the very form of practical reason. Second, Kantian constructivism suggests that this explanation of the link between moral normativity and practical rationality is corroborated by the Kantian analysis of obligation. It thus finds support in the practical phenomenology of agents who already deem themselves morally obligated and accord morality rational authority, i.e., in the point of view that gives rise to the Normative Question in the first place. I argue that Kantian constructivism therefore provides a non-foundationalist vindication of morality's rational authority.

4.1 The Form of Practical Reason

According to Kantian constructivism, morality's being normative is grounded in our possessing autonomy of the will. In short, our autonomy explains why morality is normative for us. I now want to suggest that Kantian constructivism's view of the source of moral normativity also provides an explanation of that normativity's link to practical rationality.

If we take the idea of a critique of reason seriously, we cannot accept any putative verdict of reason as a self-evident foundation for rational thought. This might seem to suggest that the project of a vindication of reason is doomed to failure. O'Neill formulates

the worry on behalf of those who are sceptic about the project of a critical vindication of reason (2015: 1):

Surely any attempt to vindicate standards or principles of reason must fail, because nothing can count as a vindication or justification, unless it is itself reasoned. Yet, if it is reasoned, it will presuppose and so cannot vindicate principles or standards of reason. But if it is not reasoned, it will fail to vindicate principles or standards of reason.

Fortunately, however, our situation is not as dire as it might appear. So far, we have only considered the approach of vindicating the rational authority of a substantive claim (e.g., ‘I am rationally required not to torture people for fun’) by *deducing* it from another, more fundamental substantive claim (e.g., ‘I am rationally required not to inflict gratuitous pain’). If we follow this approach, the best we can hope for is that, eventually, the deduction of our initial substantive claim will bottom out in a substantive claim that seems self-evident. Lest we find ourselves in an infinite regress, we will treat this substantive claim as an axiom of rational justification. But if our chain of rational justification bottoms out in an axiom, then, of course, a complete vindication of rational authority is impossible.

It is important to see that this problem arises because the axiom with which we attempted to justify the rational authority of our initial substantive claim is *itself* a substantive claim. There is nothing we can say about *why* a claim of a certain substance should determine what we are rationally required to do, other than that ‘it just *does*’. In Scanlon’s words, some verdicts of reason are judged to be “clearly correct” on account of their “substantive merits” (2014: 14). On this approach, reason is not fully *transparent* to itself because the standards by which we make these fundamental substantive judgments lie outside the reach of reason’s critical scrutiny.

This means that we cannot provide an ultimate rational vindication of reason’s verdicts—or, at least, none that does justice to the idea of a critique of reason—by deducing these verdicts from a positive, substantive foundation. However, we might still be able to show that a claim is at least *consistent* with the exercise of reason. Reason might provide us with a *negative* standard by ruling out certain claims as irrational on the grounds that they are *inconsistent* with the unhindered workings of our faculty for criticism. By following such a negative standard, we might at least make sure that reason does not *overstep its boundaries* by asserting or accepting claims whose foundations have not themselves been rationally examined (cf. O’Neill 2015: 27-37).

Kant seems to express a similar, if not identical, thought in the following passage (*CPR* A 711/B 739):

Where, as in the case of pure reason, we come upon a whole system of illusions and fallacies, intimately bound together and united under common principles, its own and indeed negative lawgiving seems to be required, which, under the title of a discipline, erects a system of precautions and self-examination out of the nature of reason and the objects of its pure employment.

Of course, reason's 'negative lawgiving' cannot be based upon some substantive claim that strikes us as self-evident. However, if this negative lawgiving proceeds on the basis of the very *form* or *constitution* of reason, it can provide us with a standard for critique without undermining itself. For, then, we have a minimal standard of reason that rules out certain propositions because they are incompatible with the very idea of reason (cf. O'Neill 2015: 162-196, *WOT* 8:144).

Now, this is precisely the role played by the Categorical Imperative, according to Kantian constructivism. In this view, the supreme principle of morality, which determines the content of all our moral obligations, is the constitutive principle of the will. That is, it rules out maxims as impermissible if they are irreconcilable with the form of practical reason. According to Kantian constructivism, the fact that the Categorical Imperative is the constitutive principle of practical reason, i.e., the fact that we possess autonomous wills, is what *grounds* morality's being normative.

Kantian constructivism escapes the problem that, as we have seen, haunts source externalism, as well as non-Kantian versions of source internalism and voluntarism, because it does not depict reason as *receptive*. According to Kantian constructivism, practical reason's very constitution provides us with a normative standard by which we can rule out the willing of certain maxims as irrational. Instead of relying on substantive inputs that serve as axioms for the operations of reason, Kantian constructivism locates the source of morality's normativity in the form of our will, and thus in the conditions of the possibility of practical reasoning. Thus, when it comes to answering the Normative Question, Kantian constructivism does not attempt to explain why our moral beliefs are relevant to our faculty for rational decision-making in terms of the content of these beliefs. Kantian constructivism bridges the gap between moral normativity and practical reason from the opposite direction, as it were, by explaining the former in terms of the constitution of the latter. This reversal is part of Kant's ambition to bring about a Copernican Revolution with regard to philosophy's approach to reason (*CPR* B xvi-xviii, cf. Sensen 2011: 2).

We might say that, when confronted with the need for an explanation of the link between moral normativity and practical reason, Kantian constructivism points out that it has already provided one. After all, autonomy of the will, which Kantian constructivism identifies as both determining the content of our moral obligations and providing the source of their normativity, is the form of practical reason. For Kantian constructivism, the task of answering the question of the source of moral normativity and the task of answering the Normative Question are one and the same. A single set of conceptually linked notions—autonomy, the form of law, and the motive of duty—delivers an account of the source of moral normativity, its link to practical rationality, and the content of our moral obligations.

4.2 Two Challenges

Kantian constructivism promises to vindicate the rational authority of moral obligations by depicting moral normativity as having its source in the form of practical reason. However, if it is to fulfil this promise it must overcome two challenges.

The first challenge is to show that autonomy of the will really *is* the form of our practical reason. To vindicate that we are rationally required to do what we believe we morally ought to do, it is not sufficient to point out that a certain formal characterisation of practical reasoning can serve both as the source of moral normativity and as a bridge between that normativity and practical reason. For all we know, this characterisation might not accurately capture our faculty of practical reason. In more Kantian terms, it might be just another instance of reason *overstepping its boundaries*, making up a mere “phantasm”, a “chimerical idea without truth” (*G* 4:445).

Within Kant’s own framework, the question of whether the formal characterisation of our will as autonomous is accurate amounts to the question of whether we possess transcendental freedom. This is because Kant takes transcendental freedom to be both necessary and sufficient for autonomy of the will (*CPrR* 5: 3-4). But Kantian constructivism, as I characterise it, remains silent on the relation between freedom and autonomy. Accordingly, for Kantian constructivism, the question is whether we possess autonomy of the will, whatever this might entail for the metaphysics of free will.

The second challenge is a version of David Enoch’s Shmagency Objection to constitutivist explanations of normativity (2006, 2011). As Enoch notes, his objection can be

summed up with the following quote from David Lewis (Enoch 2006: 171, Lewis 1996: 307):

Why care about objective value or ethical reality? The sanction is that if you do not, your inner states will fail to deserve folk-theoretical names. Not a threat that will strike terror into the hearts of the wicked! But who ever thought that philosophy could replace the hangman?

Enoch's objection to constitutivism is that, even if certain principles can be shown to be constitutive of a certain activity, then by violating these principles we will merely cease to *count* as engaging in that activity. But this will vindicate the notion that we are rationally required to comply with these principles only if we are *already* required to engage in the relevant activity. Otherwise our failure to engage in this activity will only be a failure to be describable in certain 'folk-theoretical names', e.g., as 'acting'.

To clarify Enoch's Shmagency Objection, it might help to see how it applies to a concrete example, such as Korsgaard's constitutivism (Korsgaard 2009a). I outline and discuss Korsgaard's view in more detail in the next chapter, but for now the only important thing to note is that Korsgaard takes the aim of self-constitution to be constitutive of agency such that, if we did not have that aim, we would not qualify as agents. Enoch argues that the person who raises the Normative Question might say (2006: 179):

Perhaps I cannot be classified as an agent without aiming to constitute myself. But why should I be an agent? Perhaps I can't act without aiming at self-constitution, but why should I act? If [Korsgaard's] reasoning works, this just shows that I don't care about agency and action. 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. I am perfectly happy performing shmactions—nonaction events that are very similar to actions but that lack the aim (constitutive of actions but not of shmactions) of self-constitution.

Now, I have not characterised Kantian constructivism as the view that self-constitution is constitutive of agency. Indeed, I shall reject such a characterisation in the next chapter. Nevertheless, the general strategy underlying Enoch's Shmagency Objection might also seem to apply to the answer that Kantian constructivism, on my construal, gives to the Normative Question. After all, I have been arguing that Kantian constructivism answers the Normative Question by invoking the Categorical Imperative as a formal, constitutive standard of practical reason. Thus, the person who raises the Normative Question might

reply: ‘Perhaps acting on the Categorical Imperative is constitutive of being a practical reasoner or having a will, but why am I rationally required to care about being a practical reasoner or having a will?’

Note that this challenge resembles the objection that I, reconstructing Korsgaard’s arguments, raised against source externalism (see Section 2 above).³¹ The objection was that beliefs about stance-independent properties, by themselves, cannot yield the conclusion that we are rationally required to act in a certain way. After all, we will not know how to apply our beliefs about such properties in rational decision-making unless we already follow a principle that links them to our faculty of practical rationality. The analogous objection to Kantian constructivism is that the properties of ‘being a practical reasoner’ or ‘having a will’, although they are instantiated within our own agency rather than in the external world, still seem to be mere *descriptions* of activities (cf. Korsgaard 2008e: 325). And, unless we are already guided by a principle for the application of such descriptions in rational decision-making, we will not be able to conclude that we are rationally required to respond to them in a certain way. In Korsgaard’s own remarkable words, we suddenly seem to be “in need of a reason to be rational” (2008b: 53, cf. Wallace 2004: 455). Therefore, the objector concludes, the Kantian constructivist view of the sources of moral normativity is just as incompatible with the idea of a critique of reason as its source externalist competitor.

4.3 *Philosophy as Defence*

In what follows, I respond to the two challenges to the Kantian constructivist answer to the Normative Question outlined above. In this sub-section, I respond to the first challenge by clarifying the ambitions and dialectical strategy of the Kantian constructivist answer. In the following sub-section, I show that, on the picture that emerges, the Kantian constructivist answer is not vulnerable to the Shmagency Objection after all.

Recall that the Normative Question arises from the point of view of an agent who already has moral beliefs, is motivated to act on them, and believes that she is rationally required to do so (see Chapter 2, Section 3). That agent asks whether there is an account of the relation between practical reason and moral normativity that vindicates the authority that she attributes to morality. Accordingly, in order to answer the Normative Question, we need not convince an amoralist, i.e., someone who has no prior commitment to morality, that she is rationally required to do what morality demands.

³¹ For the same reasons, it also resembles the objection that I raised against non-Kantian versions of source internalism and voluntarism in Section 3.

The Kantian constructivist answer to the Normative Question, as I conceive of it, exploits this dialectical situation by citing the moral beliefs and motivation of the agent who raises the Normative Question in support of the formal characterisation of practical reason advocated by Kantian constructivism. More specifically, this answer appeals to the Kantian analysis of obligation, which I outlined in Chapter 1, to argue that the agent's moral beliefs and motivations can be shown to entail that she does indeed possess an autonomous will. It then uses the notion that the Categorical Imperative, the supreme principle of morality, is the constitutive principle of the autonomous will to explain why the agent is rationally required to do what she believes she morally ought to do, in the way outlined in Section 4.1.

On this construal, the Kantian constructivist answer to the Normative Question employs a strategy that resembles Kant's argument from the "fact of reason" (*CPrR* 5:30-31). Kant thinks that we are bound by the moral law if and only if we possess the agential capacity of autonomy of the will. Moreover, he thinks that autonomy requires nothing less than transcendental freedom (*CPrR* 5:29). That is, it requires the will to be able to cause actions spontaneously, i.e., without being determined by empirical laws of nature. However, although theoretical reason does not *rule out* the possibility of transcendental freedom, it also does not *confirm* that we are indeed free in this sense (*CPrR* 5:3). In light of this, Kant proposes a strategy that aims to establish transcendental freedom *indirectly*. We may not have a direct insight into our freedom, he argues, but we are consciously aware of being bound by the moral law. Kant suggests that we can appeal to this awareness to establish that we are indeed transcendently free.

He illustrates this by way of an example. Imagine an agent whose prince demands of him that, on pain of losing his life, that he "give false testimony against an honorable man" (*CPrR* 5:30). Kant invites us to ask, from the agent's point of view, whether it would be *possible* for him to refuse the prince's demand. Kant predicts the following response (*ibid.*):

He would perhaps not venture to assert whether he would do it or not, but he must admit without hesitation that it would be possible for him. He judges, therefore, that he can do something because he is aware that he ought to do it and cognizes freedom within him, which, without the moral law, would have remained unknown to him.

The underlying idea is that the agent realizes that he ought, morally, not to give false testimony in this case. Since, according to Kant, moral obligation requires autonomy of the will, the agent's awareness that he *ought* not to give false testimony allows him to

judge that he does indeed possess autonomy of the will, and thus transcendental freedom.³² Kant sums up this strategy by stating that our awareness of the moral law functions as a fact of reason (*CPrR* 5:31):

Consciousness of this fundamental law may be called a fact of reason because one cannot reason it out from antecedent data of reason, for example, from consciousness of freedom (since this is not antecedently given to us) and because it instead forces itself upon us of itself as a synthetic a priori proposition.

The Kantian constructivist answer to the Normative Question resembles Kant's argument from the fact of reason because, instead of asserting directly that we possess autonomy of the will, it appeals to the point of view of agents who already have moral beliefs and are motivated to act on them to elicit the Kantian constructivist characterization of the form of our will. The Kantian constructivist answer to the Normative Question is therefore driven by a "practical insight" on the part of the agent who raises the question, to borrow a term from Rainer Forst (2011: 37). Appealing to the Kantian analysis of obligation, Kantian constructivism asserts that this practical insight involves a concept of obligation which presupposes that the agent is morally obligated only if she possesses autonomy of the will.

Forst suggests that this practical insight marks our *identification* with the conception of practical reason as autonomous willing put forward by Kantian constructivism (2011: 37):

Only this practical insight leads to a moral self-understanding and a *moral identity* that is developed in cognitive, volitional, and affective ways, or rather, to a

³² Strictly speaking, Kant here infers transcendental freedom *directly* from the agent's judgment that he morally ought to refuse the demand, without invoking autonomy (cf. Kleingeld 2010: 68-69). However, this inference is not valid without the premise, which Kant takes himself to have established two paragraphs before, that the moral law is the law that the will is to itself and that, *therefore*, action on that requires transcendental freedom (*CPrR* 5:28-29). Without this premise, it is not clear what the agent's judgment that he morally ought to refuse the demand has to do with the practical use of *pure* reason (see Chapter 5, Section 2.1, below). Indeed, in Pauline Kleingeld's reconstruction of Kant's argument at 5:30, the crucial premise is that the agent's deliberation "presupposes a normative principle, and that this normative principle is conceived as independent from the inclinations" (2010: 68). This is just another way of saying that the principle in question is not heteronomous but autonomous.

moral character. In this way we can see why talk of a categorically valid morality and of “unconditioned” duties is not empirically unreachable.

Similarly, Rawls contends that the Kantian constructivist claim that acceptance of the Categorical Imperative is the formal principle of our will finds support in a conception of the person which, in turn, “is *elicited* from our moral experience and reflection” (2000: 240, italics added).

Accordingly, the Kantian constructivist approach to the Normative Question not only depicts the normativity of morality as having its metaphysical source in the form of our will. It also cites our moral beliefs and motivations, in conjunction with the Kantian analysis of obligation, in support of the claim that our will exhibits the form that provides the source of moral normativity. It thus mirrors Kant’s remark that, while freedom is the “*ratio essendi*” of the moral law, our awareness of being bound by the moral law is the “*ratio cognoscendi*” of freedom (*CPrR* 5:4, footnote).

Now, it is important to note that this answer is not supposed to provide a *foundationalist* vindication of the rational requirement to act as we believe we morally ought to act. That is, the practical insight or fact of reason that drives the Kantian constructivist answer does not serve the role of an axiom from which this rational requirement can be inferred. If this was the role assigned to the practical insight, then either this practical insight would amount to a self-evident rational intuition or it would give rise to the Normative Question all over again. In the former case, the Kantian constructivist answer to the Normative Question would be structurally analogous to the source externalist answer. In the latter case, the Kantian constructivist answer would lead into an infinite regress.

Instead, the practical insight serves the role of *eliciting* the characterization of our will as autonomous which *then* explains why we are rationally required to do what we believe we morally ought to do. We might say, then, that the practical insight *supplements* the answer to the Normative Question which is provided by the Kantian constructivist claim that the form of our will is the source of moral normativity. Importantly, it can serve this role only because it corresponds to the moral beliefs and motivations contained within the point of view from which the Normative Question arises in the first place.

To appreciate this point, recall our dialectical situation. The Normative Question asks, from the point of view of an agent who already accepts that she is rationally required to do what she believes she morally ought to do, how the rational authority that she thereby attributes to morality can be vindicated. Kantian constructivism responds that

the agent is rationally required to do what she believes she morally ought to do because morality's normativity has its source in the form of her practical standpoint. As we saw in the previous subsection, the agent may now raise the question of whether her practical standpoint really has the form that Kantian constructivism ascribes to it. In response, the Kantian constructivist argues that the relevant formal characterization of the agent's practical standpoint is elicited by the agent's own moral beliefs and motivations, without which the Normative Question would not have arisen in the first place. This move is legitimate because, in order to answer the Normative Question, we need not provide an *ultimate* vindication of morality's rational authority that could convince any agent, independently of her prior attitudes to morality. Instead, our goal is to explain the rational authority of morality *to* an agent "who is already inside the moral life", as Stern puts it (2010: 8).

The Kantian constructivist answer to the Normative Question qualifies as an instance of what Rawls calls *philosophy as defence* (2000: 321). Philosophy as defence is philosophy employed in the service of the "practical faith" of ordinary agents (*ibid.*). That is, it serves to sustain faith in the tenability of our moral ideals. As the label suggests, philosophy as defence does not carry out this task by providing *positive* arguments in favour of these ideals, and thus, as it were, launching an offensive. Instead, it proceeds from within the practical standpoint of ordinary agents, showing these ideals to be at least defensible. Insofar as there is nothing that conclusively undermines these ideals, this strategy justifies us in continuing to uphold them, at least for practical purposes.

Rawls traces the notion of philosophy as defence to the religious apology of Leibniz (Rawls 2000: 106-107). He describes Leibniz's strategy as follows (Rawls 2000: 107):

Faith is defended by maintaining that, from the point of view of faith, the objections raised against it fail to show it to be either unreasonable or incoherent. To affirm the faith, one need not prove its beliefs. Rather, it suffices to rebut objections, and for this it suffices to state certain possibilities showing that the objections may be false. This establishes that the objections are not conclusive, and so faith stands.

As an instance of philosophy as defence, then, the Kantian constructivist answer to the Normative Question does not aim to provide a positive case for the rational authority of moral motivation. Rather, it points out that the practical standpoint of agents for whom the Normative Question arises contains a practical insight which, as shown by the Kantian analysis of obligation, elicits a formal characterisation of that practical standpoint which in turn explains why it is rationally required to act on one's moral motivation. It

employs a *non-foundationalist* strategy insofar as it does not attempt to provide a vindication of morality's rational authority from an Archimedean point, but reassures us that our faith in this authority is tenable because it forms part of a self-standing vision, or 'story', that links our faculty of practical reason to moral normativity in a way that explains why it is rationally required to do what one believes to be one's obligation (cf. Bagnoli 2014: 318, Forst 2011: 32, Neiman 1994: 135-139, Rawls 2000: 268).

Importantly, it follows that the Kantian constructivist answer to the Normative Question is *fallible*. While this answer reassures the agent who raises the Normative Question that there is a story by which she can explain the rational authority of morality to herself, it does not guarantee that this story is true. After all, although our practical standpoint contains a 'practical insight' which coheres with the notion that we possess autonomous wills and are therefore rationally required to act on our moral beliefs, this practical insight might be based on an illusion, autonomy might be a mere 'phantasm'.³³

There are at least two possible scenarios in which this would be the case. First, while I am not committed to the claim that autonomy requires transcendental freedom, I have not rejected it either. If this claim is true, and if we do not possess transcendental freedom, then practical reason might be a receptive faculty after all. Second, even if autonomy does not require transcendental freedom, we might be deceiving ourselves about our own deliberations, imagining that we can act on the motive of respect for the law, while in fact we invariably act on our strongest antecedent desire. After all, that autonomy does not require a metaphysically demanding kind of free will does not mean that it cannot be undermined by unconscious psychological mechanisms.

4.4 Avoiding the Shmagency Objection

I have argued that the Kantian constructivist answer to the Normative Question is to be understood as an exercise in philosophy as defence. I now want to suggest, as such, it turns out not to be vulnerable to the Shmagency Objection.

³³ By recognising this possibility, I depart from Kant. In the *Critique of Practical Reason*, he states that the fact-of-reason argument is sufficient to establish that we possess pure practical reason (*CPrR* 5:3, 15-16, also see note 22 in the previous chapter). Accordingly, Kant thinks that the (failed) project of a critical vindication of pure practical reason, pursued in the third section of the *Groundwork*, is no longer needed. Unlike the Kant of the *Critique*, I think that a critical vindication of pure practical reason is not superfluous. Yet, unlike the Kant of the *Groundwork*, I think that the best we can do is to provide a non-foundationalist vindication, as laid out in this chapter.

To see this, we need to consider how the Shmagency Objection represents the dialectic of the Normative Question. About Korsgaard's constitutivism, Enoch writes (2006: 179):

[C]onsider Korsgaard's hope of grounding a reply to the skeptic in what is constitutive of action. We are to imagine, then, someone who remains indifferent when we tell him that his actions are immoral or irrational. He then reads Korsgaard and is convinced that self-constitution is a constitutive aim of action [...]. Do we have any reason to believe that now he will care about the immorality or irrationality of his actions?

The Shmagency Objection, then, assumes that the Normative Question is raised from the point of view of a sceptic who is *indifferent* to the moral or rational status of her actions. In order to convince such an agent that she should act in a certain way, one would have to produce an argument that begins from normatively neutral foundations. Indeed, this is how Enoch interprets the constitutivist strategy: it purports to infer normative requirements from a merely descriptive account of the metaphysics of some activity, such as agency or practical reasoning (Enoch 2006: 172). However, argues Enoch, even if the sceptic agrees that, to count as performing a certain activity, she is required to act in a certain way, she can ask why she should be performing that activity in the first place. The *same* scepticism that made her question the importance of the moral or rational status of her actions will make her question the importance of counting as performing some activity. In Enoch's words (2006: 179):

Has Korsgaard put us in a better spot vis-a-vis this why-be-an-agent (rather than a shmagent) problem than we were vis-a-vis the why-be-moral or why-be-rational challenges with which we—or at least Korsgaard—started?

However, my remarks in the previous subsection revealed that the Kantian constructivist answer to the Normative Question, as I conceive of it, does *not* address a sceptic who is *indifferent* about the moral or rational status of her actions. Instead, it aims to reassure an agent who already accords rational authority to morality that this authority can be vindicated. In order to provide this reassurance, it points out that the agent's own moral beliefs and motivation elicit a characterisation of her own will as autonomous, which in turn explains why she is rationally required to do what she takes to be her moral obligation.

Enoch might object that, even if this *does* convince the agent that she is rationally required to act as she believes she morally ought to, she *should* not be convinced.

Whatever made the agent raise the Normative Question in the first place should also make her doubt that she is rationally required to act on her moral beliefs just because otherwise she would not count as being engaged in ‘autonomous willing’ (cf. Enoch 2006: 180). After all, why is she required to engage in that activity in the first place?

Now, not only does this objection falsely presuppose that the Normative Question asks for an explanation of the rational requirement to act on one’s moral beliefs from the point of view of someone who is indifferent towards morality, it also presupposes that it is possible to give such an explanation. However, Kantian constructivism holds that morality is normative only for agents who possess autonomous wills. And the practical standpoint of an autonomous will *presupposes* (at least implicitly) the normativity of the moral law. After all, autonomy is the will’s characteristic of being a *law* to itself, which requires the ability to act from mere respect for the lawful form (see Chapter 1, Section 4.2 above). Accordingly, an agent who is truly indifferent towards morality does not possess an autonomous will and, therefore, is not bound by moral obligations. She does not possess pure practical reason and, accordingly, cannot be criticised as irrational when she fails to engage in autonomous willing and violates the constitutive principle of that activity, the moral law.

5. Clarificatory Remarks

I argued that Kantian constructivism, unlike source externalism and non-Kantian versions of source internalism and voluntarism, provides an answer to the Normative Question. Most of my arguments so far have been concerned with discussing potential answers to the Normative Question, rather than with relating the discussion to other questions in practical philosophy or responding to potential objections. I hope to make up for this now with a few clarificatory remarks.

5.1 *Philosophy as Defence and Kantian Constructivism*

I have noted that understanding the Kantian constructivist answer to the Normative Question as an exercise in philosophy of defence helps us to see how it overcomes two challenges: the challenge of showing that we possess autonomous wills and the challenge posed by the Shmagency Objection. Now, one might wonder if source externalism—and non-Kantian versions of source internalism and voluntarism, for that matter³⁴—could

³⁴ For simplicity, I will restrict myself to the comparison between Kantian constructivism and source externalism. However, everything I say about the latter equally applies to non-Kantian

not equally avail themselves of the strategy of philosophy as defence. If they could, then Kantian constructivism would not seem to have any advantage over these views with respect to the Normative Question. After all, as I explained in Section 4.2, the Shmagency Objection resembles the objection that I raised against source externalism in Section 2. Thus, if Kantian constructivism can avail itself of philosophy as defence to avoid the Shmagency Objection, then perhaps source externalism can avail itself of the same strategy to avoid my objection from Section 2.

Is there something special about Kantian constructivism, then, such that only this view can avail itself of the strategy of philosophy as defence in order to provide an answer to the Normative Question? The answer is yes. What makes Kantian constructivism so special is that it puts forward a *formal* rather than a substantive characterisation of the principle of reason that it takes to vindicate the rational authority of moral normativity. It explains this authority by showing how moral normativity is linked to the very constitution of our practical reason. Whereas source externalism attempts to explain the link between moral normativity and our faculty of practical reason by invoking the *content* of our beliefs about the former, Kantian constructivism regards the former as having its source in the very *form* of the latter. It is important to recall that Kantian constructivism's claim about the source of moral normativity already fully explains why it is rationally required to act on one's moral beliefs. The contribution made by the strategy of philosophy of defence is to *relate* this explanation to the practical standpoint of the agent who raises the Normative Question.

One way of bringing out the difference between source externalism and Kantian constructivism is to say that the latter offers an analysis of obligation according to which our moral obligations come with their own *rationale*, so to speak, as to how they relate to our practical reason. That is, according to Kantian constructivism, our awareness of being morally obligated does not just play the role of a *substantive* input to the effect that we are obligated to perform or omit an action; rather, it offers a 'practical insight' into the *relation* between moral normativity and the form of practical reason.

Another way of bringing out the difference is to note that, in one sense, the practical insight by which the Kantian constructivist answer to the Normative Question elicits its formal characterisation of practical reason need not be *veridical*. In particular, the practical insight in question could involve an agent's belief that she is obligated to φ , where φ -ing is not in fact morally required of her, and still facilitate Kantian

versions of source internalism and voluntarism (and, in fact, any view that conceives of practical reason as a receptive faculty).

constructivism's answer to the Normative Question. As long as the agent conceptualises obligations in the way supposed by the Kantian analysis of obligation, her belief will elicit the formal characterisation of practical reason that yields a critical vindication of morality's rational authority. One would hope, of course, that the agent will recognise that her belief is false (because it is incompatible with the Categorical Imperative). But failure to recognise this would not render her belief *non-moral*. After all, unlike source externalism, Kantian constructivism does not consider a belief's moral character to depend upon its tracking the correct substantive considerations (which are in turn determined by stance-independent normative properties). Rather, what makes a belief moral is its employment of a concept of obligation that depicts obligation as a matter of form rather than substance.³⁵

Source externalism cannot avail itself of the strategy of philosophy as defence in the same way as Kantian constructivism because it does not link moral normativity to a formal characterisation of practical reason. Source externalists could at most offer a coherentist argument to the effect that their claim that it *just is* rationally required to act as one believes one morally ought to act stands in a relation of mutual support with the moral beliefs and motivations of the agent who raises the Normative Question. But this does not constitute an answer to the Normative Question. After all, it only *affirms*, and to some extent supports, the claim that we are rationally required to act according to our moral beliefs. But it does not offer any *explanation* of why this is so. After all, the source externalist account still regards moral normativity as delivering substantive inputs to practical reasoning, and therefore as lying beyond the scrutiny of reason (see Section 2 above).

³⁵ This also suggests a resolution of the *Euthyphro*-style dilemma that Kantian constructivism supposedly faces (Shafer-Landau 2003: 42-43, cf. *Euthyphro* 10a). The supposed challenge is that if the formal constraints on willing do not correspond to an independent moral reality, "it is hard to see why the outcomes of the [Categorical Imperative] should be definitive of morality" (Shafer-Landau 2003: 43; for a related objection, see Darwall et al. 1993: 143). On the other hand, if these constraints do correspond to an independent moral reality, then the form of our will is not the source of moral normativity. Note that Kant faced a similar objection in his own lifetime (Pistorius 2007: 27). The resolution is that the formal constraints on willing are authoritative for morality because the concept of moral obligation is the concept of a requirement that binds us by the mere form of our will. Indeed, according to Kantian constructivism, without the formal constraints on willing, there would be no criterion to tell the moral from the non-moral (Bagnoli 2016: 1234, cf. *CPrR* 5:60).

5.2 Non-moral Normativity

Given the way I have interpreted the Normative Question, there is no obvious reason why this question should not arise for non-moral parts of the normative domain, such as epistemic or prudential normativity. Indeed, we would expect non-moral normativity to be linked to our rational faculties just as closely as moral normativity, and the idea of a critique of reason should motivate us to ask for an explanation of that link in the non-moral as well as in the moral case. However, Kantian constructivism is a view about the source and content of moral normativity or, more specifically, moral obligation. Accordingly, it only provides an answer to the Normative Question for that part of the normative domain. This might strike some as disappointing. One might have expected a view which answers the Normative Question for one part of the normative domain to answer it for all other parts as well. Unfortunately, I have no argument to remedy this disappointment completely. Nevertheless, I would like to make a few remarks to at least alleviate the worry.

The first thing to note is that, while one could raise this point as an objection to the Kantian constructivist answer to the Normative Question, one could also draw the conclusion that the normative domain is not as uniform as one might have expected. In particular, one might conclude that there is an answer to the Normative Question for the phenomenon of moral obligation which does not apply to other normative phenomena. While this leaves open the possibility that there is *another* answer to the Normative Question that applies to *all* parts of the normative domain, it assigns a special status to moral obligation.³⁶ For those drawn to the Kantian analysis of obligation, this conclusion should not come as a surprise. After all, that analysis implied that obligations, unlike certain other phenomena such as reasons that merely recommend a certain course of action, bind the will by virtue of their *form*. Accordingly, they allow for a critical vindication of their rational authority in a way in which other normative phenomena do not.

The second thing to note is that the Normative Question seems more pressing in the area of moral obligation than in other areas. In particular, if morality could not be rationally vindicated, this might have more revisionist implications for our practices than a similar lack of rational vindication would in the case of prudential or epistemic normativity. After all, it seems that the latter parts of the normative domain are more closely aligned with our immediate self-interest. It is a commonplace that violating our moral

³⁶ Darwall reaches a similar conclusion with regard to the “hard” problem of moral normativity (2016).

obligations, in many cases, is not immediately disadvantageous for us along non-moral dimensions. Violating epistemic or prudential norms, by contrast, is almost invariably disadvantageous along several non-epistemic and non-prudential dimensions. Admittedly, this remark is vague and speculative. Unfortunately, however, I cannot provide a more detailed examination of this issue within the limits of this thesis.

Finally, while I have located the Kantian analysis of obligation in the moral domain, some might argue that the same notion of obligation finds application in other normative domains as well. This is highly speculative, and I am not aware of any convincing attempt to extend the Kantian analysis of obligation beyond the moral domain.³⁷ But it is worth noting that if such an attempt was to succeed, the Kantian constructivist answer to the Normative Question could be extended to cover non-moral normative phenomena as well.

5.3 *Humean Constructivism*

In my discussion of the answers that different views of the sources of moral normativity give to the Normative Question, I discussed Sharon Street's view as a version of source internalism and implicitly characterised it as attempting to explain both the normativity of morality and the rational requirement to do what we believe we ought to do by appeal to agents' *substantive* motivational states (see Section 3 above). One might object that I have thereby neglected the formal element in her theory which, according to Street, makes it a Humean version of *constructivism* (2010, 2012).

According to Street's Humean constructivism, understood as a view about the sources of normativity, all normativity has its source "within the practical point of view, where the practical point of view is given a formal characterization" (2012: 40). By the practical point of view, Street means "the standpoint occupied by any creature who is in the state of mind of 'valuing'" (ibid.). She adds that to give this point of view a formal characterisation is to give "an account of the attitude of valuing that does not itself presuppose any substantive values but rather merely explicates what is involved in valuing anything at all" (ibid.). Street argues that the "logical and instrumental requirements" that govern practical reason are normative because they are *constitutive* of the attitude of valuing (ibid.) According to her constructivism, the normativity of specific *substantive* principles or facts is explained by their "following" from such logical and instrumental requirements (ibid.).

³⁷ Perhaps Korsgaard's account could be interpreted as such an attempt (e.g., 2008b, 2009a).

Street distinguishes her Humean constructivism from Kantian constructivism by its treatment of morality. Unlike Kantian constructivism, Humean constructivism does not hold that the mere form of the practical standpoint, by itself, entails any substantive principles. Accordingly, whether morality, considered as a set of substantive requirements, is normative for an agent depends on “her particular, contingently given, evaluative starting points” (2012: 41). All that the form of the practical standpoint requires is that certain *relations* (of consistency, etc.) hold between the various contents of our practical point of view, i.e., the various things we ‘value’.

This latter aspect of Humean constructivism is the reason why I did not discuss it as an approach to the Normative Question—at least not as one that is distinct from source internalism. Humean constructivism cannot be characterised as locating the source of *moral* normativity, *qua* moral normativity, in the form of the practical point of view. It considers this form to be the source of the normativity of certain substantive requirements. But the moral nature of these requirements is completely irrelevant as far as the formal characterisation of the practical point of view is concerned. Morality is just one among a range of concerns that might be part of our ‘evaluative starting point’.

Therefore, when it comes to explaining why we are rationally required to do what we believe we are morally required to do, Humean constructivism does not offer an explanation in terms of formal principles of practical reason but must revert to the *substantive inputs* which, once they are ‘plugged into’ these formal principles, yield substantive requirements. As far as the Normative Question is concerned, then, Humean constructivism can be considered a source internalism of the non-Kantian variety discussed in Section 3.

5.4 *The Normative Question and Epistemology*

In Chapter 1, I left open the question of whether Kantian constructivism takes a stance on the epistemology of moral normativity (see Chapter 1, Section 3.3). Since the Normative Question, and the Kantian constructivist answer to that question, might seem to involve epistemological issues, some clarification on this issue might be in order.

The Kantian constructivist answer to the Normative Question reassures the agent that she is rationally required to do what she believes she morally ought to do. Part of this answer involved an appeal to the agent’s moral consciousness (or ‘practical insight’) to support the notion that her practical reason has the form that Kantian constructivism takes to be the source of moral normativity. This might seem to suggest that Kantian constructivism illuminates how we can have access to moral truths. However, we should be careful not to overestimate Kantian constructivism’s merits in this respect.

The first thing I wish to point out is the fallibility of the Kantian constructivist answer to the Normative Question. While this answer gives us some reassurance that we are indeed rationally required to act on our moral motivation, it does so by providing a non-foundationalist argument to the effect that faith in the rational authority of morality is *defensible* from within the point of view of the agent who raises the Normative Question. This does not *prove* the rational authority of morality. For example, as I pointed out in Section 4.3, it could turn out, either on metaphysical or on psychological grounds, that we do not possess autonomous wills.

Second, even though the Kantian constructivist strategy reassures us, to some extent, *that* we are rationally required to do what we believe we morally ought to do, it does not illuminate how we might have *access* to this truth—or, for that matter, *any* normative truths. Recall that a will's possessing autonomy presupposes its being bound by the form of law (see Chapter 1, Section 4). In other words, it logically (but not metaphysically) *presupposes* the normativity of the Categorical Imperative. In a specific sense, then, belief in our autonomy is itself a normative belief. Accordingly, the perennial question of how we have access to normative truths re-emerges as the question of how we have access to the truth that we possess autonomous wills (cf. Hussain and Shah 2006: 284-285).

Third, and connectedly, although Kantian constructivism's view of the source of moral normativity is intimately connected with a view of the content of our moral obligations—namely, that this content is given through an application of the Categorical Imperative to our maxims—it does not explain *how* we can know this content. After all, to say *that* we are morally required to φ is not to say how it is *possible* that we know that we are morally required to φ .

6. Conclusion

In this chapter, I argued that Kantian constructivism, unlike other views of the source of moral normativity, provides an answer to the Normative Question. This is because Kantian constructivism identifies the form of our practical reason as the source of moral normativity. Kantian constructivism's source claim serves as an explanation of the rational requirement to act as we believe we morally ought to act because it finds support within the practical standpoint of the agent for whom the Normative Question arises. The moral beliefs and motivations of the agent, *via* the Kantian analysis of obligation, support the claim that the agent possesses autonomy of the will. By employing this strategy—the strategy of philosophy as defence—the Kantian constructivist answer to the Normative Question avoids Enoch's Shmagency Objection.

Now, like the argument for Kantian constructivism outlined in Chapter 1, this answer relies on the Kantian analysis of obligation, which is supposedly elicited from within the practical standpoint of ordinary agents. As I noted in Chapter 1, the starting point of this analysis is the notion that an agent is obligated to φ only if she can φ from the motive of duty. The case for Kantian constructivism therefore crucially depends on whether the practical standpoint of ordinary agents can be shown to involve this rather demanding concept of obligation.

In the remainder of the thesis, I discuss two rival accounts of the practical standpoint that are put forward by theorists sympathetic to Kantian constructivism. Korsgaard's *first-personal account* identifies the practical standpoint that motivates the Kantian analysis of obligation with the standpoint that we all must take up, *qua* agents. By contrast, the *second-personal account*, put forward by Darwall, holds that the Kantian analysis of obligation must take off from within the second-person standpoint, viz. "the perspective you and I take up when we make and acknowledge claims on one another's conduct and will" (Darwall 2006: 3). While the second-person standpoint is also a first-personal standpoint—it is an agent's perspective on her second-personal relations—Darwall argues that it is not *merely* first-personal because it involves second-personal concepts which cannot be derived from the constraints of mere agency (Darwall 2006: 10). I shall side with Darwall by arguing that the first-personal account fails to motivate the Kantian analysis of obligation whereas the second-personal account succeeds.

PART II: THE PRACTICAL STANDPOINT: FIRST OR SECOND-PERSONAL?

Chapter 4: Why the First-Personal Account Fails

1. Introduction

Korsgaard argues that insofar as we are agents *at all*, our practical standpoint already involves the demanding concept of obligation that kicks off the Kantian analysis of obligation—and thereby implicitly affirms the Kantian notions of autonomy and the Categorical Imperative. More specifically, she claims that the constitutive aim of action or agency is to constitute ourselves as agents, as the authors of our own actions. In order to do so, we must choose our actions in light of a conception of our practical identity, a normative conception of who we are. But if we are to succeed at constituting ourselves as the authors of our actions, Korsgaard contends, our adoption of a conception of practical identity must be guided by the Categorical Imperative.

In this chapter, I argue that Korsgaard's first-personal account fails to motivate the Kantian analysis of obligation. Section 2 outlines Korsgaard's account. Section 3 explains why the first-personal account fails. My main argument is that the first-personal type of reasoning that all agents as such must engage in does not establish the link between obligation and autonomy that is needed to get the Kantian analysis of obligation off the ground. As we shall see, the necessary presuppositions of this type of reasoning are compatible with its being determined by the independent value or desirability of outcomes or states of the world.

2. The First-Personal Account

Korsgaard holds that the first-person standpoint of practical deliberation that we all take up *qua* agents implicitly involves the concept of obligation that gets the Kantian analysis off the ground. Arguments for this claim can be found both in her 2009 book *Self-Constitution: Agency, Identity, and Integrity* (henceforth '*Self-Constitution*') and in her earlier *The Sources of Normativity* (henceforth '*Sources*'), from 1996. Although the contents of the two works are largely compatible with one another, *Self-Constitution* seems to have a slightly different focus and, in some places, to provide an updated version of arguments found in *Sources* (cf. Korsgaard 2009: 20-26). In this section, I will try to provide an outline of what I take to be Korsgaard's mature view by combining insights from both books while following *Self-Constitution* wherever the two depart from one another.

2.1 Practical Identities and the Form of Law

Korsgaard's main line of argument starts from the observation that there is an important asymmetry between the way we might conceive of action from the third-person standpoint of an observer and the way we must think about what we're doing when *we* act, from the first-person standpoint. From the third-person standpoint, it might seem as if an agent always acts on her strongest motive, and that this is all there is to agency. From the first-person standpoint of the agent, however, it seems "as if there is something over and above" her motives, namely *the agent*, who must choose which motive to act on (1996: 100).

According to Korsgaard, this is because our consciousness has a "reflective structure" which puts us at a certain critical *distance* from our motives (1996: 100). We cannot think of ourselves as merely the sum of our motives, and thereby absolve ourselves from the need to deliberate about what to do. In Korsgaard's words, "human beings are condemned to choice and action" (2009: 1). Accordingly, *choosing* which of our motives to act on is "our *plight*: the simple inexorable fact of the human condition" (2009: 2).

However, according to Korsgaard, not only does the reflective structure of our consciousness put us at a certain critical distance from our motives, it also forces us to take *control* of our choices and actions (2009: 1). After all, as a result of our reflective consciousness, we cannot possibly conceive of our choices as determined by something outside our own practical deliberation. In other words, we cannot approach deliberation thinking that our choices have already been made for us. Korsgaard here deliberately echoes Kant's remark that we must act "under the idea of freedom" insofar as "one cannot possibly think of a reason that would self-consciously receive guidance from any other quarter with regard to its judgment" (*G* 4:448, Korsgaard 1996: 94).

Since we cannot think of our actions as being determined by anything other than our own choice, we must conceive of ourselves as the sole "authors" of our actions (Korsgaard 2009: 20). Therefore, the only way to choose a course of action, and thus to choose acting on the corresponding motive, is to *make this motive our own*. In other words, we need to *identify* with the motive and the corresponding action. It follows that, in order to act, we must choose a conception of *ourselves*. More specifically, we must choose what Korsgaard calls a conception of our "practical identity" and defines, roughly, as "a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking" (Korsgaard 1996: 101). For example, I might conceive of myself under the practical identity of a philosopher. Under this conception of myself, I will regard some courses of action (e.g., that of revising

a chapter of my thesis) as worth choosing and others (e.g., that of playing video games) as less so.

Korsgaard argues that practical identities are constituted by a set of principles or rules (1996: 101-104, 2009: 21). That is, a conception of your practical identity is associated with a set of principles such that, if you do not follow these principles (at least to a certain extent), that conception ceases aptly to describe your actions. For example, if I do not follow the principles that come with the practical identity of a philosopher to some extent at least (say, if I never actually do any philosophical work), I fail to be a philosopher. In this way, according to Korsgaard, our practical identities generate reasons and obligations (1996: 101-102). They make certain demands on us because if we fail to comply with certain principles, we fail to qualify as bearers of these identities. They are the *sources* of our reasons and obligations (Korsgaard 1996: 103-104).

Now, if the first-person standpoint is to elicit the Kantian analysis of obligation, it must somehow, implicitly or explicitly, involve the notion that an agent is obligated to φ only if her judgment *that* she is obligated to φ can motivate her to φ (see Chapter 1 above). In other words, we are obligated to φ only if we can φ from duty. Is this the notion of obligation that Korsgaard takes to be implicit in our relation to our practical identities? Korsgaard seems to think that it is; for she goes on to link our relation to our practical identities to the notion of *autonomy*.

To illustrate the link between obligation and identity, Korsgaard invites us to imagine that we have two ‘selves’, a “thinking self” and an “acting self”, where the former urges the latter to conform to our conception of practical identity (1996: 104). If the acting self fails to do so, then the thinking self can punish it through emotions like guilt, regret, repentance, and remorse. Thus, there is a relation of authority between the two selves: “the acting self concedes to the thinking self its right to government” (ibid.). Since the thinking self can thus be seen as a *legislator* and the conceptions of practical identity it lays down can be seen as *laws*, Korsgaard argues that by choosing and acting on our conception of practical identity we are a law to ourselves, i.e., we exercise autonomy of the will (ibid.)

But now we may wonder whether the notion of autonomy that Korsgaard is appealing to here is the Kantian notion of autonomy in the strict sense of “the characteristic of the will by which it is a law to itself (independently of the characteristic of the objects of willing)” (G 4:440). After all, as we saw in Chapter 1, this notion of autonomy amounts to the ability to shape one’s actions by “[n]othing other than *the representation of the law* in itself” (G 4:401). And one might suppose that when I choose and act in accordance

with my practical identity as a philosopher, my motive is that I *want* to be a philosopher rather than ‘the mere representation of the law in itself’.

However, Korsgaard notes that the picture of action that we have seen her draw so far is not yet complete (1996: 107). The problem is that everything that we have said so far is compatible with all our practical identities being *contingent*. I happen to have chosen to conceive of myself as a philosopher, but I could equally have chosen a different identity. This raises a question: why should I take seriously the obligations that come with any of my practical identities? Indeed, it seems that this question is prompted by the very reflective structure of our consciousness that gave rise to our need for a conception of practical identity in the first place: just as I cannot conceive of my choices as determined by my motives and instead must *choose* which motive to act on, I cannot conceive of my choices as determined by a conception of practical identity that I happen to have adopted and instead must *choose* a conception of practical identity myself. Korsgaard’s thought seems to be that I cannot conceive of myself as the mere *locus* of choices which ultimately determined by contingent factors such as my motives or practical identities. *I* am not really acting if my choices are determined by contingencies which are not, ultimately, under my control. But this train of thought threatens to lead us into an infinite regress. After all, the conception of practical identity which determines our choice of motive must in turn be chosen in light of a (higher-order) conception of practical identity, and so on (1996: 120).

We seem to have a dilemma, then: on the one hand, we must act on a conception of practical identity lest we yield control over our actions to whatever happens to be our strongest motive. That is, we need a *principled* way of choosing which of our motives to act on. On the other hand, we cannot act on a conception of practical identity that is itself contingently chosen; while this would amount to choosing our actions in a principled manner, it would also amount to yielding *control* over our principle of choice to contingent factors that lie outside our practical deliberation.

According to Korsgaard, this means that we must choose our conception of practical identity in line with the Categorical Imperative test (1996: 97-98, 107-108). After all, what we are looking for is a principle that can govern our choice of actions, but which is not itself adopted on the basis of some contingent motive, for then our choice would be determined by something outside our practical deliberation. All that remains, as the basis of our choice of a principle, is that it *can* serve as a principle that governs our choice, and thus independently of its contents. We might say, then, that it must be chosen by virtue of its form rather than its substance. More specifically, in our choice of practical identity, we must be guided by the mere *form of law*. Substituting ‘conception of

practical identity' with 'maxim' and associating the form of law with the possibility of being willed as a universal law (see Chapter 1, Section 4.1), it follows that we must follow the Formula of Universal Law: "act only according to that maxim through which you can at the same time will that it become a universal law" (*G* 4:421, italics deleted).

Korsgaard takes this to imply that there is one practical identity that is not contingent (1996: 120-123). The same problem that gives rise to our need to adopt *any* conception of practical identity in the first place—that is, the fact that we must act 'under the idea of freedom'—also renders it necessary for us to adopt the *specific* practical identity of *someone who faces this problem*. Accordingly, the validity of our contingent practical identities, viz., their ability to solve our problem of having to choose, *depends* on our adoption of this fundamental identity as a someone who deliberates about what to do. Since *qua* bearers of this fundamental identity we must follow the Categorical Imperative, our contingent practical identities are therefore valid only if they are compatible with the Categorical Imperative. Now, since adopting and acting on a specific maxim is required by the Categorical Imperative only if we can act on it from the motive of duty, it follows that our contingent practical identities serve as sources of genuinely normative obligations only if they can be chosen and acted on from duty (see Chapter 1, Section 4.1). Indeed, Korsgaard seems to think that to ϕ *because* it is required by one's practical identity is an *instance* of acting from duty (2009: 22). This, then, is how Korsgaard takes the Kantian concept of obligation to be entailed by the necessary presuppositions of the first-person standpoint of practical deliberation that we inevitably take up insofar as we are agents.

2.2 Interlude: Korsgaard on the Formula of Humanity

We have already seen that, according to Korsgaard, the reflective structure of our consciousness gives rise to the need for a conception of practical identity, which in turn gives rise to the need to act in accordance with the Categorical Imperative. Unlike Kant, however, Korsgaard draws a distinction between the Categorical Imperative, which she identifies with the Formula of Universal Law, and "the moral law" (1996: 99, 2009: 80). While she believes that the Formula of Universal Law is the law of a will that must act 'under the idea of freedom', she denies that it suffices to yield the full range of moral obligations we have. Her rationale for this divergence from Kant is that, while the Formula of Universal Law requires that we act only on maxims that can be willed *universally*, it does not specify the *domain* for which our maxims are to be universalizable. Specifically, as Korsgaard reads it, the Formula of Universal Law demands that we only act on maxims that can equally be adopted by all other *beings like us*. But different

classes of beings are like us under different descriptions. And, by itself, the Formula of Universal Law does not tell us which of these descriptions the relevant one is.

As I argued in Chapter 1, the Formula of Universal Laws implies that our maxims to be universalizable in the sense that we must be able to will them to hold for all *autonomous wills*. Thus, I do not think that we need an argument for the moral law in addition to one for the Formula of Universal Law (*pace* Korsgaard). I only mention Korsgaard's separate argument for the Formula of Humanity for the sake of providing a complete overview of her position. Accordingly, this sub-section does not play a crucial role in the overall argument of this thesis.

Korsgaard suggests that in order to establish the moral law from the starting point that our consciousness as agents has a reflective structure, we need to show that we must adopt the practical identity of citizens in a kingdom of ends (1996: 100). Thus, her aim is to establish that, *qua* agents, we are committed to Kant's Formula of the Kingdom of Ends (*G* 4: 439, cf. Korsgaard 2009: 80):

[A]ct in accordance with the maxims of a member giving universal laws for a merely possible kingdom of ends.

By the "kingdom of ends", Kant means a "systematic union of all rational beings" through a common law (*G* 4:433). And that common law is the so-called Formula of Humanity (*G* 4:429, italics deleted):

So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means.

Therefore, Korsgaard's goal of showing that all agents must view themselves as citizens in a Kingdom of Ends is tantamount to establishing that all agents must determine their wills by the Formula of Humanity.

As we have seen, Korsgaard argues that the one practical identity that we *must* adopt insofar as we are agents is that of a being who needs to adopt a practical identity in order to act. She takes this to imply that we must regard our identity as "a human and rational being" as normative (2009: 24). After all, in her framework, "a human being is an animal who needs a practical conception of her own identity, a conception of who she is which is normative for her" (1996: 123). It is thus by virtue of our *humanity*—in this sense of the term—that we need practical identities in the first place. Accordingly, by adopting practical identities, and regarding them as normative, we implicitly accord normativity to our humanity also. Korsgaard writes (2009: 24):

I must take *some* ways of identifying myself seriously, or I won't have any reasons at all. Insofar as I take that fact—the fact that I need some way of identifying myself—to be a reason, I express the value I set upon myself as a human and rational being. So 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. For by doing that we are endorsing a reason that arises from our rational nature—namely, our need to have reasons.

Korsgaard notes that one could argue that valuing one's own humanity does not yet commit one to valuing the humanity of *others* and that, hence, endorsing one's identity as a human being is not yet to determine oneself by the Formula of Humanity (1996: 131-132). In reply, Korsgaard argues that reasons and obligations are not private but public (1996: 136-145, cf. 2009: Chapter 9). That is, if something is normative for one agent, it is normative for *any* agent. In this sense, all reasons and obligations are *shareable*.

Korsgaard purports to establish this by an argument that she takes to be analogous to Wittgenstein's argument against the possibility of a private language. She reconstructs Wittgenstein as arguing that the meaning of a term is normative, in the sense that one's use of the term can be correct or incorrect (1996: 136-137, cf. Wittgenstein 2009: §256-258). Now, if there was a private language, i.e., a language that only a single speaker understands, then the aptness of this speaker's use of any given term of the language would simply be determined by whether or not she feels like using it. But then there is no possibility for this speaker to go wrong. According to Korsgaard's reconstruction, this is what leads Wittgenstein to conclude that language must be public. Korsgaard argues that, if it is the normativity of language that warrants the rejection of the possibility of a private language, then the normativity of reasons and obligations must likewise warrant the rejection of their privacy. If they were private, then an agent would be normatively warranted to act in a certain way if and only if she felt like acting in this way. Thus, there would be no possibility of acting wrongly. Therefore, reasons cannot be private mental entities; they must be shareable.³⁸ According to Korsgaard, just as it is almost impossible

³⁸ Note that Korsgaard here makes a rather questionable inference. That is, from the notion that normativity cannot have its source in an individual's mind she infers that all reasons must be shareable, in the sense that they are either reasons for everyone or no reasons at all. But it is not obvious that this follows: even if it is true that the notion of a normative reason is incoherent without the companion notion of a community that *fixes* its normativity, it can still be the case that some reasons are normatively *binding* only on some agents.

to hear words in a language we understand as mere noise, it is almost impossible to regard the reasons of another person as mere non-normative force (1996: 138-145).

Korsgaard goes so far as to claim that reasons and obligations are *relations* between reasoners. Even in the case of an agent obligating herself, her obligation consists in the relation between her “thinking self” and her “acting self”, where the former has legislative authority over the latter (1996: 138, also see Section 2.1 above and Chapter 7, Section 6 below). Since Korsgaard takes this to mean that reasons are always reasons for all agents, she concludes that the fact that agents must value their own humanity means that all agents must acknowledge each other’s humanity as valuable. Evidently, Korsgaard takes this to establish that all agents must follow the Formula of Humanity.³⁹

2.3 Action as Self-Constitution

We have seen how Korsgaard attempts to derive the Categorical Imperative, as well as what she calls ‘the moral law’, from the necessary presuppositions of the first-personal standpoint that we inevitably take up insofar as we are agents. In *Self-Constitution*, she points out that the same argument can also be formulated in terms of the metaphysical constitution of *action*. Again, Korsgaard starts from the observation that, when we act, we view ourselves as something over and above our motives, someone who *chooses* which of our motives to act on (2009: 75). According to Korsgaard, this means that the reflective structure of our consciousness *divides us into parts* (2009: 213). Specifically, it separates us from our motives in such a way that, if we give in to our strongest motive on every occasion, our behaviour will not be attributable to *us*, as *agents*, and therefore will not qualify as *action*. When we act, we are trying to *reunite* ourselves by integrating our motives into a coherent conception of ourselves which we thereby construct. On Korsgaard’s account, therefore, action has the constitutive aim of *self-constitution*.

In this version of Korsgaard’s argument, the Categorical Imperative is identified as the constitutive principle of self-constitution, and thus of action. The idea is that if we choose to act on a given motive—and adopt a corresponding practical identity—without following the Categorical Imperative, we will constitute ourselves *badly* (2009: 75-80).

³⁹ Several authors have remarked that it is not clear how the notion that we must regard all human beings, and all the projects that they adopt as part of their conceptions of practical identity, as equally *valuable* supports the Kantian requirement to respect the dignity of persons as a *constraint* on our actions as opposed to a consequentialist ethics of protecting and promoting human beings and their projects as much as possible (e.g., Cumminskey 2008, Pauer-Studer 2010: 301, Sensen 2011: 137).

This is because by selecting practical identities that cannot be willed as universal laws, we act on motives that we happen to have, as a result of our contingent psychological make-up. Instead of taking up *authorship* of our actions by backing them up with a conception of *ourselves* as a unified, self-determining being, we act on a conception of practical identity that only represents a *part* of us. In other words, instead of controlling our actions through an exercise of our faculty for practical deliberation alone, we let ourselves be determined by forces that are contingent from the point of view of our identity as a human, rational being.

Korsgaard labels a will that lets itself be *wholly* determined by contingent motives in this way a “particularistic will”. She writes (2009: 76):

A truly particularistic will must embrace the incentive in its full particularity [...]. But this means that particularistic willing eradicates the distinction between a person and the incentives on which he acts. And then there is nothing left here that is the *person*, the agent, that is his self-determined will as distinct from the play of incentives within him. If you have a particularistic will, you are not one person, but a series, a *mere heap*, of unrelated impulses.

Korsgaard concludes from this that a truly particularistic will is *impossible*. After all, if the distinction between the person and her incentives is collapsed, there is no *agent* to whom the resulting movements could be attributed. Hence, these movements cannot be interpreted as *actions*. As human beings, however, we *must* act. Accordingly, we must always implicitly be *trying* to follow the Categorical Imperative in our practical deliberations.

This concludes my exposition of Korsgaard’s argument for the claim that the Categorical Imperative, and the associated notions of autonomy and the motive of duty, are entailed by the necessary presuppositions of the first-person standpoint of agency. In the next section, I shall explain why this account fails.

3. Against the First-Personal Account

We have seen how Korsgaard purports to show that the first-personal perspective of practical deliberation that we inescapably take up *qua* agents elicits the starting point for the Kantian analysis of obligation. In her view, our need to be guided by a conception of our practical identity entails an implicit acknowledgement of the Kantian notions of the motive of duty, autonomy, and the Categorical Imperative. However, as I now want to show, Korsgaard’s first-personal account fails. I want to make three claims in particular. First, deliberation from the first-person standpoint of agency does not necessarily

presuppose that our reasons and obligations have their source in our conception of practical identity. Accordingly, the first-person standpoint does not involve an implicit acceptance of the Categorical Imperative. Second, and connectedly, Korsgaard's argument equivocates two different senses of 'practical reason' and related terms. While it is true that, in *one* sense of the term, practical reason is impossible unless we possess autonomy of the will, this is not the sense of the term in which it is inescapable to reason practically. Third, even if the Categorical Imperative were the constitutive principle of the first-person standpoint, the inescapability of that standpoint would not suffice to show that all agents implicitly acknowledge that principle as binding. This can only be established by appeal to a 'practical insight', on the part of agents, into the obligating force of the law.

3.1 *The Naïve First-Personal Reasoner*

In rejecting Korsgaard's claim that the first-person standpoint necessarily involves an implicit acceptance of the Categorical Imperative, I largely follow an argument by Darwall (2006: Chapter 9). He begins by reminding us that autonomy, which is a necessary condition of being capable of acting from duty, and thus of being bound by the Categorical Imperative, is "the characteristic of the will by which it is a law to itself (independently of any characteristic of the objects of willing)" (G 4:440). This means that we are autonomous if and only if we are capable of being guided by "[n]othing other than *the representation of the law in itself*" (G 4:401). Darwall points out that, therefore, autonomy amounts to the capacity to act on a principle which "is not itself 'empirical' (based on a prior desire) or derived from the value (e.g., the desirability) of an independent outcome or state" (2006: 220. cf. G 4:21). After all, through autonomy, pure reason itself is supposed to be practical.

Accordingly, if the first-person standpoint of agency necessarily presupposed that we possess autonomous wills (and hence that we are bound by the Categorical Imperative), then it would rule out as unintelligible, not only actions performed from a desire for some outcome or state, but also actions performed from "any putative evaluation of some outcome or state" such as the belief that some state of the world is intrinsically valuable (Darwall 2006: 221). The first-person standpoint would render these actions *unintelligible* insofar as one could not be guided by such considerations and still consider oneself the *agent* from whose *authorship* they spring. After all, in Korsgaard's view, our actions cannot fully be imputed to *us*, and can therefore not fully be regarded as *actions*,

if the reasoning leading up to them is guided by anything other than our conception of practical identity.

However, Darwall objects, it seems perfectly intelligible, from within the first-person standpoint, for one's deliberations to be guided by the putative value or desirability of outcomes or states of the world. To illustrate this, he invokes what he calls the "*naïve (first-personal) practical reasoner*" (2006: 216-217, cf. FitzPatrick 2005: 680):

For such an agent, desires (and related mental states like pleasure and enjoyment) seem responsive to the value or desirability of outcomes or possible states of affairs. The agent reasons from premises drawn from the content of her beliefs and desires, not from anything about their form.

The naïve practical reasoner, then, takes her desires to track the value or desirability of their objects. If a naïve practical reasoner desires to watch a film, for example, she will regard the state of affairs in which she watches a film to be desirable or of value. It is important to note, however, that the practical reasoner does not *infer* this evaluation from the premise *that* she has this desire. Instead, her desires are "backgrounded" in her deliberation (Darwall 2006: 216, cf. Pettit and Smith 1990). That is, she is guided by the *content* of her desires, i.e., the way they present the world to her, rather than by her *having* these states. That is, the naïve practical reasoner is guided by the putative value or desirability of watching a film—as presented to her by her desires—rather than by the realisation that she has such a desire and that, therefore, watching a film must be desirable or of value. In short, the naïve practical reasoner reasons "from the perspective of [her] desires" (2006: 217).

Clearly, the naïve practical perspective does not presuppose that the agent possesses autonomy of the will. After all, the naïve practical reasoner's will is far from being "a law to itself (independently of any characteristic of the objects of willing)" (*G* 4:440). She acts in accordance with what she takes to be the desirability or value of states of the world. Accordingly, nothing in her first-person practical standpoint requires her to adopt a purely formal principle of reason such as the Categorical Imperative.⁴⁰

⁴⁰ In fact, things might be even worse for Korsgaard's argument, for it is questionable whether we must invariably suppose that the actions we choose are objectively good or supported by reasons at all (Timmermann 2006a: 79, Wallace 2004: 459). It is not obvious that we must even take up the naïve practical perspective, and that we might not instead engage in a mode of reasoning that regards itself as purely determined by antecedent motivational states.

Although Darwall does not say so explicitly, his remarks about the naïve practical perspective amount to a rejection of Korsgaard's claim that, *qua* agents, we possess a reflective consciousness which separates us from our motives in such a way that we must choose which motive to act on in light of our conception of practical identity. The naïve practical reasoner does not regard her practical identity as the source of the normativity of her reasons and obligations (cf. Darwall 2006: 230-231). Instead, she regards certain objects as possessing value or desirability independently of her conception of herself. The only role assigned to her agency is that of responding to the independent value and desirability of these objects by promoting or sustaining them.

Note that a naïve practical reasoner might still question whether the actions that her desires suggest to her really serve to promote these valuable objects. Or, again, she might try to gain a better grasp of the features of these objects that account for their value (Darwall 2006: 228). But these kinds of critical reflection are merely epistemic in nature as they are concerned with the reasoner's ability to *discern* the value of the consequences of actions; they do not question whether the considerations in virtue of which these consequences are supposed to be valuable are normative in the first place. That is, in a naïve practical perspective, the reasons for which one takes something to be a reason are not the object of critical reflection.

Note also that all that is required for Darwall's central claim to be true is that, from within the first-person standpoint, we need not conceive of our actions as guided by a conception of our practical identity *all the way down*. That is, it could even be true that, *qua* agents, we cannot conceive of our particular choice of action as determined by our strongest motives, and that we must instead regard it as an expression of our conception of practical identity. It would still not follow that the first-person standpoint necessarily involves the Kantian notions of autonomy and the Categorical Imperative. After all, Korsgaard's argument for that claim requires that even our conception of practical identity itself is chosen in light of some higher order practical identity of ours, and so on, such that, ultimately, we must choose in light of our most fundamental, non-contingent identity as a human being. Therefore, Darwall need not even rely on the intelligibility of the naïve practical perspective as outlined above. Instead, it suffices that it be intelligible for our practical reasoning to be 'naïve' at the level of choosing our conception of practical identity. That is, all that needs to be intelligible for Korsgaard's argument to fail is that the sequence of practical identities bottoms out with a practical identity that is chosen, not by a formal principle of practical reason, but by the perceived desirability or value of some outcome or state of affairs (cf. Fitzpatrick 2005: 666-668).

3.2 Agency and the Critique of Reason

One might object that what I have just argued contradicts some of my remarks in Chapters 2 and 3. More specifically, one might argue that by invoking the idea of a critique of reason in my motivation for the Normative Question I have also provided support for Korsgaard's claim that, from within the first-person standpoint, we are positioned at a critical distance from our motives. After all, I argued that the idea that no standard possesses the authority of reason unless it withstands the critical scrutiny of reason is incompatible with conceptions that depict reason as a *receptive* faculty. Accordingly, a choice that is determined by factors that lie outside the scrutiny of reason will not count as a reasoned one. But then it seems to follow that Korsgaard is right: we cannot view our choice as determined by factors, such as our antecedent motives, that lie outside our own practical deliberation. If our choice is to be *rational*, it must be guided by a conception of *ourselves*, a representation of the principle by which *we choose* how to act. And this conception of ourselves must in turn be chosen in a rational way, lest it become a factor that lies outside our practical deliberation and thereby undermine the rational credentials of our choice. Ultimately, it seems, our choice must be guided by a formal principle of practical reason which expresses the constitution of the will as such.

The problem with this line of argument is that it presupposes that a critical vindication of reason is *possible*. It is true that the idea of a critique of reason on which I relied in Chapters 2 and 3 suggests that it is constitutive of a standard's possessing rational authority, i.e., of its being a verdict of *reason*, that it withstand the scrutiny of reason. And this means that it must be generated by the very form of reason. Applying this to the practical domain, it follows that if the form of practical reason as such does not generate any practical principles, then rational choice is impossible. But this line of argument warrants the conclusion that we are obligated by the Categorical Imperative only if we *assume* that rational choice is possible.

Korsgaard would reply that we *must* assume that a critical vindication of reason is possible because “[w]e must act, and we need reasons in order to act” (2009: 24). In other words, the need to act is *inescapable*, it is “our plight” (2009: 2). But if acting is inescapable, and acting requires a formal principle of practical reason, then there *must* be a formal principle of practical reason. That this is Korsgaard's view is perhaps clearest from the following two passages (1996: 123-124 and 125, respectively):

Rational action exists, so we know it is possible. How is it possible? [R]ational action is possible only if human beings find their own humanity to be valuable.

But rational action is possible, and we are the human beings in question. Therefore we find ourselves to be valuable.

If you think reasons and values are unreal, go and make a choice, and you will change your mind.

Korsgaard claims that the possibility of rational action can be inferred from its *existence*. Therefore, according to Korsgaard, the success of a critique of reason is guaranteed by the existence of rational action. When asked how we can be so sure that rational action *does* exist, Korsgaard responds that it must exist insofar as we act, and acting is inescapable.

Now, if Korsgaard's argument is sound, then our need to act ascertains our autonomy. However, as we saw in Chapter 3, although autonomy is necessary for the success of a critical vindication of practical reason, it might nevertheless be a mere "phantasm", a "chimerical idea without truth" (*G* 4:445). In other words, although the idea of a critique of reason entails that practical reason is impossible unless there is a formal principle of practical reason, it might yet turn out that our will is not autonomous and that, therefore, there is no formal principle of practical reason. Thus, the idea of a critique of reason, as such, is compatible with the result that practical reason is impossible.

But where, then, did Korsgaard's argument go wrong? After all, is it not true that action presupposes the possibility of practical reason and that acting is inescapable? I submit that the flaw with this argument is that it equivocates different senses of 'practical reason', 'action', and related terms. It is true that the idea of a critique of reason entails that nothing can be truly rationally justified unless the principles of reason themselves can withstand the scrutiny of reason. And this, in turn, implies that practical reason is not possible unless there is a formal principle of practical reason. And it is also true that action is inescapable, and that in order to act we need to reason practically. However, this latter sense of 'practical reason' is not the sense in which, according to the idea of a critique of reason, practical reason is impossible without autonomy. This is what the argument from the naïve practical perspective shows. While it is true that, in order to act, we must take ourselves to be responsive to *normative considerations* of some sort, it does not follow that, when we act, we are inescapably guided by a conception of our practical identity which can be willed as a *universal law* (cf. Fitzpatrick 2005: 666-668). Instead, we might view ourselves as responding to the value or desirability of states of the world, as these are presented to us by our desires. Hence, it is not true that, unless we

are guided by the notions of autonomy and the Categorical Imperative, we are not acting—at least in the sense of ‘acting’ in which it is inescapable for us to act.⁴¹

This, then, is how I can sustain the notion of a critique of reason without supporting Korsgaard’s claim that the necessary presuppositions of the first-person standpoint elicit the Kantian analysis of obligation. While it is true that practical reason, in the sense employed by the idea of a critique of reason, is impossible if we do not possess autonomous wills, it is not true that we *must* regard ourselves as autonomous insofar as we are agents. For the sense of ‘practical reason’ employed in the idea of a critique of reason is not the *minimal* sense of the term which is employed in the (true) statement that it is inescapable for us, *qua* agents, to use practical reason. Connectedly, the support that the idea of a critique of reason lends to the Normative Question does not stem from the claim that action and practical reason are *inescapable*. Instead, it relies entirely on the *attractivity* of the more demanding notion of reason that it appeals to.

This more demanding notion of reason roughly corresponds to what Rawls labels “the Reasonable”, distinguishing it from “the Rational” (1980: 528). By “the Rational”, Rawls means someone’s capacity “to secure the advancement of their conception of the good” (*ibid.*). We might say that even Darwall’s naïve practical reasoner possesses this agential capacity. While such an agent does not necessarily act in a self-interested manner (i.e., she does not act with the sole aim of fulfilling her desires *because* they are her desires), she still acts in accordance with her desires insofar as these desires present certain objects as valuable or desirable for their own sake. What the naïve practical reasoner lacks, however, is the Reasonable, which corresponds to Kant’s notions of autonomy and pure practical reason (Rawls 1980: 532).

Indeed, Kant himself distinguishes between “pure practical reason”, i.e., a practical reason that is capable of being a law to itself by its mere form, and empirical practical reason or “empirically conditioned reason”, i.e., a practical use of reason that is not determined by reason itself, but by aspects of the agent’s empirical psychology (*CPrR* 5:16, cf. Timmermann 2006a: 80-85). Connectedly, Kant sometimes distinguishes

⁴¹ It is worth noting that a similar criticism of Korsgaard’s approach was suggested by Reath (2006b: note 51). Reath argues that Korsgaard equivocates ‘law’ in her derivation of the Categorical Imperative by inferring the need for a universal and necessary practical principle from the need to act on a more or less general principle (cf. Fitzpatrick 2013: 57). Moreover, he contends that Korsgaard makes no real use of Kant’s notion of autonomy and instead settles for a less ambitious conception of freedom. Darwall seems to raise what is essentially the same objection by invoking the naïve practical perspective.

between the “predisposition to humanity”, which renders us “a rational being”, and the “predisposition to personality”, which is “our receptivity for the moral law, as an incentive, sufficient by itself” (*Rel* 6:26-28). Through these distinctions, Kant acknowledges that, in principle, there could be a being that is rational in the sense that it is capable of something akin to means-ends reasoning while lacking pure practical reason, i.e., the capacity to shape actions by the mere representation of law as such.

3.3 *The Fact of Reason and the Inescapability of Agency*

I have argued that the first-person standpoint that we all take up *qua* agents does not elicit the Kantian analysis of obligation because we can deliberate practically, and thus be agents, without employing the concepts of pure practical reason (i.e. autonomy), the motive of duty, and the Categorical Imperative. Put another way, acceptance of the Categorical Imperative is not constitutive of the first-person standpoint that is inescapable for us insofar as we are agents.

I now want to argue that Korsgaard’s emphasis on the *inescapability* of the first-person standpoint of agency is ill-advised in the first place. Even if the Categorical Imperative was the constitutive principle of a practical standpoint that is inescapable for us insofar as we are agents, it still would not follow that we implicitly acknowledge that we are obligated to follow the Categorical Imperative whenever we deliberate about what to do. In general, from the fact that a practical standpoint is inescapable for all agents we cannot infer that all agents implicitly acknowledge the *normativity* of the constitutive principle or aim of that standpoint.

The problem is related to Enoch’s Shmagency Objection, which was discussed in the previous chapter (Sections 4.2 and 4.4). Recall that, according to the Shmagency Objection, Korsgaard’s claim that the aim of self-constitution is constitutive of agency does not suffice to establish that there is a reason to engage in self-constitution. Enoch’s argument is that, unless we already have a reason to be agents (rather than shmagents), the constitutive aim of agency does not give us any reasons. Now, one might expect Korsgaard’s claim that agency is *inescapable* to be a promising reply to the Shmagency Objection. After all, if being shmagents is not an option for us, then perhaps there is a reason to pursue the constitutive aim of agency even if we do not have a prior reason to be agents. However, Enoch remarks (2006: 188):

Think again about finding yourself playing a game of chess and assume for now that for some reason you cannot quit—not that you should not quit but that you cannot quit. And assume that sacrificing a pawn is the thing you have most chess-

related reason to do (it best promotes your chances of checkmating your opponent or some such). Well, do you have a reason to sacrifice a pawn? Not, it seems to me, if you don't have a normative reason to play or win the game, and this even if you can't quit. For you can continue playing or 'going through the motions,' grudgingly, refusing to internalize the aims of the game.

The idea is that, even if the activity of playing chess is inescapable, you do not have a reason to perform the actions that are most conducive to the constitutive aim of playing chess (checkmating your opponent) unless you have a prior reason to play chess. Therefore, even if it is not an option to play 'shmess' rather than chess (because playing chess is inescapable), you can still play chess "half-heartedly", without acknowledging the constitutive aim of playing chess as providing you with reasons (2006: 187).

Enoch persuasively argues that the reason why the inescapability reply to the Shmagency Objection fails is that the necessity involved in the inescapability of playing chess is of the wrong *kind* to render it *normatively* necessary to pursue the constitutive aim of playing chess (2006: 187, 189, cf. FitzPatrick 2005: 673, Parfit 2006: 372-376, Wallace 2004: 456, 2012: 37). As Enoch notes, Korsgaard does not fully specify the nature of the necessity involved in the inescapability of agency (Enoch 2006: note 42). While she distinguishes the necessity in question from causal, logical, and rational necessity, her positive characterisation of it is exhausted by the remark that it is "a necessity you are *faced* with" and "our *plight*: the simple inexorable fact of the human condition" (Korsgaard 2009: 1-2). I hope it is clear by now that the realisation that I am 'faced' with the necessity of acting, by itself, does not entail that I must consider myself bound by the constitutive principle of agency.

We can see, then, that emphasising the inescapability of a practical standpoint—that is, *any* practical standpoint—is of no help in establishing that an agent implicitly considers herself bound by the constitutive principle of that standpoint. This reinforces a point I already made in Chapter 3: in order to elicit the Kantian analysis of obligation, it does not suffice to show that autonomy and related Kantian notions are necessary for the possibility of practical reason, considered as a descriptive 'folk-theoretical' notion. Instead, we have to appeal to a 'practical insight', akin to Kant's fact of reason, by which agents acknowledge that they are obligated in the sense of obligation that kicks off the Kantian analysis. In other words, we need to show that ordinary agents conceptualise the obligations that they *already* take themselves to have in accordance with the Kantian analysis. It needs to be established, not that all agents are engaged—by some non-normative kind of necessity—in an activity constituted by acceptance of the Categorical

Imperative, but that ordinary agents' deliberative field is already *normatively* structured in a way that presupposes the Categorical Imperative as the supreme principle of obligation. We need to show that the practical standpoint of ordinary agents is “already colored—so to speak—by the moral ought” (Sensen 2011: 116).

Now, Korsgaard could in principle decide to abandon the ambition of eliciting the Kantian analysis of obligation by appealing to the inescapability of the first-person standpoint, and instead argue that all agents, by virtue of deliberating from within the first-person standpoint, have a practical insight of the sort just described. However, her account would still fail. After all, as the argument from the naïve practical reasoner demonstrates, it is not true that the kind of practical deliberation that agents, *qua* agents, must minimally engage in elicits the conceptual framework of the Kantian analysis of obligation. I lay out a more promising approach in the next chapter.

4. Conclusion

In this chapter, I argued that the first-personal account according to which the necessary presuppositions of the first-person standpoint that all agents as such take up elicits the Kantian analysis of obligation fails. In a nutshell, the problem is that merely first-personal reasoning is too minimal to involve the ambitious Kantian notions of the motive of duty, autonomy, and the Categorical Imperative. We have also seen that *any* argument that attempts to elicit the Kantian analysis of obligation by appealing to the inescapability of some type of reasoning must fail. We cannot vindicate the Kantian framework by appeal to the claim that accepting the Categorical Imperative is associated with some feature that all agents necessarily have, where the necessity in question is non-normative. Instead, we must appeal to a ‘practical insight’ on the part of agents, i.e., some *normative* notion that they are already committed to, in order to get the Kantian analysis of obligation off the ground.

In the next chapter, I argue that the second-personal account successfully appeals to such a practical insight. It does so by highlighting that, insofar as we participate in the practice of second-personal address, i.e., the person-to-person address of purportedly legitimate claims and demands, we already view ourselves as obligated in the sense of obligation that the Kantian analysis takes as its starting point. I also rebut several objections to the effect that the success of the second-personal account saves the first-personal account, either because the former is indistinguishable from the latter, or because first-personal reasoning necessitates second-personal reasoning.

Chapter 5: The Second-Personal Vindication of the Kantian Analysis of Obligation

1. Introduction

In order to get the Kantian analysis of obligation off the ground, we need to appeal to a ‘practical insight’ on the part of agents which catapults them into a mode of reasoning involving the concept of obligation that forms the starting point of this analysis. As we saw in the previous chapter, this mode of reasoning must go beyond the merely first-personal mode that is inescapable for all agents. The reason is that the interrelated notions of autonomy, the Categorical Imperative, and the motive of duty are not entailed by the practical concepts that we must inescapably employ insofar as we are agents.

In this chapter, I follow Darwall in arguing that the mode of reasoning that Kantian constructivists should appeal to in order to elicit the starting point of the Kantian analysis of obligation is that which takes place within the second-person standpoint, viz., “the perspective you and I take up when we make and acknowledge claims on one another’s conduct and will” (Darwall 2006: 3). Accordingly, the fact of reason should be construed as an encounter with demands addressed person-to-person that prompts us to reason in ways that implicitly involve the Categorical Imperative and presuppose our autonomy.

Several theorists contend that, even if second-personal concepts are required to get the Kantian analysis of obligation off the ground, it does not follow that the first-personal account fails. These theorists argue either that the second-person standpoint cannot be distinguished, conceptually, from the first-person standpoint, or that the two standpoints, while conceptually distinct, are always taken up at the same time. I shall discuss the arguments for these claims in detail and reject all of them. I conclude that the second-person standpoint is not only necessary to motivate the Kantian analysis of obligation, but also conceptually and extensionally distinct from the standpoint of agency as such.

The argument of this chapter unfolds as follows. Section 2 introduces the second-personal account and argues that it offers an interpretation of the fact-of-reason argument that vindicates the Kantian analysis of obligation. Section 3 rebuts three arguments to the effect that the success of the second-personal account saves the first-personal account because the first- and second-person standpoint cannot be taken up separately from one another.

2. The Second-Personal Account

In this section, I show that in order to elicit the Kantian analysis, we should appeal to the second-personal nature of obligation. In other words, in order to show that agents have a fact-of-reason-like ‘practical insight’ involving the Kantian notions of autonomy, the Categorical Imperative, and the motive of duty, we should appeal to the second-personal dimension of our practical deliberation, i.e., its sensitivity to demands addressed person-to-person.

I start by outlining Darwall’s claims that, first, the concept of obligation is intimately linked to the notion of accountability and, second, the fact-of-reason argument will not be successful if it fails to emphasise this fact. Then, I explain how this notion of accountability elicits the Kantian analysis of obligation. Subsequently, I argue that the same notion is also linked to some of the more patient-centred aspects of Kantian ethics (such as the dignity of persons) and thereby illuminates the relationship between the different variations of the Categorical Imperative. Finally, I show how the second-personal account helps us to provide a successful version of the fact-of-reason argument.

2.1 *Obligation and Accountability*

As we saw the previous chapter, Darwall agrees that the first-person standpoint of agency does not necessarily involve autonomy and the Categorical Imperative and that, instead, we need to appeal to a more direct encounter with the moral law, something akin to Kant’s fact of reason. Darwall argues that, in order to show that the fact of reason prompts us to reason in a way that presupposes autonomy of the will, we need to emphasise the second-personal character of obligation (2006: 239-243).

Recall that Kant illustrates his fact-of-reason argument with the example of an agent whose prince demands that he “give false testimony against an honorable man”, on pain of losing his life (*CPrR* 5:30). Kant argues that, while the agent might not be confident that he *will* refuse his prince’s demand, he must nevertheless judge that he *can* refuse it. And this, according to Kant, is because the agent is aware that he is *obligated* not to give false testimony against an honorable man. According to Kant, this shows that from our awareness of being bound by the moral law, the “fact of reason”, we can infer that we possess autonomous wills (*CPrR* 5:31). Now, Darwall objects that, as it stands, it is not clear that the fact of reason establishes autonomy of the will (2006: 235-239). After all, almost anyone will agree that, if we are obligated to φ , then we are able to φ , given that ‘ought’ implies ‘can’. From the proposition that we are *able to* φ , however, it does not follow that we can determine ourselves to φ by “[n]othing other than *the*

representation of the law in itself” (G 4:401). In other words, it is not clear that the fact of reason shows that we can ϕ from the motive of duty—which is what it has to show if it is to establish that we possess autonomous wills.

According to Darwall, the fact of reason can serve the purpose of eliciting the Kantian analysis of obligation only if it exploits obligation’s conceptual link to *accountability*. He argues that, “[w]hen we are morally obligated, we are not morally free to act otherwise; members of the moral community have the authority to hold us responsible if we do” (2006: 27). This is not to say that, as a conceptual matter, failure to fulfil one’s obligations warrants a physical sanction or penalty (Darwall 2006: 250, 2017: 268). Instead, obligation implicates a distinct kind of *authority*, the authority to hold the agent accountable by *demanding* that she fulfil her obligations. An agent’s failure to fulfil her obligation without excuse renders her *blameworthy* in the sense that she is legitimately held accountable, by the victim of her wrongdoing or by representative members of the moral community, at least through the adoption of Strawsonian reactive attitudes such as indignation and resentment (Darwall 2006: Chapters 4 and 5, cf. Strawson 1968). According to Darwall, the adoption of such Strawsonian reactive attitudes amounts to an implicit *address* of a purportedly legitimate demand vis-à-vis the agent because these attitudes come with an “implicit RSVP”; that is, they call on the agent to acknowledge and act in accordance with the demand in question (2006: 145).

Importantly, Darwall argues that the authority implicated by moral obligation is an irreducibly second-personal, practical authority (2006: 11-15). That is, it is not exhausted by the standing to demand compliance with standards of conduct that are valid independently of this standing. Instead, it includes the authority “to create a distinctive reason for compliance” that would not exist *but for* this authority (2006: 11). Darwall contends that any account that attempts to ground moral obligations in something other than this irreducibly second-personal authority will face a *wrong-kind-of-reasons problem* (2006: 14-17). This is because nothing but the irreducibly second-personal authority to demand that an agent ϕ will warrant holding the agent accountable for failing to ϕ . Therefore, given the link between obligations and accountability, nothing but irreducibly second-personal considerations will provide the right sort of rationale for taking an agent to be under an obligation to ϕ .

Darwall illustrates this claim with an example (2006: 5-10). Suppose that Arnold is standing on top of Bella’s foot. Darwall notes that Arnold can acknowledge several types of practical reasons or requirements to withdraw his foot. First, Arnold could recognize that by standing on Bella’s foot he is making Bella worse-off than she would otherwise be, and thus bringing about a sub-optimal state of the world. Since this reason

depends solely on facts about states of the world, and Arnold's ability to manipulate them, Darwall labels it a "state-of-the-world-regarding reason" (2006: 6). Note that Arnold's state-of-the-world-regarding reason could equally serve as a reason for any other agent who is able to withdraw Arnold's foot, and is thus an *agent-neutral* reason. Indeed, state-of-the-world-regarding reasoning need not even favour withdrawing Arnold's foot since another course of action might bring about an even more valuable state of the world. Hence, Arnold's state-of-the-world-regarding reason is independent of his *accountability* to the moral community. Accordingly, the only authority that we can invoke by 'addressing' this reason with Arnold has a purely *third-personal*, epistemic character. It entails the standing to *advise* or *counsel* Arnold on facts about states of the world and his ability to affect them, but it does not provide the conceptual resources to warrant holding him to account if he fails to withdraw his foot.

This is not to say that Arnold cannot possibly be held accountable for failing to act in accordance with his state-of-the-world regarding reasons. Darwall's point is that state-of-the-world-regarding reasons, by themselves, cannot *warrant* holding Arnold accountable. To establish that Arnold is accountable for withdrawing his foot from the premise that he has state-of-the-world-regarding reason to do so, one would have to add an additional premise, e.g., that Arnold is accountable for doing whatever he has state-of-the-world-regarding reason to do. And, importantly, this additional premise cannot itself be based on state-of-the-world-regarding reasoning. For the value of the state of the world that is brought about by holding Arnold accountable is a reason of the wrong kind to hold him accountable. After all, an agent's accountability for ϕ -ing conceptually entails that she would be blameworthy if she failed to ϕ , lacking adequate excuse. And an agent's blameworthiness for failing to ϕ does not depend on the value of the outcome of blaming her (Darwall 2006: Chapter 4, Strawson 1968).

We have seen, then, that state-of-the-world-regarding reasons cannot capture the phenomenon of moral obligation. After all, if they are the wrong kind of reason to hold Arnold accountable, then, in light of the conceptual link between obligation and accountability, these reasons cannot ground moral obligations.

What is even more important for our purposes is that, if we read Kant's fact-of-reason argument as depending on the premise that the agent has a state-of-the-world-regarding reason not to give false testimony against his fellow citizen, then this argument will not be successful in establishing that the agent possesses autonomy of the will. To see this, consider that even the naïve practical reasoner, whom we encountered in the previous chapter, could recognize such a reason. In the example employed in Kant's fact-of-reason argument, the naïve practical reasoner's desires might present a state of the

world in which the accused goes free, or a state of the world in which the act of giving false testimony against an innocent person is not performed, as more desirable or valuable than the alternatives. Accordingly, the naïve practical reasoner might acknowledge a state-of-the-world-regarding reason not to give false testimony against the innocent person. However, as we have already seen, the perspective of the naïve practical reasoner does not involve the Kantian notions of the motive of duty, autonomy, and the Categorical Imperative. Accordingly, state-of-the-world-regarding reasoning does not vindicate the Kantian analysis of obligation.

Now, Arnold might remove his foot on the basis of a second type of normative consideration. In particular, he might recognize that he is under a *requirement* to withdraw his foot. Such a requirement might have its normative source in irreducibly normative properties. Note that, since such a requirement functions as a *constraint* on Arnold's behavior and binds *him*, in the first person, to remove his foot, it is an *agent-relative* consideration. As such, it might appear to be a more promising candidate for capturing the phenomenon of obligation than state-of-the-world-regarding reasons. Nevertheless, the only authority that we might invoke in addressing such a requirement with Arnold is the first-personal authority that *his* requirements have over *him*. We would thus merely *remind* him of a standard that he already (at least implicitly) considers himself to be subject to, like a requirement of logic. Consequently, Arnold might recognize this requirement, and recognize the corresponding authority, without thereby recognizing his *accountability* in the sense that is involved in the notion of obligation.

Requirements of this form are therefore just as inadequate to capture the sense in which it is obligatory for Arnold to remove his foot as state-of-the-world-regarding reasons. Again, there is no conceptual impossibility in holding Arnold accountable for complying with such a requirement. But his accountability cannot be *warranted* by such a requirement. Instead, it must be based upon Arnold's blameworthiness should he fail to withdraw his foot, lacking adequate excuse. And this blameworthiness is independent of any (merely first-personal) requirement to blame Arnold.

But not only is this type of practical requirement unsuitable to ground moral obligation, it also cannot figure in the fact-of-reason argument if this argument is to elicit the Kantian analysis of obligation. After all, the only agential capacity that can be inferred, on the basis that 'ought' implies 'can', from the proposition that we are under a requirement (of this sort) to ϕ is the capacity to ϕ . It does not follow that we are capable of ϕ -ing *from duty*, and thus possess autonomy of the will. Nothing in the fact-of-reason argument, so construed, rules out that the agent acts from the recognition that he *just* is required not to give false testimony, where the normative force of the requirement in

question has its source in the irreducibly normative properties of the act of giving false testimony. Indeed, Darwall points out that this type of first-personal requirement is consistent with the (decidedly source externalist) metaethical picture of a Rossian deontologist, which explicitly rejects the Kantian analysis of obligation (Darwall 2006: 239, Ross 2002: 5-6).

However, there is a third type of normative consideration on the basis of which Arnold might withdraw his foot. Namely, Arnold might recognize Bella's authority to *demand* that he withdraw his foot. Arnold would thereby recognize a consideration that is *irreducibly* second-personal because it is based upon the very authority that Bella invokes in demanding Arnold's compliance. Darwall labels this consideration a "second-personal reason" because its "validity depends on presupposed authority and accountability relations between persons and, therefore, on the possibility of the reason's being addressed person-to-person" (2006: 8). Unlike the state-of-the-world-regarding reason, this second-personal reason is *agent-relative* in that it will not serve as a reason for anyone but Arnold. Connectedly, it will not serve as a reason for Arnold *not* to remove his foot.

Arnold's second-personal reason is the right kind of reason to warrant blaming Arnold if he fails to remove his foot without adequate excuse; for its existence is conceptually linked to Arnold's accountability. Accordingly, this reason captures the sense in which Arnold is *obligated* to withdraw his foot. The second-personal account, according to which obligation is conceptually linked to an irreducibly second-personal authority, thus avoids the wrong-kind-of-reasons problem that haunts non-second-personal accounts of obligation. But second-personal reasons do not only help us to capture the phenomenon of obligation. As I am about to argue, they also illuminate the fact-of-reason argument and vindicate its ability to establish the agent's autonomy of the will.

2.2 Accountability and Autonomy

In order to show that autonomy of the will can be established if the fact-of-reason argument is supplemented with the notion of a second-personal reason, Darwall notes that "second-personal address", i.e., the practice of giving one another second-personal reasons by making demands on one another, has certain "normative felicity conditions" (2006: e.g., 265-266). That is, any instance of second-personal address, if it is to be intelligible, presupposes that certain conditions are satisfied. In case these conditions are not satisfied, the address "misfires", and thus fails to create a second-personal reason (Darwall 2006: 44).

Importantly, Darwall thinks that second-personal address presupposes that the addressee can hold herself accountable for acting as the addresser demands. To see this, recall that, by addressing Arnold second-personally, Bella is implicitly holding him accountable. But holding someone accountable differs in important ways from coercing or otherwise “goad” him (Darwall 2006: 49-52). In goading an agent, we are attempting to get him to act in a certain way by announcing that we will act in a way or bring about a state of the world that he regards as desirable (or undesirable) if he complies (or fails to comply). That is, we are appealing to the ways in which the agent evaluates actions or states of affairs *independently* of our addressing him. By contrast, when we hold an agent accountable, we are trying to influence his behaviour by getting him to appreciate, freely and rationally, the *legitimacy* of our demand. Insofar as we announce sanctions for non-compliance (or rewards for compliance) at all, we are not merely expecting the agent to change his behaviour according to his independent evaluations of these sanctions (or rewards) but trying to get him to acknowledge the legitimacy of these measures (Darwall 2006: 250).

Above we have seen that, by addressing Arnold second-personally, Bella attempts to give him a reason whose validity derives from her irreducibly second-personal authority to address it. In light of the difference between holding accountable and goading, this means that by addressing Arnold second-personally, Bella is trying to get him to determine himself, freely and rationally, *by her authority* to address him in this way. Connectedly, Darwall notes that second-personal authority is, most fundamentally, the “authority to demand respect *for this very authority*” (2006: 14, italics added). Therefore, when addressing Arnold second-personally, Bella presupposes that he can *internalise* her demand. That is, Arnold must be able to reflect Bella’s authority within his deliberation and address himself by *holding himself accountable* for complying with her demand (Darwall 2006: 111-115). In Darwall’s terminology, Bella’s second-personal authority implies the Arnold’s “second-personal competence” (2006: 21). But this means that the demand must be addressable from a perspective that Bella and Arnold, addresser and addressee, can *share*. This shared perspective is what Darwall calls the “second-person standpoint” (2006: 114). It provides a nexus for persons to engage with one another in a mode of reasoning that is irreducibly second-personal in that its sole standard of validity is the shared authority to address one another second-personally.

How does any of this help us elicit the Kantian analysis of obligation? The answer is that, as I shall demonstrate shortly, second-personal competence entails autonomy of the will (Darwall 2006: 275). For a brief sketch of the argument, note that in addressing Arnold second-personally, and thus holding him accountable, Bella presupposes, not

only that Arnold can freely and rationally acknowledge the legitimacy of her demand, but also that he can determine himself by her demand so as to *act* accordingly. In Darwall's words, holding persons accountable presupposes "that there was a process of reasoning they could have engaged in by which they could have held themselves responsible and determined themselves to act as they should have" (2006: 241). Now, recall that the second-personal reason that Bella addresses in holding Arnold accountable rests on her irreducibly second-personal authority, and is thus independent of non-second-personal considerations such as whether or not the *content* of her demand is desirable, valuable or required independently of her demanding it. We might say that Bella's demand is valid "independently of any characteristic of the objects of willing" (G 4:440). But this means that Arnold's ability to appreciate and act on the considerations that legitimise Bella's demand, i.e., his second-personal competence, involves the ability to shape his actions in light of considerations that do not concern the *substance* of his volition. This, in turn, implies that Arnold must be able to determine himself by considerations that pertain to his will, *qua* will, i.e., by virtue of its *form*. In other words, his will must be susceptible to certain considerations merely because they take the form of demands—independently of *what* those considerations are, and of what *other* reasons he might have for acting as they require.

As this brief sketch of the argument indicates, the basic idea is that an agent must possess autonomy of the will to be second-personally competent. I now want to give a longer, more detailed argument and thereby convey a deeper appreciation of *why* autonomy is required for second-personal competence, while also linking this competence to the motive of duty and the Categorical Imperative.

Start by noting, again, that a second-personally competent agent is capable, not only of performing the act that her second-personal reasons demand, but of freely and rationally appreciating the legitimacy of this demand and directing herself by this appreciation. It follows that she must be capable, first, of ascertaining the legitimacy of the demand by her own *reasoning*, and second, of being *motivated* by this reasoning (Darwall 2017: 266). But we have noted above that the second-personal authority by which second-personal reasons are created is *irreducibly* second-personal in that it does not depend on any non-second-personal considerations, such as the value or desirability of outcomes. And since the second-personal authority that creates the accountable agent's second-personal reason must be *shared* between the addresser and the addressee of that reason, it is not the authority of the addresser, *qua* the particular person she is, but second-personal authority *as such*. So, the reasoning by which the accountable agent ascertains the legitimacy of the demand and motivates herself to act on it must involve

the judgment that the agent must act as demanded because it *can* be demanded of her, *qua* second-personally competent agent. In short, the relevant reasoning is irreducibly second-personal and the corresponding motive is a *sui generis*, second-personal motive.

From this, it can be inferred that the validity of second-personal reasons is determined by the Categorical Imperative and that a second-personally competent agent must be capable of acting from the motive of duty. To see this, recall that to φ from the motive of duty is to φ , motivated by the recognition that one is obligated to φ (see Chapter 1 above). But if one is obligated to φ only if one can φ from duty, then to φ from the motive of duty is to φ , motivated by the recognition that one *can* φ from duty. In this way, the volition involved in acting from duty turned out to have a *reflexive* structure. Now, we have seen that an agent is second-personally competent only if she can determine herself to φ by holding herself accountable for φ -ing. And this means that she must be capable of φ -ing, motivated by the recognition that she *can* legitimately be held accountable for φ -ing. But if legitimately being held accountable *presupposes* that one can determine oneself by holding oneself accountable, it follows that to determine oneself to φ by holding oneself accountable for φ -ing is to φ , motivated by the recognition that one *can* hold oneself accountable for φ -ing. Therefore, the volition involved in determining oneself by holding oneself accountable has the *same* reflexive structure as that involved in acting from the motive of duty.

Indeed, I claim that the volition involved in holding oneself accountable must be identical to that involved in acting from duty. After all, in φ -ing, motivated by the recognition that she can hold herself accountable for φ -ing, the agent is motivated, not by independent aspects of the substance of her volition, but by its being demandable from the second-person standpoint (Darwall 2017: 278). Now, since the second-person standpoint is a perspective that all second-personally competent agents can share, it follows that she is motivated by the recognition that it can be demanded of her that she φ , *qua* second-personally competent agent (Darwall 2006: 269-271). But this means that what motivates her is her free and rational acceptance of *every* second-personal agent's determining herself in the same way—i.e., to φ under the circumstances in which the agent finds herself—by holding themselves accountable from the second-person standpoint. The principle or standard by which the agent holds herself accountable must, then, be fit to be *universal* in the sense that the agent could freely and rationally accept that it governs any second-personally competent agent's volition. This means that, on the second-personal account, the content of our second-personal reasons, and thus of our obligations, is determined by the Formula of Universal Law: “*act only according to that maxim through which you can at the same time will that it become a universal law*” (G 4:421,

cf. Darwall 2006: 304-309). And this, in turn, confirms that the volition involved in determining oneself by holding oneself accountable from the second-person standpoint is identical to the volition involved in acting from duty. It is a matter of shaping one's actions by "[n]othing other than *the representation of the law in itself*" (G 4:401, cf. Darwall 2006: 275). We can thus see why second-personal competence requires Kantian autonomy of the will: it requires the capacity to act on principles that bind the will of every person by its mere form.

2.3 Authority and Equal Dignity

So far, I have argued for the claim that the second-personal account supports the Kantian analysis of obligation by pointing out that second-personal competence requires autonomy of the will—or, equivalently, the capacity to act from respect for the law—and that our second-personal reasons are determined by the Formula of Universal Law. My focus has thus been on Kantian notions that pertain most directly to the agent who is under moral obligation. As I want to show now, the conclusion that the second-personal account supports the Kantian analysis can be reinforced by showing that this account also accommodates Kantian notions that emphasise the role of moral patients, such as the dignity of persons as ends-in-themselves. Consequently, the second-personal account also illuminates the relationship between different parts of the Kantian framework, including the various formulations of the Categorical Imperative.

I have noted that, when she holds Arnold accountable, Bella is presupposing that Arnold can freely and rationally appreciate the legitimacy of her demand and thereby respect her second-personal authority. Now, Darwall argues that the capacity to respect another's second-personal authority is always reciprocal. In other words, in presupposing that Arnold is capable of respecting Bella's second-personal authority, Bella presupposes that Arnold also possesses second-personal authority, which Bella is capable of respecting. For this reason, according to Darwall, the second-personal account of obligation supports a conception of morality as *equal* accountability (2006: 101):

According to this conception, moral norms regulate a community of equal, mutually accountable, free and rational agents as such, and moral obligations are the demands such agents have standing to address to one another and with which they are mutually accountable for complying.

To see that the capacity to respect someone's second-personal authority is always reciprocal recall that, in presupposing that Arnold possesses second-personal competence, Bella presupposes that he can hold himself accountable for complying with her demand

from a perspective that they both *share* (Darwall 2006: 259). Since Arnold's ability to hold himself accountable from this shared perspective is what makes Bella's demand legitimate, it follows that Arnold thereby exercises the *same* kind of authority that Bella exercises when making her demand on Arnold.

This shows that, in addressing Arnold second-personally, Bella presupposes that Arnold, too, possesses second-personal authority. But does it also show that Bella is, in turn, capable of respecting that authority by determining herself in accordance with Arnold's demands? To see that it does, consider that, since Bella presupposes that Arnold can determine himself to act as her demands direct by exercising his second-personal authority, she implicitly acknowledges that there is a *limit* to what she can legitimately demand of Arnold (Darwall 2006: 259). That is, she acknowledges a distinction between demands that she can legitimately make of him, i.e., ones that he can be held accountable for complying with from the second-person standpoint, and ways of addressing Arnold that could not be justified from within that shared perspective and that, consequently, do not count as legitimate uses of her second-personal authority. Instances of address of the latter sort are essentially attempts to *bypass* Arnold's second-personal competence and are therefore to be considered instances of goading rather than of holding Arnold accountable. In acknowledging that different ways of addressing Arnold are legitimate or illegitimate depending on their being addressable from the shared perspective of all second-personally competent agents, Bella is already implicitly holding herself accountable for complying with the standards of that perspective. In this way, an agent's second-personal authority presupposes her second-personal competence.

It is important to note that Bella implicitly acknowledges that both she and Arnold equally possess second-personal competence and authority even *prior* to making any specific demands of him. This implicit acknowledgment already follows from Bella's regarding Arnold as a potential addressee of second-personal address, someone who is receptive to demands addressed person-to-person. After all, by regarding Arnold in this way she already views him as someone with the standing to demand that she relate to him in a certain way—in an addressive and justifying, rather than in a goading fashion. Accordingly, insofar as *all* persons regard each other (and themselves) as potential participants in the practice of second-personal address, then—regardless of whether they have ever explicitly engaged with each other in this way—they already view themselves as equal members of a moral community who possess an equal second-personal standing to address demands at one another.

Darwall points out that the distinction between legitimate and illegitimate ways of relating to someone, which is implicit in the normative felicity conditions of second-

personal address, does not only mean that we must regard ourselves as equal members of a community of mutually accountable beings. It also tells us something about what we are mutually accountable *for*. That is, it also gives rise to a *substantive* requirement that all participants in the practice of second-personal address, as such, are obligated to follow. In particular, we are obligated to act only in those ways that we can justify to all persons from within the second-person standpoint (Darwall 2006: 262). To violate this requirement with respect to any person is to relate to that person in a way that bypasses her (second-personal) standing as an equal member of the moral community. Darwall points out that this requirement can be understood as expressing the thought that every person is assigned a “sphere of freedom” in which no one may legitimately interfere with their conduct (2006: 262). After all, by violating this requirement with respect to a person, one fails to address them as a free and rational, second-personally competent agent, and instead attempts to “determine” them in a way that cannot be justified to them from within the second-person standpoint (ibid., cf. Darwall 2013c, 2013e).

Note that although the relevant ‘sphere of freedom’ is concerned with *external* freedom, it is ultimately based on a type of *internal* freedom, namely the agent’s capacity to hold herself accountable from the second-person standpoint. External acts of interference with the agent’s conduct are violations of the agent’s ‘sphere of freedom’ if and only if the agent cannot appreciate their legitimacy from within the second-person standpoint.

According to Darwall, the conception of morality as equal accountability can be interpreted as a form of contractualism, according to which “an act is wrong if the act would be disallowed by a principle no one could reasonably reject our holding one another accountable for complying with” (2006: 301, cf. Hill 1989, Kumar 2003, Scanlon 1998, Wallace 2013). And he suggests that this conception also provides the most promising interpretation of the Kantian requirement to respect the dignity of persons, as expressed in the Formula of Humanity: “*So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means*” (G 4:429, Darwall 2006: 306).⁴²

⁴² As Sensen points out, Kant does not think that persons are ends in themselves *because* they have dignity. Rather, they possess dignity because they are ends in themselves and, as such, must be respected (2011: Chapters 3 and 5). As I shall argue in the next chapter, the second-personal account is compatible with this reading. However, following large parts of the literature and Darwall’s way of phrasing his arguments, I shall occasionally speak as if persons are to be respected as ends in themselves because they possess dignity within this chapter.

In order to motivate this claim, Darwall notes that, when we treat someone as a mere means, there is a sense in which we fail to take seriously her rational nature. At best, we are utilising her rational capacities in a purely instrumental fashion (2006: 305). He suggests that what goes wrong in these cases can, again, be illustrated with the contrast between goading a person and addressing her second-personally (see Section 2.2 above). When we threaten or otherwise goad someone, there is a sense in which we are appealing to her ability to *reason* by counting on her to act in a certain way on the basis of her independent evaluation of outcomes or actions. But we are still merely utilising her rational nature in an instrumental fashion, and thus using her as a mere means, insofar as we fail to make our actions *justifiable* to her from within the second-person standpoint. We fail to treat her as a ‘free and rational’ being because we aim to ‘determine’ her from the outside, as it were, in a way in which she could not determine *herself* by holding herself accountable.

Now, I have said that an agent can determine herself by a certain principle or standard by holding herself accountable only if this principle or standard is universal in the sense that *any* second-personally competent agent could be held accountable for complying with it. This means that, if I am to treat a person as an end in herself, I must relate to her in a way that is consistent with her determining herself by the Formula of Universal Law. Thus, the second-personal account vindicates the following *dictum* of Kant’s (*G* 4:438):

[E]very rational being, as an end in itself, must be able to view itself as at the same time universally legislating with regard to any law whatsoever to which it may be subject, because it is just this fittingness of its maxims for universal legislation that marks it out as an end in itself [...].

Respecting a person’s dignity, then, requires relating to her in a way that is consistent with her exercising autonomy of the will. Now, this means that the actions that are allowed by the system of obligations that governs the moral community according to Darwall’s conception of morality as equal accountability are those that are compatible with every person’s holding herself accountable for complying with this system of obligations from the second-person standpoint. But I have also said that, if our actions are to be justifiable to others from within the second-person standpoint, our volition must pass the test of the Formula of Universal Law; in other words, our actions must be compatible with our autonomy. It follows that those actions that are compatible with the Formula of Universal Law, and thus with our *own* autonomy, are also those actions that are compatible with respect for the dignity of persons as ends in themselves, and thus with

everyone's autonomy. This thought is expressed in Kant's notion of the kingdom of ends (*G* 4:438):

[The dignity of the rational being] brings with it that it must always take its maxims from the point of view of itself, but also at the same time of every other rational being as legislating (which are therefore also called persons). Now in this way a world of rational beings (*mundus intelligibilis*) as a kingdom of ends is possible, and possible through their own legislation of all persons as members.

This notion leads Kant to the Formula of the Kingdom of Ends: "every rational being must so act as if through its maxims it were at all times a legislating member of the universal kingdom of ends" (*G* 4:438). Accordingly, Darwall sums up his conception of morality as equal accountability as follows (2006: 101):

[N]orms of moral obligation are "laws" for a "kingdom of ends", which structure and define the equal dignity of persons as beings who may not be treated in some ways and must be in others and who have equal standing to demand this second-personally of one another.

2.4 Dignity and the Fact of Reason

We are now in a position to see how the second-personal account allows us to reconstruct the fact-of-reason argument in such a way that it supports the Kantian analysis of obligation.⁴³ Recall that Kant asks us to imagine an agent whose prince demands that he give false testimony against an innocent person, on pain of losing his life (*CPrR* 5:30). And from the fact (if it is a fact) that the agent is obligated to refuse the prince's demand, we

⁴³ By arguing that the second-personal account supports Kantian constructivism, I reject Darwall's claim that the second-personal account is not committed to a constructivist metaethics (Darwall 2006: 293-299). Darwall claims that his theory is equally compatible with "recognition theories", according to which "deliberation is a matter of registering considerations whose status as reasons is independent of any ideal of deliberation" (2006: 297). His argument for this claim is that "there seems no doubt whatsoever that we experience feelings that present themselves as responses to reasons grounded in dignity, that is, as instances of Kantian *reverentia*" (2006: 298). However, this argument seems to establish, not that second-personal reasons are compatible with recognition theories, but that recognition theorists should—on the basis of their own premises about the veridicality of recognition—accept that some reasons have their source in the second-person standpoint, and thus abandon their view that all reasons are independent of any ideal of deliberation. I say more on the metaethical implications of the second-personal account in the next chapter.

are supposed to conclude that he *can* refuse it. And this, in turn, is supposed to show that we must suppose the agent to possess autonomy of the will. However, as we saw above, it is not immediately clear why it follows, from the notion that the agent *can* refuse, that the agent is capable of *autonomous* volition.

The second-personal account closes this gap by directing our attention to the conceptual link between obligation and accountability. That the agent is obligated to refuse the prince's demand implies that he can legitimately be held accountable, i.e., that he is blameworthy, if he fails to do so without excuse. This implies that the agent can direct himself to refuse the prince's demand by holding himself accountable from the second-person standpoint. From this, it follows that the agent is capable of acting from the motive of duty, and therefore possesses autonomy of the will. Indeed, as I have pointed out in the previous section, the second-personal account's emphasis on the link between obligation and accountability also helps us to connect the fact of reason, and the Kantian analysis of obligation that it elicits, with parts of the Kantian framework that concern the role of moral patients, namely the notions of dignity and the kingdom of ends.

Now, the second-personal account's key advantage over the first-personal account is that it does not attempt to derive autonomy of the will and related Kantian notions from the (non-normative) necessity of agency and practical deliberation.⁴⁴ The second-personal account acknowledges that, in order to vindicate the Kantian analysis of obligation, we need to appeal to a 'practical insight' on the part of agents, i.e., the way in which they conceptualise obligations that they already take themselves to have. The second-personal framework fulfils this need by appealing to the notion of second-personal authority. In particular, the second-personal account points out that we already acknowledge second-personal authority, be it vested in our own person or in someone

⁴⁴ It is worth noting that this feature allows the second-personal account to avoid another objection that threatens the first-personal account. Fitzpatrick objects that, on Korsgaard's account, "ethics is not fundamentally about goodness, rightness, and rationality, as one might have thought; it is [...] really all about *being a unified agent rather than a disunified heap*" (2013: 44; cf. Fitzpatrick and Watson 2002: 361). And this, according to Fitzpatrick, "is more a change of subject than an account of ethics or normativity" (ibid.) The second-personal account avoids this objection because it does not appeal to the (non-normative) necessity of being a unified agent in order to elicit the Kantian analysis of obligation.

else, and thereby implicitly presuppose the dignity and autonomy of all members of the moral community. Accordingly, Darwall writes (2006: 242):

What is fundamental in this way of thinking is the second-personal authority that I have been claiming is an essential aspect of the dignity of persons [...]. Autonomy of the will and the necessity of a formal deliberative process [such as the Categorical Imperative test] derive from this more fundamental idea—as necessary conditions for the possibility of second-personal authority and as necessary to mediate second-personal relations—rather than vice versa.

Note that, although a *direct* encounter with the moral law (through second-personal address) is necessary to elicit the Kantian analysis of obligation, the second-personal account has the implication that this encounter need not take place in an *explicitly moral* context.

After all, the practice of addressing purportedly legitimate demands extends to apparently nonmoral cases (Darwall 2006: 80-82). Darwall employs the example of a sergeant addressing commands at her troops (*ibid.*). Not only is this example not what we would consider a paradigmatically moral case, it also involves the asymmetric, hierarchical authority of one person over others. As such, it might not immediately strike us as involving an encounter with the moral requirement to respect the *equal* dignity of all persons. However, Darwall argues that, insofar as the sergeant purports to address a second-personal reason with her troops, i.e., insofar as she tries to address a reason that would not exist but for her legitimate authority over the troops, she must implicitly presuppose that her troops could direct themselves in accordance with this reason by holding themselves accountable from within the second-person standpoint.⁴⁵ Conversely, by acknowledging the sergeant's command and internally holding themselves accountable, the troops implicitly presuppose their equal second-personal authority. In this way, both the sergeants and the troops are in a fact-of-reason situation in which they gain a practical insight into their obligation to determine themselves by the moral law, if only implicitly.⁴⁶

⁴⁵ Of course, it is possible that she is not trying to address a second-personal reason. She might simply be trying to *goad* the troops by uttering something that appears to be a command but ultimately functions like a threat. The key feature of genuine second-personal address that would be missing in that case is that the demand in question is addressed *as legitimate* (Darwall 2010: 257).

⁴⁶ Empirical evidence for the claim that a second-personal understanding of morals pervades our practices is compiled by Darwall and de Kenessey (2014).

Importantly, this means that, although the second-personal account abandons the ambition of eliciting the Kantian analysis of obligation from a morally neutral starting point (such as the inescapable standpoint of agency), it asserts that autonomy of the will and related Kantian concepts are employed in practices that are not explicitly cast in moral terms, let alone in terms of a Kantian morality of respect for the equal dignity of all persons.

Of course, the second-personal account is thereby committed to the claim that contexts that involve a genuine instance of second-personal address are at most *apparently* nonmoral contexts. Ultimately, any successful instance of second-personal address creates a moral obligation. After all, to neglect a person's legitimate demand always amounts to a failure to respect her second-personal authority and, thus, her equal dignity as a member of the moral community. Note that, in this way, the second-personal account not only broadens the range of cases that involve a fact-of-reason type encounter with the moral law, but it also increases the domain in which the concept of moral obligation applies. It thereby widens the scope of the Kantian constructivist claim about the source of moral normativity, which, after all, is exclusively concerned with the phenomenon of moral obligation (see Chapter 1 above).

3. The Irreducibility of Second-Personal Reasoning

The argument in favour of the second-personal account and against the first-personal account rests on the premises that, first, only second-personal reasoning vindicates the Kantian analysis of obligation and, second, second-personal reasoning is fundamentally distinct from first-personal reasoning. Darwall expresses the second premise as follows (2006: 12):

These notions—second-personal authority, valid claim or demand, second-personal reason, and responsibility to—therefore comprise an interdefinable circle; each implies all the rest. Moreover, I contend there is no way to break into this circle from outside it. Propositions formulated only with normative and evaluative concepts that are not already implicitly second-personal cannot adequately ground propositions formulated with concepts within the circle.

However, there are some theorists who, while granting the first premise (that the second-person standpoint is required to elicit the Kantian analysis), reject the second premise. Their arguments against the second premise come in two forms. Korsgaard and Tamar Schapiro argue that the second-person standpoint is *integrated* into the first-person standpoint that all agents as such take up (Korsgaard 2007, Schapiro 2010). That is, they

doubt that there is any conceptual gap between first-personal and second-personal notions to begin with. By contrast, Christoph Hanisch argues that, although second-personal reasoning is conceptually distinct from first-personal reasoning, we are nevertheless required to *take up* the second-person standpoint insofar as we reason first-personally at all (2014). Thus, while granting Darwall's notion of a circle of irreducibly second-personal concepts, Hanisch rejects the claim that "there is no way to break into this circle from outside it" (Darwall 2006: 12). I discuss Korsgaard's, Schapiro's, and Hanisch's arguments in turn. I argue that all three philosophers fail to appreciate the full force of the *irreducibly* second-personal nature of second-personal concepts.

3.1 Korsgaard: The Second-Person Within

Korsgaard concedes Darwall's central insight that obligation is second-personal in the sense that it is conceptually tied to accountability (2007: 10). Nonetheless, she defends her first-personal account by arguing that the presuppositions of the second-person standpoint which elicit the Kantian analysis of obligation are also necessary presuppositions of the first-person standpoint (2007: 14-20). Korsgaard thus characterizes herself "both as someone who thinks all reasons are second personal and also as someone who thinks 'moral obligations can be grounded in the constraints of first-personal deliberation alone'" (Korsgaard 2007: 10, citing Darwall 2006: 11). She advances four arguments in favour of her view.

First, Korsgaard argues that "the reflective structure of self-consciousness inevitably places us in a relation of authority over ourselves and that we are as a consequence also accountable to ourselves" (2007: 10, cf. 1996: 104). In favour of this claim, Korsgaard essentially repeats her argument for the first-personal account (which I discussed in this previous chapter), adding that this argument can be read as establishing that we encounter the moral law through being addressed by "the voice of the second-person within" (2007: 23). From this, she infers that Darwall's argument from accountability to autonomy, which I outlined in Section 2.2 above, can take off from within the first-person standpoint which all agents as such take up. Indeed, Korsgaard contends that, since Darwall's argument employs the idea that we cannot be obligated to act in a certain way unless we can hold ourselves accountable to do so (see Section 2.2 above), he is already committed to the view that we need not take up a standpoint which is distinct

from the first-person standpoint to realize that we are autonomous (Korsgaard 2007: 11, cf. Darwall 2006: 111-112).⁴⁷

However, as we saw in Chapter 4, Darwall makes the case that first-personal reasoning, to be intelligible, need not conform to Korsgaard's conception of reflective consciousness. It is entirely possible, within the constraints of the first-person standpoint, to be a naïve practical reasoner whose deliberations are determined by the desirability or value of states of the world, as these are presented by one's desires. Therefore, while it is true that an agent can intelligibly be held accountable only if she can hold herself accountable, there is nothing in the necessary presuppositions of the first-person standpoint alone that prompts her to regard herself as accountable to anyone—be it herself or other persons—in the first place. As Darwall puts it in his reply to Korsgaard, “I can view things in my ‘deliberative field’ in relation to me without relating to anything in it as a you” (2007: 55).

Korsgaard's second argument challenges the intelligibility of the naïve practical perspective by questioning the coherence of the notion of state-of-the-world-regarding reasons (2007: 14-15, see Section 2.1 above). Since Darwall accepts that at least some reasons—namely, second-personal reasons—are grounded in relations of accountability and authority, Korsgaard raises the question what Darwall takes state-of-the-world-regarding reasons to be grounded in. What, if not the agent's authority over herself, is the source of their normativity?⁴⁸

However, this argument of Korsgaard's does not seem to constitute a genuine objection to Darwall's account. It is important to note that Darwall does not defend the naïve practical perspective as an attractive characterization of practical reason. Rather,

⁴⁷ Korsgaard concedes, however, that Darwall's notion of the second-person standpoint explains how it is possible that we are obligated *to others* (2007: 9). This suggests that she considers Darwall's work to provide an elaboration of her claims about the shareability of reasons (see Chapter 4 above, and Chapter 7, Section 6 below).

⁴⁸ Note that Schapiro makes a similar objection to Darwall's employment of the naïve practical perspective, particularly the way in which desires are “backgrounded” in this perspective (Schapiro 2010: 231-235). More specifically, Schapiro contends that the naïve practical perspective would not be regarded as intelligible by someone like Kant, who thinks first-personal reasoning includes the critical reflection on one's desires. Accordingly, the naïve practical standpoint cannot be employed in an internal critique of Kant. Darwall makes a reply to this objection that resembles my reply to Korsgaard's objection. In particular, he argues that he does not aim at providing an *internal* critique of Kant. Rather, he criticizes Kant for not presenting an argument for autonomy based on premises that *all* his readers can accept (Darwall 2010: 254-255).

he employs it as an example of a form of practical reasoning that, while conforming to the constraints on merely first-personal reasoning, does not presuppose autonomy. For his argument to succeed, then, he need not vindicate the existence of genuinely normative state-of-the-world-regarding reasons. All he needs to assert is that reasoning from the naïve practical perspective would be an intelligible form of practical reasoning and that someone who only ever acknowledged and acted on state-of-the-world regarding reasons would not fail to be an agent.

Besides, it is not immediately clear why the notion of state-of-the-world-regarding reasons is supposed to be incoherent or implausible. Therefore, even if Darwall had to defend the existence of genuinely normative state-of-the-world-regarding reasons in order to motivate his critique of the first-personal account, it is not clear that his argument could not succeed. Regarding the source of the normativity of state-of-the-world-regarding reasons, it would be open to him to adopt a source externalist view, according to which the source of their normativity lies in irreducibly normative properties. So, unless Korsgaard makes the case against source externalism on the basis of the necessary presuppositions of the first-person standpoint, she cannot rule out state-of-the-world-regarding reasons as incoherent or implausible.

Korsgaard's third argument seems to be meant to achieve precisely this (2007: 15-20). The aim of the argument is to defend both (what she identifies as) Kant's derivation of autonomy from within the first-person standpoint and his fact of reason argument. She follows Kant in holding that, when we act, we cannot conceive of our reason as "consciously receiving direction from any other quarter" (*G* 4:448, Korsgaard 2007: 15). Yet, she notes, "a will that simply accepted the authority of state-of-the-world-regarding reasons *would be* governed by something alien to itself" (2007: 15, italics Korsgaard's). Therefore, according to Korsgaard, the notion of state-of-the-world-regarding reasons is incompatible with the necessary presuppositions of agency. These presuppositions require that, when we act, we consider ourselves guided by considerations that bind the will as such, i.e., by its mere form. Accordingly, when the agent in the fact of reason example realizes that she can refuse the prince's demand because this is what she *ought* to do, this realization already entails that the agent's will possesses the capacity to be a law to itself (Korsgaard 2007: 17-19).

The problem with this argument is that, like Korsgaard's earlier argument from *Sources*, it relies on an equivocation of 'agency', 'will' and related terms (see Chapter 4 above). It is true that a *will*, in the Kantian sense of that term, cannot consider itself to be determined by something other than its own form. But as soon as we accept this usage of the term 'will', the question is no longer whether the will is autonomous, but whether

every agent, in the minimal sense of ‘agent’ employed by the first-personal account, possesses a will (cf. Darwall 2006: 223). If we want to elicit the Kantian analysis of obligation, it will not help to simply *define* the will as a faculty that has the characteristic of autonomy. For then it becomes questionable whether agents do indeed possess this faculty. Now, Darwall agrees with Korsgaard that the fact-of-reason argument is our best hope to show that we already presuppose that we possess this faculty. But, as opposed to Korsgaard, he believes that, if the fact-of-reason argument is to be successful, it needs to emphasize the *second-personal* character of moral obligation (2006: 235-241, 2007: 57). Hence, if Korsgaard is to show that the second-person standpoint, conceived as separate from the first-person standpoint, is unnecessary, it will not be sufficient to claim that a Kantian will is autonomous by definition. Instead, she needs to show that the reasoning that leads Darwall to the conclusion that the first-personal strategy fails is flawed.

Indeed, Korsgaard’s final argument seems to be an attempt to show that there is something wrong with the argumentative strategy that Darwall employs. Specifically, she argues that Darwall’s criticism of Korsgaard’s arguments can also be turned against the second-personal account (2007: 20). In particular, she raises doubts concerning the necessity of autonomy for acting on second-personal reasons. Could we not devise a naïve practical reasoner who takes certain features of persons as indicators of authority, and thus as warranting an attitude of submission to that authority? Such a reasoner would not view this attitude as self-imposed, but as being warranted by features of the world.

Darwall replies that it is hard to see how the naïve second-personal reasoner that Korsgaard envisages could intelligibly be held accountable (2007: 58-59). As we saw in section 2.2, Darwall argues that genuine accountability entails that one can freely determine oneself as demanded by holding oneself accountable from the second-person standpoint. Conversely, an agent of the kind Korsgaard is envisaging—who regards the authority of requirements as flowing from features of the world—cannot truly be held accountable. For such an agent will not endorse the normativity of the reasons in question as grounded in relations of authority and accountability she can see to be justifiable to her as an equal member of the moral community. Korsgaard’s naïve second-personal reasoner is thus analogous to Arnold, when he withdraws his foot because this way he can bring about a state of the world which is more valuable than the present one. While Arnold acts as Bella has the authority to demand that he act, he does not act *on* this consideration.

3.2 Schapiro: Desires as Demands

Tamar Schapiro concedes Darwall's claim that the Kantian analysis of obligation is elicited only through an encounter with second-personal reasons. She credits Darwall with highlighting "a genuine gap in Kant's and Korsgaard's arguments" by pointing out the link between autonomy and morality on the one hand, and accountability on the other (2010: 230). She sums up this central lesson by stating that "[u]nless the deliberative standpoint is inherently a standpoint from which we are summoned second-personally, our taking it up cannot make us aware of our second-personal autonomy" (2010: 233).

Unlike Darwall, however, Schapiro does not conclude from this that the Kantian analysis cannot be elicited by the constraints of first-personal reasoning. By contrast, she argues that, within the first-person standpoint, we *are* addressed second-personally. In particular, she suggests a conception of desire on which "[d]esire is a *demand on the will*, made in light of a motivationally-efficacious form of consciousness" (2010: 234, italics Schapiro's). By 'desire' Schapiro is referring to the particular kind of mental state which Kant calls "inclination", viz. the "dependence of the desiderative faculty on sensations" (*G* 4:413).⁴⁹ Indeed, Schapiro claims that her conception of desires as demands on the will is suggested by Kant's moral psychology (2010: 234).

Schapiro starts by arguing against the notion that desires are "backgrounded" in rational first-personal deliberation, which Darwall invoked in his characterization of the naïve practical perspective (see Chapter 4, Section 3.1, cf. Pettit and Smith 1990). She contends that this view is either put forward as "cognitivist", in which case it takes a desire to be a representation of the desired state of the world, or as "non-cognitivist", in which case it takes a desire to be akin to a sensation (2010: 234). Schapiro claims that both versions of the view face difficulties when trying to explain that desire is something that can be *satisfied*. By contrast, she argues, Kant employs a distinction between our rational self and our animal nature that provides an account on which our desires are drawn into the focus of critical reflection: our rational self ponders whether it should act on the desires of our animal nature (cf. *CPrR* 5:61). As a result, desires are experienced as principles that purport to motivate us to act in a certain way, and, thus, as satisfiable (2010: 234-235). Importantly, according to Schapiro, "in bringing desires into the foreground, *qua* motivational principles, we become aware of them as making purportedly *authoritative claims* to direct our wills" (2010: 235, italics added). In this view, our

⁴⁹ What Schapiro calls desires therefore belong in the class of what Rawls calls "object-dependent" (as opposed to "principle-dependent") desires (Rawls 2000: 45-48 and 148-152).

overcoming of our animal nature not only allows us to reflect on our desires, it also presents them as making demands on us. For once we step back from our desires and reflect on them we realise that they attempt to take control of our actions.⁵⁰

Schapiro admits that desires are not paradigmatic instances of second-personal address (2010: 235-236). First, she notes that desires use the practice of second-personal address “abusively” (cf. Darwall 2006: 265). That is, they do not constitute legitimate demands as they do not satisfy the necessary presuppositions of second-personal address. I suppose that Schapiro makes this concession since desires stem from our animal nature and only our rational nature has dignity, i.e., the authority to make demands. Second, she observes that the “you” that we recognize in dealing with our desires is actually part of ourselves. Nevertheless, she argues that our encounter with our own desires is second-personal, and thus suffices for us to enter the second-person standpoint, since their “capacity to address an abusive claim still presupposes some form of second-personal standing” (2010: 235).

In his reply, Darwall concedes that some desires are phenomenologically more akin to instances of address than the notion that desires are “backgrounded” suggests. For example, an intense feeling of hunger often presents itself, not as the piece of advice “eating would be good”, but as the quasi-second-personal demand “satisfy me” (2010: 257). However, he notes that to constitute even abusive instances second-personal addresses, these desires would have to present themselves as *legitimate*, and it is not clear that they do. In fact, they might rather fall within the class of address that, although it takes the form of demands, does not quite qualify as second-personal in Darwall’s technical sense. As examples of such addresses, he mentions “[p]ure expressions of force” such as “Move over Buddy” (2010: 257, cf. Darwall 2007: 55).

I think that this reply by Darwall highlights exactly what is wrong with Schapiro’s proposal. That is, although our desires might be seen “as purporting to have authority to govern our whole selves”, their ‘governance’ would not be governance by the rule of law, but rather that of an external force driving the will (Schapiro 2010: 235). To use Darwall’s analogy again, their authority would not be like the status that a sergeant might claim with respect to her troops, but rather like the power that a coachman exercises over his horses. As I said in Section 2, second-personal accountability presupposes that the

⁵⁰ Note the similarity between Schapiro’s account of desires and Korsgaard’s remarks on the reflective structure of consciousness, which lead her to suggest that our particular practical identities function as “standing sources of incentives as well as principles in terms of which we accept and reject proposed actions” (2009a: 22, also see Chapter 4 of this thesis).

accountable agent is capable of freely appreciating, and being motivated by, the legitimacy of a demand. No such thing seems required for giving in to the dictates of our desires.⁵¹

3.3 Hanisch: *Second-Personality and Self-Constitution*

Christoph Hanisch argues that all agents as such must reason second-personally because the constitutive aim of agency—which, following Korsgaard, he identifies as self-constitution—cannot be achieved without taking the second-person standpoint (2014). Unlike Schapiro and Korsgaard, however, he does not attempt to show that the second-person standpoint is *integrated* in the first-person standpoint. To the contrary, he emphasizes its distinctness from the standpoint of first-personal deliberation “as a free-standing autarkic source of normative requirements” (2014: 110). In particular, he thinks that the second-person standpoint is the locus of reasons that exist, and are accessible, only through actual, *interpersonal* interaction (*ibid.*).⁵²

Nonetheless, Hanisch attempts to establish that we must take up the second-person standpoint insofar as we are agents. For this, he relies upon Korsgaard’s work on agency and self-constitution (Korsgaard 2009, see Chapter 4 above). Recall that Korsgaard contends that, in deliberating about what to do, not only must we adopt a conception of ourselves, but we also need to make sure that this conception is integrated into one coherent whole with all our other practical identities (2009: 21). In support of this claim, she argues that for behaviour to be interpretable as an *action*, it needs to be attributable to a single *agent* (2009: 18-19). Our behaviour will not be thus attributable, however, if it exemplifies an incoherent, disintegrated set of self-conceptions.

Also recall that, according to Korsgaard, to have a certain conception of one’s identity is to be governed by this conception in one’s choice of *maxims*, i.e., principles of action stating that one will perform a certain *act* for the sake of a certain *end* (2009: 8-

⁵¹ There is a further reason why at least Kantians should be suspicious of Schapiro’s proposal. Specifically, as I have already noted in the context of Korsgaard’s first-personal account (see Chapter 4 above), Kant distinguishes between our capacity to act *rationally*, i.e., which entails the capacity to resist immediate inclination, and our capacity for *morality*, i.e., pure practical reason (*Rel* 6:26-28). Thus, within Kant’s framework, it would be surprising if our elevation from animal nature was enough to make us aware of our autonomy. After all, it seems that Kant could in principle allow for the possibility of a creature that is rational in that it is capable of being guided by practical reason, yet not moral in the sense of being capable of recognizing and acting on principles of *pure* practical reason (cf. Timmermann 2006a: 80-85).

⁵² I argue against this interpretation of the second-person standpoint in the next chapter.

14, 18-26). Thus, practical identities are “standing sources of incentives, as well as principles in terms of which we accept and reject proposed actions” (2009: 22). Note also that, according to Korsgaard, it is not the case that we *have* a determinate practical identity prior to our actions, which is then *expressed* through the maxims we choose (2009: 19). Instead, as long as we do not make any choices our identities are indeterminate. It is only when we take action that, through the maxims we choose, we *constitute* who we are.

In order for our self-constitution to be successful, then, it is vital that it is at least possible for us to act on the maxims we choose. Hanisch emphasizes that this means that successful self-constitution requires that we can be efficacious in bringing about the ends we set as part of the maxims we choose⁵³ (2014: 114). Now, he observes that for an agent to be efficacious in achieving her ends, the world she acts upon must follow certain regularities. In particular, “surrounding agents pose a potential source of interference that endangers every agents’ self-constituting activities, especially her attempt to conceive of certain ends as something that can be brought about by acting in a specific way” (ibid.) It is therefore necessary for an agent’s self-constitution that she can predict, with some degree of reliability, that others will not interfere with her activities. Hence, Hanisch argues, “practical deliberation presupposes that certain interpersonally shared principles of non-interference are held constant” (2014: 115).

Hanisch suggests that this necessary presupposition of shared principles of conduct is what commits self-constituting agents to taking up the second-person standpoint and, thus, assuming their own and others’ second-personal authority and competence (2014: 116-118, 122-126). Moreover, it requires them to recognize others as *independent* agents and, thus, as originators of their own maxims and actions. This forces them to engage in an actual social interaction with their fellow agents. It is thus not sufficient for self-constitution to solipsistically carry out second-personal reasoning by oneself (2014: 117).

The main flaw in Hanisch’s argument is that it assumes what it is supposed to show: that agents unavoidably presuppose that they are autonomous. This flaw is symmetrical to the one I identified in Korsgaard’s practical identities argument in the previous chapter (Section 3.2). In fact, the way Hanisch fleshes out his account of the first-person standpoint makes the flaw in Korsgaard’s account of practical identities and self-

⁵³ Of course, Hanisch does not suggest that if we are to constitute ourselves, our actions must always be successful. What he proposes is rather that, in order to *intelligibly* engage in self-constitution, i.e., to make any *meaningful* choices of maxims, we at least need to be able to assume that our actions on these maxims *can* be successful.

constitution even more apparent. Hanisch suggests that in order for *action* to be possible we must assume interpersonal normative principles that bar against chaotic interference with our plans. And these principles are supposed to be binding on every human being since every human being *must act*. However, it is obviously not the case that, if there are no interpersonal normative principles of non-interference, we cease to ‘act’ in any sense of that word in which it is *inevitable* for humans to act.

Again, Korsgaard’s argument equivocates two senses of the word ‘action’. In the first sense, action is an unavoidable part of the human condition: we must deliberate about what to do, we cannot cease to act. In the second sense, action implies the authorship of a fully constituted agent, someone with a determinate identity who chooses maxims in light of the system of normative principles she endorses as part of her self-conception. Now, as Darwall’s argument naïve practical perspective highlights, these two senses of ‘action’ are not the same. Hence, from the fact that action in the first sense is inevitable we cannot conclude that action in the second sense is non-optional.

That the way Hanisch depicts the first-person standpoint already includes second-personal competence can also be seen from the details of his argument for the claim that, in adopting maxims, we need to presuppose that others are *second-personally* competent. Hanisch contends that we cannot just view others as forces of nature that are held in check by the shared principles of non-interference we must rely on (2014: 122-126). Instead, according to Hanisch, we need to rely on others’ ability to freely endorse the legitimacy of our shared principle and bring themselves to act on them. To establish this conclusion, he relies on Wittgenstein’s considerations against the possibility of a private language (cf. Wittgenstein 2009: esp. §256-258). He takes these considerations to show that, whenever we follow a normative principle, we must presuppose an *equal* normative competence in others.⁵⁴ But, obviously, the conclusion that we must assume others to be *second-personally* competent follows only if we already assume *ourselves* to be second-personally competent. By contrast, if we take ourselves to act in accordance with the principles in question from sympathy for our fellow human beings, or from a realization of the fact that non-compliance will erode social stability, nothing forces us to view others as autonomous. Therefore, if Hanisch’s argument is to show that agents must take up the second-person standpoint in their interaction with others, it must assume that they already engage in second-personal reasoning.

⁵⁴ Also see Korsgaard’s similar argument for the publicity of reasons, which I outlined in Chapter 4, Section 2.2.

Indeed, if Hanisch did not implicitly assume that agents view themselves as second-personally competent *prior* to their encounter of others, his argument would amount to an attempt to provide a non-second-personal reason for entering the second-person standpoint. More specifically, if Hanisch did not assume that agents pursuing the task of self-constitution are already reasoning second-personally, then the fact (if it is a fact) that reasoning second-personally is required to complete the task of self-constitution would provide a technical or instrumental reason for taking up the second-person standpoint. However, Darwall would object—and, ironically, Hanisch seems to agree—that such a reason would be “the wrong kind of reason” to employ second-personal concepts (Darwall 2007: 59, Hanisch 2014: 119, also see Section 2.1 above). Indeed, I would make an even stronger claim by arguing that the second-person standpoint is not the kind of standpoint that we can *choose* to take up at all. I say more on this in the next chapter. For now, we can conclude that Hanisch does not succeed at showing that the necessary presuppositions of the merely first-personal standpoint of agency as such require that we take up the second-person standpoint.

4. Conclusion

In this chapter, I hope to have shown that the second-personal account successfully elicits the Kantian analysis of obligation, and thereby supports the two lines of argument in favour of Kantian constructivism presented in Chapters 1 to 3 by exploiting the link between obligation and accountability. Not only are second-personal concepts required to employ Kant’s fact-of-reason argument, but these concepts also form a separate, irreducibly second-personal domain of practical reason which we cannot reason ourselves into, as it were, from non-second-personal premises. Accordingly, the success of the second-personal account cannot be used to revive the first-personal account that I rejected in Chapter 4.

In the next two chapters, I reply to various potential objections to the second-personal account that might be raised from a Kantian, or Kantian constructivist, point of view. Most notably, one might argue that, if moral normativity has its source in the form of second-personal reasoning, then moral obligation is *optional*. After all, it seems possible to conceive of a hermit who never makes or acknowledges any demands, and thus never comes to reason second-personally. I respond to this objection by arguing that, although second-personal reasoning is fundamentally distinct from merely first-personal reasoning, the second-person standpoint is taken up in purely self-regarding as well as in other-regarding contexts.

Chapter 6: Kantian Constructivism and the Second-Person Standpoint

1. Introduction

In the previous chapter, I argued that we need to appeal to the irreducibly second-personal dimension of practical reasoning in order to elicit the Kantian analysis of obligation. In other words, we can vindicate autonomy of the will and related Kantian notions only by appeal to the practical concepts that are intimately linked to the idea of legitimately being held accountable by the moral community. Moreover, this idea does not form part of the necessary presuppositions of the first-person standpoint of practical deliberation which is inescapable for us *qua* agents.

In this chapter, I discuss several objections to the second-personal account that one might raise from a Kantian constructivist or, more generally, Kantian perspective. I argue that all of these objections fail to show that the second-personal account in some way undermines the Kantian constructivist programme.

In Section 2, I discuss a bundle of objections that are concerned with the relationship that holds between the Kantian notions of law, respect, and persons. All these objections, in some way or another, point out that it is essential to the Kantian framework that obligations depend on a first-personal encounter with the form of our own practical reason, as opposed to the presence of, or our interaction with, other people. I reply that the second-personal account does not characterise obligations as dependent, in some problematic way, on our interactions with other people. In Section 3, I turn to the objection that the second-personal account renders moral obligation optional because, in principle, one might never come to reason second-personally. I argue that, on its most plausible interpretation, this objection is concerned, not with the optionality of the standpoint of moral deliberation, but with the potential unavailability of this standpoint to isolated individuals. As I show in Section 4, this issue is related to the worry that the second-personal account does not accommodate Kantian duties to the self. I leave my reply to these connected worries for the next chapter.

2. Respect, Persons, and the Law

In this section, I respond to three objections that, in some way or another, accuse the second-personal account of locating the source of moral obligations in something other than the form of practical reason. All three objections are concerned with the role of respect, persons, and the law in the second-personal account.

2.1 Second-Personal Authority as Fundamental

As I have pointed out in the previous chapter, Darwall regards the dignity of persons, understood as second-personal authority, as the fundamental idea in his framework. He writes (2006: 242):

What is fundamental in this way of thinking is the second-personal authority that I have been claiming is an essential aspect of the dignity of persons [...]. Autonomy of the will and the necessity of a formal deliberative process [such as the Categorical Imperative test] derive from this more fundamental idea—as necessary conditions for the possibility of second-personal authority and as necessary to mediate second-personal relations—rather than vice versa.

One might take this to imply that second-personal authority plays a *foundational* role in Darwall's picture. That is, one might interpret Darwall's statement that second-personal authority is the 'fundamental idea' as indicating that the legitimate demands of those who possess this authority furnish the *inputs* to practical reasoning and thereby determine what the agent is rationally required to do, without themselves being subject to the critical scrutiny of reason (see Chapters 2 and 3 above). If this interpretation was correct, the second-personal account would not seem to support a Kantian constructivist view of the sources of moral normativity at all. After all, this interpretation suggests that the dignity of persons plays the role of a substantive value, or a set of substantive requirements, that possesses normative force independently of the form of our will (cf. Sensen 2011: 120-121, Timmermann 2014, Waldron 2012: 26).

This interpretation can be fleshed out in two different ways. First, one might take second-personal authority to be the standing to demand compliance with independently valid norms. In this view, the second-personal nature of obligations is exhausted by their being so demandable. This interpretation favours a source externalist metaethics since it suggests that the second-personal authority does not play a role in *grounding* obligations. Jay Wallace and Ariel Zylberman defend an account along these lines (Wallace 2007, 2013, Zylberman 2014, 2017, *pace* Darwall 2007, 2013b). Second, one might argue that obligations are grounded in second-personal authority but deny that second-personal authority is limited by the form of second-personal address. This view does not favour source externalism because it renders the normativity of moral obligation dependent on the addresser's *will*. By conceiving of the addresser's demands as substantive inputs to practical reason whose normativity does not depend on the *form* of the will, however, it instead favours a non-Kantian version of source voluntarism. This is the view that Wallace and Zylberman attribute to Darwall (Wallace 2007, Zylberman 2017). On

this view, moral obligations are determined by “the actual claims (implicit or explicit) that members of a moral community make on one another” (Zylberman 2017: 926). According to this view, then, members of the moral community have the power to create obligations by an act of will, without being limited by any formal constraints on willing, such as the normative felicity conditions of second-personal address.⁵⁵

However, neither the source externalist nor the non-Kantian source voluntarist alternative is correct as an interpretation of the second-personal account. Indeed, these views are incompatible with the irreducibly second-personal nature of the authority which, according to the second-personal account, underwrites the equal dignity of persons. To see this, recall that, if the demands that persons can legitimately make on one another were valid independently of their authority to address them, this authority would amount to a merely epistemic, first- or third-personal authority (see Chapter 5, Section 2.1). Accordingly, insofar as the authority that gives rise to moral obligations is to be an *irreducibly* second-personal authority, it cannot be exhausted by the standing to demand recognition of independent reasons, requirements, or values. Instead, second-personal authority involves the standing to *create* reasons that otherwise would not exist (Darwall 2006: 11).

Most fundamentally, then, second-personal authority amounts to the “authority to demand respect *for this very authority*” (Darwall 2006: 14, italics added). It follows that, in the second-personal framework, respect for the dignity of persons is not to be understood, primarily, as a matter of *treating* persons in the right way, by following a set of substantive requirements (ibid., *pace* Kamm 1992, Nagel 1995). This is certainly *part* of what is involved in respecting persons; for their second-personal authority gives rise to legitimate substantive demands. Most fundamentally, however, respecting second-personal authority is a matter of *relating* to persons in the right way, by acting on maxims that are compatible with their equal standing as members of the moral community endowed with second-personal authority and second-personal competence.

The demands that members of the moral community can legitimately make on each other, then, are not rendered legitimate by their *substantive* merits. Instead, they give rise to obligations only because, and insofar as, they spring from second-personal authority. This rules out the source externalist version of the interpretation sketched

⁵⁵ I am not aware of any theorist who defends this view. Of all the positions advanced in the literature, perhaps Susan Wolf’s “Social Command Theory of moral obligation” is the view that resembles this non-Kantian kind of source voluntarism most closely (2015).

above. After all, on this interpretation, second-personal authority is exhausted by the authority to demand compliance with *independently* valid demands.

The above observations also rule out the source voluntarist version of this interpretation. After all, as we have seen, a demand can be considered as springing from a person's second-personal authority if and only if it is justifiable, from within the second-person standpoint, to all members of the moral community.⁵⁶ Furthermore, a demand is justifiable in this sense if and only if it could be made of any second-personally competent agent *as such*. But this means that the demand's validity depends on the addressee's being able to hold herself accountable for complying with it from within the second-person standpoint through an exercise of her autonomy. In other words, it depends on the addressee's being able to follow the Categorical Imperative in complying with the demand.

This thought can be expressed in terms of a distinction that Darwall draws between two kinds of second-personal authority (2006, 2007, 2013). Darwall distinguishes between the "individual authority" by which a person demands compliance with obligations that are *owed* to her and the "representative authority" by which all members of the moral community can demand compliance with an obligation owed to someone else or, indeed, to no one in particular (2013b: 32-39). Individual authority, on this picture, corresponds to obligations that are "bipolar" in the sense that they are obligations *to* a person, and failure to fulfil these obligations is not merely wrong but *wrongs* that person (2013b: 20, cf. Thompson 2004). If a person merely has representative authority with respect to an obligation, by contrast, the obligation is either owed to *another* person or it is not an obligation *to* anyone at all. In the latter case, it is an instance, not of bipolar obligation, but of "moral obligation *period*" (ibid.). According to Darwall, the two different kinds of second-personal authority correspond to two different ways of holding an agent accountable: while individual authority is the standing to adopt *personal* reactive attitudes like resentment towards her if she fails to comply without excuse, representative authority is the standing to take up *impersonal* reactive attitudes like blame and indignation towards her if she fails to comply without excuse (2013b: 22, cf. Darwall 2006: 67-68).

The two types of second-personal authority are importantly related to one another. In particular, individual authority *presupposes* representative authority (Darwall 2013b: 38-39). To see this, recall the case of Arnold and Bella. Arguably, Arnold *owes* his

⁵⁶ Therefore, a demand is also not rendered legitimate by the mere fact that it is expressed by a person who possesses second-personal authority. It must furthermore qualify *as an expression* of that second-personal authority.

obligation to withdraw his foot *to* Bella. Accordingly, Bella has individual authority to demand that Arnold do so, i.e., she may legitimately adopt the reactive attitude of resentment if he fails to do so. But, as we have seen, Bella's authority over Arnold presupposes Arnold's capacity to determine himself by holding himself accountable. This means that Arnold must be capable of taking up the second-person standpoint vis-à-vis himself and make a demand on himself. However, as I have also noted, a demand is justified from within the second-person standpoint only if it is such that second-personally competent beings, i.e., members of the moral community, *as such* can legitimately hold each other accountable for compliance with it (see Chapter 5, Section 2.3 above). Accordingly, the authority that Arnold exerts over himself when he holds himself accountable is, not individual authority, but the authority of a representative member of the moral community. In short, Bella's individual authority over Arnold must be mirrored by Arnold's representative authority over himself. It follows, quite plausibly, that an act *wrongs* an agent only if it is also wrong *period* (Darwall 2013b: 39).

For our purposes, the important lesson is that the second-personal account does not depict the legitimate demands of individual persons as normative independently of the form of second-personal address. Consequently, this account does not envisage the dignity of persons as a substantive value, or as a set of substantive requirements, that are normative independently of the form of our will. Any claims that persons can legitimately make on us *presuppose* our second-personal competence, and thus our autonomy, as a necessary condition of the possibility of their second-personal authority (cf. Darwall 2006: 242). In this sense, Darwall agrees with Kant that, "I can recognize that I am under obligation to others only insofar as I at the same time put myself under obligation" (Darwall 2006: 248, note 9, *MM* 6:417-418).

In light of this, dignity is a standing that persons are *marked out* as having by the formal principle of our will, i.e., the Categorical Imperative. Persons have dignity *because* we are required to make our actions justifiable to them from within the second-person standpoint, not *vice versa*. And we are required to make our actions justifiable to them in this way, not because they possess a special importance that is prior to the formal principle of our will, but because they, too, are capable of autonomy, and therefore our maxim can be willed to be a universal law only if they can co-legislate it as equal members of a kingdom of ends (see Chapter 5, Section 2.3). Darwall's account therefore confirms Kant's *dictum* that "morality is the condition under which alone a rational being can be

an end in itself; because it is possible only by this to be a legislating member in the kingdom of ends” (Darwall 2006: 102, *G* 4:435).⁵⁷

Contrary to Sensen’s interpretation of the second-personal account, then, this account agrees with his assessment that (2011: 121):

Making a claim does not by itself generate an obligation for the agent. For Kant, the bindingness arises through the qualification of the claim as a universal law, as commanded by the Categorical Imperative.

More generally, the second-personal account is compatible with Sensen’s interpretation of Kant’s account of human dignity (2011: 2):

It is not because others have a value that one should respect them, but it is because one should respect them that they have an importance and a dignity. The justification for the requirement to respect others is a direct command of reason.

But if the second-personal account is compatible with this view of human dignity, one might ask in how far dignity can be considered the “fundamental idea” in Darwall’s framework (2006: 242). I submit that it is fundamental only in the sense that second-personal authority is the notion through which we *encounter* the moral law in our practical consciousness. As I noted above, it is a crucial feature of the second-personal account that it recognises the necessity of basing a vindication of the Kantian analysis of obligation on a practical insight along the lines of Kant’s fact of reason. According to Darwall, this practical insight is an insight into second-personal authority, and hence dignity. But this does not mean that, for Darwall, dignity is *metaphysically prior* to autonomy and, thus, to our will’s being governed by the Categorical Imperative.

As I have pointed out in Chapter 1, the fact that the moral law’s normativity *presupposes* our autonomy does not rule out that our having autonomous wills *grounds* the moral law’s being normative. After all, the logical *symmetry* between autonomy and the normativity of morality does not preclude a metaphysical *asymmetry* (to the effect that autonomy is metaphysically prior to the normativity of morality) between them. To quote Sensen again (2013: 77):

⁵⁷ As Sensen points out, since it is ultimately their capacity for morality, combined with the dignity of the moral law, that marks out persons as possessing dignity, we can distinguish between persons’ “initial dignity”, which they possess insofar as they are autonomous, and their “realized dignity”, which they possess insofar as they act on the moral law and thereby realise their autonomy (2011: 168-169).

[I]f one were never prompted to deliberate morally, there would be no principle of reason. However, once one is so prompted, reason creates the law spontaneously and of its own accord.

In other words, although we cannot derive the normativity of morality from facts about agency or practical reason and, instead, encounter it more directly, through a practical insight, morality's having normative force depends on our engaging in a mode of practical reasoning guided by the notions of autonomy and the Categorical Imperative. To the extent that our encounter with second-personal authority is simply that which facilitates our consciousness of the law that reason creates 'spontaneously and of its own accord', then, the second-personal account is congenial to the Kantian constructivist picture.

2.2 *Rights and Duties*

I have just argued that, on the second-personal account, our encounter with second-personal authority is that which facilitates our consciousness of the moral law. One might object that this conflicts with central tenets of Kantian constructivism or, indeed, any framework that is Kantian in spirit. In particular, it seems to conflict with the Kantian notion that we encounter the moral law, and thus our own autonomy, as an imperative in our own practical consciousness (*MM* 6:239). As Jens Timmermann notes, this notion is what leads Kant to regard duties as prior to rights, both epistemically and metaphysically (Timmermann 2014: 136-137). Timmermann writes (2014: 137):

Rights can be used to make others do what we want them to do. In that sense rights obligate others. But they are made possible by the categorical imperative, which can be enforced only in so far as it concerns the equal external liberty of everyone.

Although the second-personal account is, by and large, not framed in terms of *rights*, the notion that second-personal demands or claims are essential to our encounter with the moral law, and thus our duties, might seem to run counter to the Kantian notion of the primacy of duties vis-à-vis rights (Timmermann 2014: 136).

However, it is important to note that the second-personal account agrees that an encounter with the moral law in our own practical consciousness is necessary to know what we are obligated to do. What the second-personal account adds is that, if our consciousness of being bound by the law is to elicit the Kantian analysis of obligation, it must involve the awareness that we can legitimately be held *accountable* if we fail to comply with the law. This, as I have argued above, is the sense in which second-personal authority is "fundamental" in the second-personal framework (Darwall 2006: 242). It does not

follow, however, that the second-personal account regards claims or demands as *prior* to obligations. Instead, it proposes that the connection between obligation and accountability is part of the very *concept* of moral obligation. Therefore, to deem ourselves under moral obligations is to deem ourselves accountable for complying with them.

Moreover, as I have pointed out in the previous subsection, the conceptual connection between moral obligation and second-personal authority does not imply that the fact-of-reason argument appeals to our awareness of a particular person's *individual* second-personal authority. That is, it does not appeal to the awareness that we *owe* an obligation to a particular person. Indeed, whether one person, *A*, possesses the individual authority to *resent* another person, *B*, when *B* fails to φ without excuse depends on whether every member of the moral community, including *B*, has the representative authority to *blame B* if she fails to φ without excuse. Accordingly, *B* cannot have a bipolar obligation to φ to *A* unless *B* is also under a moral obligation (period) to φ .

We might say, then, that the second-personal account depicts moral obligations as *prior* to individual second-personal authority. Now, the existence of a bipolar obligation is commonly taken to be a necessary condition for the existence of a right in its most paradigmatic form, i.e., that of a *claim-right* (e.g., Hart 1955, Hohfeld 1964, *MM* 6:221-229, Raz 1984). That is, it is commonly assumed that *A* has a (claim-)right that *B* φ only if *B* has an obligation to φ that is owed to *A*. Therefore, by asserting the primacy of moral obligations (period) vis-à-vis individual second-personal authority, which is conceptually tied up with bipolar obligations, the second-personal account also asserts the primacy of obligations vis-à-vis rights. In light of all this, Darwall's remark that autonomy and the Categorical Imperative are "necessary conditions for the possibility of second-personal authority" is actually strikingly similar to Timmermann's remark that rights are "made possible by the categorical imperative" (Darwall 2006: 242, Timmermann 2014: 137).

Similar considerations can be invoked to silence, or at least defuse, scepticism about the second-personal account's ability to accommodate imperfect duties (cf. Timmermann 2014: 133-136). The difference between perfect and imperfect duties, within the Kantian framework, is that while perfect duties *prohibit* all maxims that would have us perform a certain act-type, ψ , imperfect duties *command* the adoption of a certain maxim (*MM* 6:390). As a result, a perfect duty yields one and the same token obligation (not to ψ) under all circumstances. Imperfect duties, by contrast, only provide "grounds of obligation" that need to be applied to the case at hand, alongside other considerations, to determine whether they yield a token obligation and, if yes, which (*MM* 6:224, cf. Timmermann 2014: 134). Following Kant and Timmermann, I am here distinguishing

between “duties”, understood as “general prescriptive laws or rules that provide the matter of what ought to be done” and “obligations”, understood as “individual cases of being morally bound to do something” (Timmermann 2013: 42-43, cf. *MM* 6:222). Timmermann lists three conditions that must be met for an imperfect duty to yield a token obligation: first, the act that it would have us perform must be available (e.g., in the case of the imperfect duty of beneficence, there must be another person in need of my aid); second, the act must be in line with perfect duty; third, the act must not be such that, by performing it, we would neglect another imperfect duty of ours (Timmermann 2014: 134).

Timmermann explains that Kantian imperfect duties “sit uneasily with second-personal claims” because (Timmermann 2014: 134):

[F]or Kant rightful second-personal claims are indigenous to the sphere of strict or juridical duty, at least if moral obligation is meant to have those normative implications that, following Mill, Darwall intends them to have: second-personal accountability, in particular censure, punishment or blame in the case of non-compliance as the fitting reaction to a ‘wrong’.

Timmermann is here referring to Kant’s notion that an obligation is *owed* to a person only if it is enforceable (*MM* 6:231-233). Since imperfect duties need not always yield a token obligation (and when they do, it need not always be the same), they are not enforceable. Therefore, according to this line of reasoning, imperfect duties do not yield bipolar obligations. Adding the premise that an obligation is second-personal in Darwall’s sense only if it is bipolar, it follows that imperfect duties cannot be accommodated within the second-personal framework.

However, as I hope to have shown in the previous subsection, the latter premise is not true. It is not the case that all obligations that are second-personal in Darwall’s sense are, by conceptual entailment, bipolar obligations. The second-personal framework has the conceptual resources to accommodate obligations that are not owed to anyone in particular and yet conceptually tied to the notion of accountability. These obligations would be linked to representative but not individual second-personal authority. Accordingly, an agent would not appropriately be resented for violating such an obligation without excuse. Yet, it would be appropriate to blame her, in the addressive sense of blame implicated by the second-personal framework.

One might respond that even blame is an inappropriate response to a violation of imperfect duty. Indeed, Timmermann argues (2014: 135):

The crucial point is that when I decide to comply with an imperfect obligation I must do so, in a strong sense, of my own accord. That is why I acquire some ‘merit’ with regard to the beneficiary [...], and why he owes me gratitude in return.

Timmermann seems to be raising two points here. The first point is that, while one can act “in conformity with” one’s perfect duties by refraining from some external act, ψ , without doing so from a specific motive (*G* 4:398), compliance with imperfect duty is a matter of volition. This is because, as we have seen above, imperfect duties command the adoption of a certain maxim, and whether an agent has adopted a specific maxim cannot be ‘read off’ her external acts alone. Timmermann seems to suggest that blame is an inappropriate response to an agent’s failure to comply with her imperfect duties because, unlike the omission of some external act, the adoption of a maxim cannot be effected, as it were, from the outside (cf. *MM* 6:220).

However, it is important to note that second-personal blame does not aim to change an agent’s behaviour by appealing to her independent evaluation of actions or outcomes (see Chapter 5 above). That is, it does not function like a threat or temptation that persuades the agent through the prospect of a sanction or reward that she values or disvalues on non-second-personal grounds. Second-personal address is fundamentally different from this ‘goading’ type of persuasion in that it aims to get the agent to change her behaviour by appreciating that this can legitimately be demanded of her from within the second-person standpoint. Thus, second-personal blame does not attempt to sidestep an agent’s autonomy—to the contrary, it attempts to *trigger* that capacity by appealing to the agent’s respect for second-personal authority, and thus, for the law.

Timmermann’s second point is that, if an agent fulfils her imperfect duty (e.g., of beneficence) with respect to some particular person (e.g., by aiding her), then that person owes *gratitude* to the agent. This, according to Timmermann, is a problem for the second-personal account (2014: 135):

Why should anyone be grateful to a benefactor who is merely *responding* to pre-existing second-personal claims, who (what is worse) is accountable for his actions to an external authority and who (worst of all) faces sanctions if he fails to comply?

Now, I have already noted that second-personal blame does not presuppose an authority that is ‘external’ in the sense that Timmermann is implying here. That is, it does not aim

to move the agent by appealing to her independent evaluation of actions or states of affairs. Moreover, I have noted that the second-personal account need not conceptualise imperfect duties as responding to claims of individual second-personal authority. Accordingly, an agent who discharges her imperfect duty with respect to some person need not thereby be responding to claims that this person, considered as an individual, has on her.

In any case, I do not think that the notion that we are accountable for complying with our imperfect duties—be it to individuals or to the moral community—poses a special problem when it comes to accommodating the link between imperfect duties and gratitude. First, it seems that any account that depicts imperfect duties as *duties* in the full sense of that term, i.e., as binding us no less stringently than perfect duties, will be in a similar position. After all, saying that an agent deserves gratitude for doing what she had a duty to do is hardly more plausible than saying that an agent deserves gratitude for doing what she was accountable for doing. Second, it seems that anyone who regards imperfect duties as binding with the full stringency of duty should agree that we are just as blameworthy for violating imperfect duties as we are for violating perfect duties.

Finally, and most importantly, the key premise underlying Timmermann’s argument is not true. In particular, it is not true that an agent is owed gratitude for φ -ing only if she would not be blameworthy for failing to φ without excuse. Suppose that I could save a child’s life at no significant cost to myself. Suppose further that, by doing so, I would not be violating any perfect duty, nor neglecting any imperfect duty. If I failed to help, I would violate my imperfect duty of beneficence. Accordingly, it seems plausible that I would be blameworthy for failing to help without excuse. Even Timmermann should agree with this since he admits that an agent who has a maxim of *never* helping others is “morally vicious” (2014: 135, note 11)—and if I refused to help the child, then, *by hypothesis*, I would have such a maxim.⁵⁸ Moreover, Timmermann agrees that blame

⁵⁸ Timmermann seems to presuppose that ψ -ing is blameworthy only if an agent’s ψ -ing renders her a bad (or vicious) person (2014: 135). This is another indication that Timmermann is not using ‘blame’ in the same sense as Darwall. Within the second-personal framework, “[r]eactive attitudes [...] concern themselves not with a person’s overall agency, but specifically with his conduct with respect to claims or demands that other persons have standing to make of him” (Darwall 2006: 80). In other words, the concept of blame invoked by the second-personal account is an *irreducibly* second-personal concept. Therefore, even if Timmermann is right to note that blame in *his* sense is an inappropriate response to the violation of an imperfect duty, it does not follow that blame in the *second-personal* sense is an inappropriate response.

is the appropriate response to all forms of moral wrongdoing, even violations of duties to the self, which are notoriously unenforceable (2006b: 526). Yet, it also seems appropriate (for the parents of the child, at least) to show gratitude towards me if I save the child.⁵⁹ Hence, the fact that the second-personal account considers gratitude an appropriate response to an agent's φ -ing, where the agent would have been blameworthy for failing to φ without excuse, is not a reason to reject the second-personal account. After all, this verdict seems to be in line with our practices of blame and gratitude. Accordingly, if gratitude is an inappropriate response to such actions at all, then the problem is not rooted in the second-personal account but in those practices.

2.3 *Respect for Persons vs. Respect for the Law*

I have tried to show that the notion that the fact of reason must involve an encounter with second-personal authority is compatible with Kantian constructivism and the primacy of duties vis-à-vis rights. However, one might still be worried about this notion from a Kantian point of view. In particular, one might point out that the moral feeling of respect, which arises from our awareness of the principle of pure practical reason and serves as the motive for action from duty, is respect for the law. And respect for the law, the objection continues, is not to be conflated with respect for persons. By identifying the fact of reason as an encounter with second-personal authority, however, the second-personal account seems to commit this very conflation.

Timmermann provides two major arguments for the claim that the second-personal account runs afoul of the distinction between respect for persons and respect for the law (2014: 138-141). First, he argues that the encounter with the moral law that furnishes the fact of reason and gives rise to the feeling of respect takes place within our own practical reason (2014: 138-140). Timmermann notes that, while it is true that, for Kant, respect is produced when the moral law “infringes upon self-love” and “strikes down self-conceit”, it does not, fundamentally, concern the self's relation vis-à-vis *others* (Timmermann 2014: 139, *CPrR* 5:73, italics deleted). That is, the main problem with

⁵⁹ One might object that it seems appropriate for the parents to show gratitude, but not because gratitude is an apt response to my act of saving the child, morally speaking. Instead, the parents' gratitude might be taken to constitute a—morally unwarranted—overreaction which is understandable given their close relationship to the child. I am inclined to agree that something like this might explain our judgment that the parents' gratitude is appropriate. However, I do not think that this constitutes a problem for the second-personal account. To the contrary, it should make us even more wary of the supposed link between the practices of gratitude and moral blame.

uninfringed self-love or self-conceit is not that it elevates the self over *other persons*, but that it elevates the deliverances of self-love to the status of justified normative judgments, against the verdict of pure practical reason (ibid.). Indeed, the self's relation to others does not play any major role at this point in Kant's moral psychology (Timmermann 2014: 140).

Second, and connectedly, Timmermann argues that the notion of respect for persons is respect of a different kind—and plays a different role—than Darwall's notion of respect for second-personal authority (Timmermann 2014: 140-141). Darwall distinguishes between what he calls “recognition respect” and “appraisal respect” (1977, 2006: 122-126). Appraisal respect, he says, is also referred to as “esteem” and amounts to “an assessment of someone's conduct or character or of something that somehow involves these” (2006: 122). In the case of morality, we respect someone in the appraisal sense when we judge them to be a morally good agent (ibid.). Recognition respect, by contrast, “concerns not how something is to be evaluated or appraised, but how our relations to it are to be regulated or governed” (2006: 123). Relatedly, “[t]he object of recognition respect is not excellence or merit; it is dignity or authority” (ibid.). Now, when Darwall says that the fact-of-reason argument, if it is to be successful, must involve respect for persons' second-personal authority, he uses ‘respect’ in the recognition rather than in the appraisal sense (2006: 122-123). After all, the thought he wants to express is that we ought to relate to persons in a certain way merely on account of their standing as equal members of the moral community, and thus independently of the quality of their conduct or character. However, Timmermann argues that the only way in which respect *for persons* prominently figures in Kant's framework is in the shape of appraisal respect (2014: 140).

Indeed, Kant writes that “[a]ll respect for a person is actually only respect for the law (of righteousness etc.) of which he gives us the example” (*G* 4:401, footnote). From this, Timmermann infers that we respect persons only insofar as we *appraise* them as morally good agents. What is more, the reason why we have appraisal respect for morally good agents is that they represent the moral law to us, which is the only proper object of *recognition* respect (Timmermann 2014: 140). While conceding that Kant does speak of “duties of respect” to other persons, Timmermann argues that these duties are “further down the line in Kant's moral system” (ibid.). After all, these duties merely form one part of his system of duties and are therefore not concerned with the type of respect that serves as the motive of duty in general. Following Sensen, one might add that the attitude of “respect” that these duties command cannot be the motive of duty because the latter

is a feeling, and feelings, according to Kant, cannot be commanded (Sensen 2011: 119, cf. *MM* 6:449).

Nevertheless, I think that it is congenial to Kant's overall framework to interpret the motive of duty as a feeling of recognition respect for second-personal authority. Recall that, while the second-personal account agrees that the feeling of respect arises from the agent's encounter of the moral law within her own practical reason, it asserts that this encounter must involve the agent's awareness of being *accountable* for acting on the moral law (see Chapter 5 above). According to the second-personal account, the latter notion is part of the very concept of moral obligation and plays an essential role in eliciting the Kantian analysis of obligation. As I have pointed out above, the conceptual link between obligation and accountability supports the picture of morality as equal accountability because it implies that "moral norms regulate a community of equal, mutually accountable, free and rational agents as such, and moral obligations are the demands such agents have standing to address to one another and with which they are mutually accountable for complying" (Darwall 2006: 101). It follows that recognition respect for the moral law, as a matter of conceptual entailment, is recognition respect for the equal second-personal authority of all members of the moral community.

It is important to note, however, that this does not mean that the motive of duty reduces to respect for the *individual* second-personal authority of particular persons. Instead, all moral obligations, including those that correspond to someone's individual second-personal authority, presuppose the representative authority of any member of the moral community to hold the agent accountable (see Section 2.1 above). And it is crucial to understand that the moral community, according to the second-personal account, is not a group of concrete individuals. Darwall writes (2007: 64):

[T]he moral community as I understand it is not any actual community composed of actual human beings. It is like Kant's idea of a "realm of ends," a regulative ideal that we employ to make sense of our ethical thought and practice.

To see this, recall that an agent is accountable for complying with some demand only if she can hold *herself* accountable for complying with it from within the second-person standpoint. But this, in turn, implies that she is accountable for complying with the demand only if *any* second-personally competent being as such could hold itself accountable for complying with it. This, as I have tried to show in Chapter 5, evokes Kant's notion of the kingdom of ends (*G* 4:438):

[The dignity of the rational being] brings with it that it must always take its maxims from the point of view of itself, but also at the same time of every other

rational being as legislating (which are therefore also called persons). Now in this way a world of rational beings (*mundus intelligibilis*) as a kingdom of ends is possible, and possible through their own legislation of all persons as members.

By saying that to respect the moral law is to respect the second-personal authority of the moral community, then, I am not saying that the motive of duty comes down to respect for the contingent, empirical demands made by actual human beings (cf. Schaab 2018). Rather, it comes down to respect for autonomy of the will, and thus—given that autonomy amounts to the capacity to co-legislate from a shared second-personal perspective—to respect for the equal dignity of all second-personally competent beings as such.

Indeed, Timmermann admits that “we feel compelled to ascribe to others equal moral status as fellow ‘ends in themselves’, as autonomous beings equipped with transcendental freedom, which marks out persons among all other entities in the world (called ‘things’, *Sachen*)” (2014: 140). I would add that we feel compelled to ascribe the status of ends in themselves to persons, as a matter of conceptual entailment, whenever we have the feeling of recognition respect for the moral law.

Yet, Timmermann seems to think that ascribing the status of “ends in themselves” does not amount to recognition respect for persons, because “Kant does not usually make this point in terms of respect, and it does not rely on relationships with other individuals” (2014: 140-141). However, the first point seems irrelevant to our discussion because, insofar as our attitude to beings who are ends-in-themselves qualifies as recognition respect by Darwall’s definition, we are free to refer to it as ‘respect’ in this specific sense. What is more, as I have just said, it is the feeling of recognition respect for the law that prompts us, by conceptual entailment, to ascribe a different status to ‘persons’ than to ‘things’. Therefore, consistency requires that, even within Kant’s use of terminology, we speak of recognition respect for persons. As to the second point, I hope to have shown that recognition respect for the equal second-personal standing of all members of the moral community does not, as a conceptual matter⁶⁰, “rely on relationships with other individuals” in any objectionable sense either. After all, recognition respect for persons as equal members of the moral community, as I have tried to explain, is not to be conflated with respect for the individual second-personal authority of concrete others. In sum, then, there is no real dichotomy between recognition respect for the moral law and recognition respect for persons.

⁶⁰ Whether this notion relies on relationships with other individuals in another way is a question I take up in the following sections.

3. The Optionality of the Second-Person Standpoint

In Chapter 4, I argued that, even if the first-person standpoint was constituted by acceptance of the Categorical Imperative, the mere *inescapability* of that standpoint, by itself, would not suffice to elicit the Kantian analysis of obligation. Nevertheless, one might think that the notion of inescapability plays an important role in Kantian constructivism. More specifically, one might think that one of the central tenets of Kantian constructivism is that all rational beings, as such, are obligated to abide by the moral law. However, one might worry that the version of Kantian constructivism that I have presented here abandons this notion (cf. Hanisch 2014, Korsgaard 2007, Schapiro 2010).

After all, according to this version, moral obligations have their source in the form of second-personal address. And second personal address does not seem to be inescapable. In principle, a person might avoid making or acknowledging any demands, and thus never take up the second-person standpoint. Of course, it seems difficult to avoid taking up the second-person standpoint, *psychologically* speaking, as we naturally tend to hold each other accountable. Indeed, as Darwall points out, whenever we are addressed second-personally by another person, we more or less automatically enter the second-person perspective, “if only in considering how to respond” (2006: 278). However, as Darwall admits, this response is philosophically unsatisfactory as “the philosophical question is whether we *should* see things in this way [i.e., second-personally]” (ibid., italics added).

In response to this philosophical question, Darwall argues that taking up the second-person standpoint substantially enriches the way we experience our own agency (ibid.). That is, not to take up the second-person standpoint is to miss out on the awareness of a certain class of reasons for action and, hence, on one’s own capacity to act on, and reason in terms of, such reasons. More specifically, Darwall argues that, from the naïve practical perspective, it might seem as though bringing about valuable states of affairs is all there is to agency (Darwall 2006: 280-286). Our grasp of agency changes, however, when we engage in the practice of second-personal address and, consequently, take up the second-person standpoint (Darwall 2006: 287-291). For, in this way, we are made aware of requirements that purport to bind us by their mere form, as well as of our capacity to freely endorse, and self-impose, such requirements. Indeed, Darwall goes as far as to say that, once we take up the second-person standpoint, we learn that the appearances of the naïve practical stance are “an illusion” (2006: 290).

However, as Korsgaard correctly notes, it does not follow that an agent who never takes up the second-person standpoint, and therefore never attains the richer grasp of

her own agency that this standpoint enables, is making a mistake (2007: 20-23). That is, a naïve practical reasoner does not seem to fail by any standard that applies to her when she never comes to reason second-personally. After all, as Darwall himself notes, no non-second-personal reason, i.e., no reason that an agent who does not reason second-personally could grasp, could ever be the right kind of reason to hold an agent accountable for complying with moral obligations (Darwall 2006: 59). Accordingly, even if a naïve practical reasoner somehow *knew* that the second-person standpoint, once taken up, will provide her with a more accurate view of agential capacities, she would not be able to take up the second-person standpoint and hold herself accountable *for that reason*. This is because, as I explained in Chapter 5, second-personal concepts are *irreducibly* second-personal and any non-second-personal reason is the wrong *kind* of reason to employ such concepts. Therefore, Korsgaard is right to object that, “[o]n Darwall’s account, the second-person standpoint does not seem to be unavoidable, the way the standpoint of first-personal deliberation is” (2007: 22).

In more recent work, Darwall questions whether the fact that taking up the second-person standpoint is in principle avoidable means that the obligations that spring from the formal principle of that standpoint are escapable in any problematic sense (2007, 2010). He seems to suggest that the universal authority of the moral law would be endangered only if the *existence* of second-personal reasons for a given agent depended on whether she has taken up the second-person standpoint. And he assures us that “whether someone actually takes up the second-person perspective or not is irrelevant to whether or not second-personal reasons for him actually exist” (2007: 60). All that is required for an agent to be subject to the authority of second-personal reasons, according to Darwall, is that she has “the psychic capacities necessary to take a second-person standpoint” (2010: 256). Accordingly, as long as an agent is second-personally competent, even if she does not *use* her second-personal competence and therefore does not recognize second-personal reasons, she fails by standards that apply to her whenever she fails to act as these reasons direct.

However, this response to the optionality worry is not available to Kantian constructivists. After all, in this view, the form of agents’ practical standpoint is the source of moral obligations’ normativity. And, indeed, Darwall himself states that moral obligations are “grounded” in the second-person standpoint (2007: 59). This suggests that an agent who does not take up the relevant practical standpoint is not morally obligated. To quote Sensen again, “if one were never prompted to deliberate morally, there would be no principle of reason” (2013: 77). Accordingly, it is incompatible with Kantian

constructivism to assert that agents are subject to moral obligations independently of the practical standpoint they happen to take up.

I agree with Korsgaard's assessment, then, that the second-personal account does not render moral obligations inescapable in the sense in which the first-personal account aspires to do. Contrary to what Korsgaard seems to suggest, however, I do not think that this is a serious problem, even from a Kantian constructivist perspective. One might object that any account that renders it a contingent matter whether an agent is under moral obligations thereby departs from the Kantian tradition. After all, it is central to Kant's moral theory that morality binds *unconditionally* and *necessitates* us (*G* 4:413 and 4:416). However, we should distinguish the modality of the *way* in which the moral law binds us from the modality of *whether* it binds us. From the fact that, *as* moral agents, we recognize the moral law as commanding with necessity it does not follow that, necessarily, we *are* moral agents. Indeed, Kant quite clearly remarks that unless one possesses the capacities of "moral feeling", "conscience" and "respect for the law", by which we "recognize duties", one cannot be bound by morality (*CPrR* 6:399-402). And it is possible to conceive of a human being—a member of the biological species, that is—who possesses some capacity for practical reason but not pure (i.e., moral) practical reason (*Rel* 6:26-28, also see Chapter 4, Section 3.2). Therefore, no human being, *qua* member of the species, is *necessarily* under moral obligations.

What is more, as Korsgaard concedes in *Sources*, even the standpoint of merely first-personal deliberation is not strictly inescapable since one could in principle regard nothing as a normative reason for action, and thus avoid practical deliberation altogether (1996: 161-164). Since, in Korsgaard's view, we adopt practical identities and constitute ourselves as an agent by choosing to value ourselves under certain descriptions (see Chapter 4 above), she holds that "[t]he most straightforward expression of complete practical normative skepticism would [...] be a form of suicide" (1996: 161). She writes (1996: 163):

Is there an argument against that kind of skepticism, a reason not to commit suicide? In one sense, the right reply is that there can't be, since life itself is the source of reasons. In another sense, the right reply is that this is not an issue to be settled by philosophical argument alone.

I agree with Korsgaard that philosophical argument alone cannot convince an agent to take up the first-personal standpoint of agency. But why, then, would the second-personal account be under pressure to provide a philosophical argument for taking up the second-person standpoint?

In fact, I would go so far as to say that there cannot be *any* argument as to whether or not to take up the second-person standpoint. This is because taking up the second-person standpoint is not something that one can *choose*, deliberately, to do at all. To see this, consider what it would mean to deliberate about taking up the second-person standpoint. It would be a matter of trying to determine whether one has reason to apply second-personal concepts such as blame, demand, or accountability to a given situation. However, at this point, the matter is already settled. Either one considers it apt to apply second-personal concepts to the situation, which, given the nature of these concepts, means that one judges irreducibly second-personal features of the situation to render the applications of these concepts apt, or one does not consider it apt to apply these concepts. Now, in the former case, one has already taken up the second-person standpoint. And, since the second-person standpoint is a *normative* standpoint, in the sense that the form of this standpoint is the source of obligations that are normative for anyone who takes up that standpoint, this means that it is no longer an option (rationally, at least) to abandon the second-person standpoint. However, if one does *not* consider irreducibly second-personal features of the situation to warrant the application of second-personal concepts, then one cannot *decide* to take up the second-person standpoint either. After all, any *other* reason to apply second-personal concepts would be the wrong *kind* of reason to apply them. In fact, even if one tried using second-personal concepts for such a reason, one would not succeed. One would not be reasoning second-personally and therefore fail to take up the second-person standpoint.

For this reason, it is misleading to say that the second-person standpoint is ‘optional’, ‘avoidable’ or ‘escapable’. After all, taking up the second-person standpoint is not optional in the sense of something that one can choose to do or not to do. Nor is it possible deliberately to avoid or escape taking up the second-person standpoint. It would be more accurate to say that it is not necessary, conceptually or metaphysically, that a human being take up the second-person standpoint. But, as I have argued, the same is true of the first-person standpoint of agency. Why, then, does Korsgaard think that there is a special problem of ‘avoidability’ for the second-personal account?

Korsgaard might argue that the first-personal account is in a superior dialectical position because, although it does not have an argument to convince the complete *normative* sceptic, it has one to convince the *moral* sceptic—the agent who acknowledges *some* reasons for action but refuses to act morally—of taking morality seriously (1996: 163). After all, the first-personal account purports to motivate the Kantian analysis of obligation from within the first-person standpoint that we all take up insofar as we acknowledge *any* reasons for action. The second-personal account, by contrast, purports

to motivate the Kantian analysis from within the second-person standpoint, and thus from within a perspective that is taken up only by those who acknowledge a certain *kind* of reasons for action: authority-regarding reasons.

However, this dialectical advantage is insignificant. First, it does not render the first-personal account the ‘more Kantian’ account since, as I have argued above, the Kantian tradition is not centred around the ambition of showing that all *agents*—regardless of whether they engage in moral reasoning—implicitly accept moral principles.

Second, I doubt that there is a significant number of—or, indeed, that there are *any*—agents who take up the merely first-personal standpoint of agency but not the second-person standpoint.⁶¹ In fact, it seems likely that the isolation or mental illness necessary to prevent a human being from taking up the second-person standpoint would be sufficient to prevent her from taking up *any* practical standpoint whatsoever. After all, a host of human interactions, e.g., with our parents, seem necessary to instil in us even the minimal agential capacities that are necessary to take up, say, the naïve practical perspective. And it would be surprising if someone who interacted with others in this way and managed, as a result, to take up the naïve practical perspective did not also, at some point along the way, take up the second-person standpoint, if only to blame or resent others. Therefore, if there are human beings who do not take up the second-person standpoint, and their existence poses a problem for the second-personal account, then the first-personal account faces the same problem.

Finally, even if there are agents who take up a minimal practical standpoint, such as the naïve practical perspective, but not the second-person standpoint, I suspect that we would not ordinarily regard them as subject to moral obligations. I submit that an agent who was not sensitive to the legitimate demands of persons (including her own) at all—someone who never even formed reactive attitudes such as blame, guilt, and resentment—would not strike us as ‘one of us’. While neither human biology nor the constraints of agency as such require that we take up the second-person standpoint, a human agent who never took up that standpoint would nevertheless appear *inhuman* to us.

⁶¹ David Shoemaker argues that psychopaths might fail to take up the second-person standpoint while managing to take up a merely first-personal stance (2011). Unfortunately, I do not have the space necessary for a detailed discussion of his account. Suffice it to say that Shoemaker’s account is not the only one compatible with the empirical data on psychopaths. In any case, Shoemaker does not think that psychopaths can be held morally responsible. Accordingly, his account is congenial to my overall line of argument in this section.

4. Second-Personal Address, the Fact of Reason, and Duties to the Self

I have argued that we need not be troubled by Korsgaard's objection to the second-personal account if it is understood as the charge that the second-personal account renders moral obligations escapable or avoidable. But I suspect that Korsgaard's objection is in fact driven by a different, deeper worry, namely that the second-person standpoint hinges upon the *social* practice of second-personal address, and therefore cannot form a central component of (a Kantian account of) practical reason.

It might seem that if morality is ultimately grounded in a mode of interaction with others, its credentials are not accessible to each individual rational being's quiet reflection upon her own agency. The worry would then be, not that one might in principle fail to take up the practical standpoint that grounds morality, but that one cannot—even in principle—take up this standpoint unless one is initiated into a certain social practice. Strictly speaking, then, Korsgaard's worry is not that Darwall's account renders morality *avoidable*, but that it renders morality *unavailable* to some agents.

This worry, unlike the worry of optionality, might indeed be described as Kantian in spirit. After all, one tenet of the Kantian tradition seems to be that every individual that possesses the psychological capacities necessary for moral agency can potentially access the requirements of morality and thereby make herself aware of her capacity to follow these requirements, i.e., her autonomy. In other words, every such individual can encounter the moral law, and thus experience the 'fact of reason', within her own practical deliberation, independently of her interactions with other people. If the fact of reason must be interpreted second-personally, however, then it may seem as if individuals who are not part of society, or in any case do not participate in the social practice of making and acknowledging demands on one another, are denied access to the moral law.

Now, one might respond that this worry can be resolved by recalling that, on the second-personal account, the fact of reason is not facilitated—or, at least, not directly—by concrete others' *individual* second-personal authority over us. Instead, our encounter with the moral law is facilitated immediately by the *representative* authority that we, as an equal member of the moral community, exercise over ourselves (see Section 2.1 above). What is more, representative authority is prior, metaphysically and conceptually, to individual authority because someone's individual authority to demand that I ϕ depends on my (and everyone else's) representative authority to demand that I ϕ . Accordingly, it seems possible, metaphysically and conceptually, that I take up the second-person standpoint, and thus encounter the moral law in my practical consciousness, independently of my interaction with a concrete other person.

In particular, it seems that I could address myself second-personally, demanding that I comply with my moral obligations, just by invoking my representative second-personal authority as a member of the moral community. This seems particularly plausible on the assumption that there are instances of moral obligation *period*, i.e., moral obligations that are not owed to another person. For, if there are such obligations, I have exactly the same authority to address these obligations by holding myself accountable for complying with them as anyone else: the representative second-personal authority of a member of the moral community. But even if there are no such obligations and, instead, all our moral obligations are owed to particular others, it seems that I will be able to take up the second-person standpoint and hold myself to these obligations independently of being addressed by those to whom they are owed.

However, while representative second-personal authority is metaphysically and conceptually prior to individual second-personal authority, it might still seem that engaging in second-personal address with others is the *paradigmatic* way of taking up the second-person standpoint. This reading of Darwall's position seems quite common in the literature (cf. Hanisch 2014, Pauer-Studer 2010, Stern 2014, Timmermann 2014, Wallace 2007, Zylberman 2017). Indeed, Darwall himself describes the second-person standpoint as "the perspective you and I take up when we make and acknowledge claims on *one another's* conduct and will" (2006: 3, italics added). Even when he discusses the worry that the second-person standpoint might be optional he remarks that demands "from *another* free and rational agent will usually be enough to set us to thinking second-personally" (2006: 278, italics added). Thus, he does not even seem to consider the possibility of taking up the second-person standpoint in isolation from other second-personally competent agents.

Accordingly, although it is possible to take up the second-person standpoint by addressing oneself by invoking one's representative authority, this way of taking up the second-person standpoint might strike many readers of Darwall's work as non-paradigmatic, perhaps even derivative. The reason for this—and for Darwall's striking omission of that possibility in his discussion—might be that there appears to be an *asymmetry*, within the second-personal framework, between the interpersonal and the intrapersonal realm. In particular, while there are certainly obligations that we owe to concrete others, and there are possibly obligations that we do not owe to anyone in particular, there is no indication—at least, within Darwall's work—that there are obligations that we owe to *ourselves*.

Admittedly, Darwall acknowledges, following Thomas Hill, that "[g]iving little weight to one's own wishes and values, by being inappropriately deferential to those of

others, can be [...] a failure to respect ourselves” (Darwall 2013e: 121, cf. Hill 1973, Darwall 2013d). However, as Hill explicitly notes in his discussion of this topic, “[t]he duty to avoid servility is a duty to take a certain stance *towards others* and hence would be inapplicable if one were isolated on a desert island” (1973: 103, italics added). So, the relation to ourselves that underlies the duty to avoid servility, on Hill’s account, does not generate any legitimate demands on us unless we also stand in certain relations to *others*. Hence, while this duty is in a sense *self-regarding*, it is not *owed* to ourselves in the sense in which bipolar obligations to others are owed to other agents. After all, the duty against servility does not seem to be based on our *individual* second-personal authority over ourselves (if it was, it would be applicable even if we were isolated on a desert island). Instead, this duty seems to be due to our *representative* authority’s mediating and limiting the demands that *others* can legitimately make on us.

I should also note that, since the second-personal account confirms Kant’s *dictum* that “I can recognize that I am under obligation to others only insofar as I at the same time put myself under obligation”, there is a sense in which this account holds that *all* our obligations are obligations to the self (Darwall 2006: 248, note 9, *MM* 6:417-418). However, this is not the sense of ‘obligations to the self’ that is at issue here. After all, the notion that all obligations are obligations to the self *in this sense* does not remove the above-mentioned asymmetry between interpersonal and intrapersonal second-personal relations. In fact, to say that all obligations are obligations to the self *in this sense* is just another way of saying that representative second-personal authority is prior to individual second-personal authority. And we have already seen that the latter notion does not suffice to remove the apparent asymmetry between intrapersonal and interpersonal second-personal address.

It is worth noting that, even if it does not support the notion that interpersonal second-personal address is the paradigmatic way of taking up the second-person standpoint, the apparent lack of obligations owed to the self in the second-personal framework is worth discussing from a Kantian constructivist perspective. After all, the existence of obligations to oneself, and the resulting symmetry between the interpersonal and the intrapersonal realm, seems to be a central tenet of the Kantian tradition. Hence, if the second-personal account was unable to accommodate obligations owed to oneself, this might limit its ability to support a Kantian picture of morality (cf. Timmermann 2014: 131-132). I shall take up this worry, along with the worry that the second-person standpoint might be unavailable to individuals who do not engage in the social practice of second-personal address, in the next chapter.

4. Conclusion

I have argued that the second-personal account does not undermine the Kantian constructivist project. My main line of argument has been that the second-personal account does not deny that the Kantian analysis of obligation depends on an encounter of the moral law within our own practical reason. The second-personal account does not, then, claim that moral obligation has its source in our actual interactions with particular others. Instead, the second-personal authority of particular persons is mediated through, and conferred by, our own capacity to be motivated by the lawful form as such, i.e., our autonomy.

Furthermore, although I have conceded that the second-personal account, in some sense, renders moral obligation avoidable insofar as one could in principle avoid engaging in second-personal address, I have argued that this does not mean that the second-personal account has a significant dialectical disadvantage vis-à-vis the first-personal account. However, I have noted that one might read Korsgaard's objection, not as the complaint that the second-personal account renders moral obligation avoidable, but as the complaint that the second-personal account bars isolated individuals from access to moral obligations. And this complaint, in turn, seems to be related to the complaint that the second-personal account might not be able to accommodate obligations that are owed to oneself. I take up these connected complaints in the next chapter.

Chapter 7: Commitment and the Second-Person Standpoint

1. Introduction

In the previous chapter, we have seen that Korsgaard's objection might be underwritten by the worry that the second-personal account renders morality unavailable to isolated individuals by tying it to the essentially social practice of second-personal address. In this chapter, I want to put this worry to rest by arguing that second-personal address is not, fundamentally, a social practice but a mode of practical reasoning that extends to other-regarding as well as purely self-regarding contexts. Accordingly, the second-person standpoint is taken up not only in interpersonal interactions, but also in an agent's engagement with her own will.

In order to establish this conclusion, I argue that we address ourselves second-personally whenever we make a *commitment* to φ (e.g., to pursue some project or maintain some relationship). The second-personal account of commitment that I propose characterises commitments as obligations to oneself. Accordingly, the argument of this chapter establishes, not only that the second-person standpoint can be taken up in purely self-regarding reasoning, but also that the second-personal framework can accommodate obligations to oneself.

In Section 2, I introduce Chang's voluntarist account of commitments, according to which, by committing to φ , we employ the normative powers of our will to give ourselves a reason to φ which we would otherwise not have. In Section 3, I argue that, by itself, Chang's account does not have sufficient resources to reconcile the volitional nature of commitments with their normative significance. In order to remedy this problem, we should conceive of commitments of instances of second-personal address vis-à-vis ourselves. Section 4 lays out two important implications of this argument: that commitments are obligations to oneself and that there are moral constraints on commitment. In Section 5, I show how the argument of this chapter dispels the worry that the second-person standpoint might be inaccessible to isolated individuals and how, in conjunction with some remarks from the previous chapter, it suggests a more accurate understanding of the second-personal framework in general. In Section 6, I explain how my remarks on commitments relate to Korsgaard's notion of practical identities.

2. Chang on Commitment

Practical philosophers are increasingly interested in commitment as a normative relation (e.g., Chang 2013b, Ebels-Duggan forthcoming, Liberman and Schroeder 2016, Shpall 2013, 2014, Wallace 2001). One lesson that can be drawn from recent work on this topic

is that we can distinguish at least three types of commitment. First, having a certain attitude might commit us to forming some other attitude: my belief that I ought to φ commits me to intending to φ ; my belief that p , and that p implies q , commits me to believing that q . We might refer to this type of commitment as “rational commitment” (Shpall 2013, 2014). Second, we might be committed to φ by virtue of having promised someone that we would φ . This is a paradigmatic example of what we might call “moral commitment” (Chang 2013b: 76-79, Shpall 2014). Third, we might be committed to φ by being *dedicated* to φ . This type of commitment is implicated in utterances like ‘I am committed to pursuing a career in philosophy’. We might refer to this type of commitment as “internal commitment” (Chang 2013b: 76-79). Since I am here concerned with this latter type of commitment, I shall use ‘commitment’ to refer to internal commitment, except where I explicitly say otherwise.

2.1 *The Volitional Character of Commitment*

Chang proposes a thoroughly voluntarist account of commitments, according to which they “essentially involve volitional activities” (2013b: 93, cf. 2009, 2013a, 2015). On this account, when we make a commitment to φ we employ the normative powers of our will to give ourselves a reason to φ that we might otherwise not have had. Chang refers to the reasons that we create in this way as “voluntarist reasons” (2009, 2013b).

Importantly, for Chang, committing is not merely a matter of deciding to φ , or of responding to some reason-giving aspect of φ -ing, e.g., its value (2013b: 79-81). Indeed, the willing involved in committing to φ does not even take the act of φ -ing as its sole object. Instead, the volition involved in commitment is also concerned with our *reasons* to φ —and, in particular, with altering and supplementing these reasons. Specifically, when we commit to φ , we *will*, perhaps unconsciously, that some fact about φ -ing *be a reason* for us to φ (Chang 2013b: 93). In Chang’s words, it is a matter of “placing your will—your very agency—behind its being a reason” (2013b: 93). For example, by committing to a career in philosophy, I will that some fact about such a career have a normative, reason-giving significance that it otherwise would not have.

Chang is right to characterise commitment as an essentially volitional phenomenon. What is more, the volitional act that Chang identifies as central to commitment is uniquely suited to explain the normative difference between projects and relationships to which we are committed and those to which we are not committed. Let me highlight this difference with two examples.

First, imagine that whenever I have friends over for dinner, I save the cork of each bottle of wine we open. I am not particularly invested in this activity. In other words, I

am not committed to it. Rather, I acquired the habit when, one night, I incidentally forgot to throw out the used corks. After a while, a considerable number of corks have accumulated. However, one day on my way to the bin I spot the corks in a bag in the corner of my kitchen and, on a whim, pick them up and throw them out.

Next, imagine that I plan to go on a backpacking trip around the globe. Since this trip will cost some money, I start saving for it by putting a five-pound note into a jar on my kitchen table every morning. I am fully dedicated, or committed, to this project. As a result, I spend my days and nights thinking and dreaming about the adventures I might have on my trip around the world—the people I might meet and the places I might see. My bedroom walls are covered with photos of beaches in Indonesia and maps of trekking routes in the mountains of Peru. In fact, I view this project as connected to my identity, one of the things I bring up when someone asks me questions about my personality, say, in a job interview or at a party. After a few years of saving, a considerable sum of money has accumulated in my jar—almost enough to pay for the entire trip. Yet, one day on my way out of the kitchen, I spot the jar on the table and, on a whim, take out the money and, instead of using it to fund my trip, spend it on a new car.

There seems to be a considerable difference between the normative status of my behaviors in the two cases. In particular, while it is entirely unproblematic to throw out the accumulated corks in the first example, something goes wrong, at least *prima facie*, when I abandon my original project in the second example.⁶² What is more, the problem does not seem to be with deciding to spend the money on something other than travelling *per se*.⁶³ Instead, I suggest, the problem is that I abandon my travelling plans *on a whim*. It seems that, given how *invested* I am in the travelling project, I should not just drop it

⁶² I qualify the second claim as one that holds *prima facie* because, first, there might be constraints on what we can commit to (accordingly, some of our commitments—or would-be commitments—might not make a difference to our normative situation) and, second, there might be weighty reasons to go against an existing, normatively significant commitment (thus, that I have committed to φ need not be sufficient for my not- φ -ing to be problematic). I discuss these qualifications in some detail below.

⁶³ One might argue that the judgment that there is something wrong with spending the money on a car is due to the fact that a car does not possess sufficient value. Or, in fact, one might argue that it possesses a negative value on account of its harmfulness to the environment and its morally questionable supply chain. However, I contend that we would find something wrong with dropping the travelling plans on a whim regardless of what the money is spent on instead. We can thus regard ‘a new car’ as a place-holder. Fill in any valuable objects you like and dropping the project on a mere whim will still seem problematic (or so I predict).

without giving the decision some thought. Again, it is unproblematic to *change my mind*, i.e., to be swayed by considerations to the effect that investing my money in a new car is more worthwhile than spending it on travels. But our judgment seems to be different if I simply fail to carry out my project, without reconsidering whether or not my investment in the project was a good idea. The crucial normative difference between the two examples seems to be that I am *committed* to travelling around the globe but not to collecting corks. And as long as I *remain* committed to the travelling project, i.e., as long as I do not reconsider and terminate my commitment, it remains problematic to abandon the project.

Now, there are several ways in which one might try to explain the difference between these two cases—and not all of them characterize commitment as volitional, let alone in the terms Chang proposes. However, Chang persuasively argues that none of these alternative accounts manage to capture the phenomenon of commitment. Here, I can only provide a brief sketch of her arguments.

One candidate explanation suggests that someone is committed to a project or relationship if she is in a special *epistemic* position towards this project or relationship (Chang 2013b: 81-83). More specifically, to be committed to φ is to hold certain *normative beliefs* about φ -ing. I shall refer to this as ‘the normative belief account’. On the normative belief account, the reason why not everyone is equally committed to, say, travelling is that not everyone has access to the same information as committed travelers. That is, committed travelers *recognize* reasons to pursue travelling to which others are blind. However, Chang persuasively rejects the normative belief account (ibid.). The main objection she raises is that commitment is *discretionary*. That is, even if we have full information about a project or relationship, it should—to a certain extent, at least⁶⁴—be ‘up to us’ whether we commit to it or not. Normative belief, by contrast, does not seem to be the kind of thing that we can discretionally decide to have, independently of our epistemic position. Rather, given an equal epistemic position, two agents should hold the same beliefs about what normative reasons there are to pursue some project, insofar as they are both rational.

The normative belief account is perhaps most plausible as an account of the commitments involved in love relationships. Kyla Ebels-Duggan argues that, while one does not usually fall in love with a person for reasons that one can publicly articulate and defend, “one normally comes to love someone through experience of him, experience in which one appreciates directly the value that one affirms in loving him” (forthcoming:

⁶⁴ I say more about the constraints of commitment below.

3). In Ebels-Duggan's words, falling in love with a person involves a "direct appreciation" of that person, resulting in the realization that the person in question is valuable along some dimension which has hitherto gone unnoticed and cannot be communicated to those who have not had this experience (forthcoming: 1). Indeed, this account seems to make sense of sentences like 'I don't know what you *see* in him'. Yet, even here Chang's objections seem to apply. The normative belief account seems to imply that everyone has equal reason to enter a love relationship with *X* if everyone is in an equal epistemic position with respect to *X*, where this includes having had the same transformative experiences of *X*. However, how much reason we have to enter or maintain such a relationship seems to be at least partly a function of our own *choices*.

In addition to falling prey to Chang's objection, the normative beliefs account does not accurately capture the phenomenology of love relationships. When we are in love with someone we do not view ourselves as occupying a position which anyone else could, in principle, occupy as well. More specifically, we take ourselves to make a constitutive *contribution* to the love relationship, and we take this contribution to at least partly explain the reasons we have to maintain the relationship. On normative belief accounts, however, we are mere passive bystanders, bearing witness of the value of (our relationship with) *X* without making much of a difference to this value. Even though a direct appreciation might make us stand out from the crowd by enabling us to recognize reasons for being in a love relationship with *X* which are *incommunicable* or *inarticulable* (cf. Ebels-Duggan forthcoming), it still constitutes a *recognition* of a value that is already there, in the world, independently of our appreciation of it.

Chang also discusses potential explanations of commitment in terms of desires, endorsements of desires, intentions, and policies. Her rejections of all these proposals rest on what I take to be essentially the same point: instead of explaining the normative force of commitment, i.e., its potential to give us reasons, they presuppose that these reasons are already in place (2013b: 83-92).

With regard to desires, Chang argues that desiring to φ does not seem to explain the special reasons to φ that we have once we commit to φ . Our commitment to *X* will typically *give rise* to strong desires to φ , but not the other way around (2013b: 83-84). I might add that the difference between my project of collecting corks and my project of travelling around the globe also does not seem to be due to a difference between my desires vis-à-vis these projects. While it is plausible to suppose that my desire to travel is stronger than my desire to collect corks, this does not explain why it is problematic to abandon the travelling project *on a whim* whereas this seems entirely unproblematic in the case of the cork-collecting project. After all, while the difference between my strength

of desire for the two projects is a difference of degree, the normative difference between the two projects is a difference in kind.

Against the proposal that commitments might consist in endorsements of—or identifications with—certain desires that we have, Chang contends that desires are not the proper object of commitment. Our commitments are importantly *to* φ , not to our own, *antecedent* attitudes towards φ -ing. Again, certain attitudes are commonly among the *upshots* of our commitments, but they do not serve as their source or object (2013b: 84-86). Indeed, I would argue that, when I abandon my travelling project on a whim, the problem is not the way I relate to my own desires but the way I relate to the project itself.

Against the view that commitments are reducible to intentions, Chang raises the objection that, while we typically form intentions to do various things as a result of having committed to φ , our commitment to φ does not itself seem to be captured by these intentions (2013b: 86-87). And, again, it does not seem plausible to me that the difference between my two projects—particularly the fact that it seems problematic to abandon one of them, but not the other, on a whim—is explained by a difference between my intentions vis-à-vis these projects.

Finally, the view that commitments are essentially policies to treat certain considerations as reasons for action in one's deliberation, according to Chang, falls prey to the objection that treating something *as if* it is a reason does not by itself make it a reason. The *normative* force of commitment can thus not be explained by adopting such a policy but must instead be what *underwrites* such a policy (2013b: 87-92).

In contrast with these alternative proposals, Chang's account promises neatly to capture the difference between the project of collecting corks and that of travelling around the globe. Chang's notion of *placing one's will or agency* behind something's being a reason is congenial to the notions of being invested in and dedicated to a project that we naturally associate with commitments (cf. Chang 2013b: 93). Indeed, it is plausible that, in the example of the travelling project, I have willed that there be (additional) reason for me to pursue this project. In particular, by committing to pursue this project, I have willed that my *normative* relation to this project change—that it play a role in determining what I should do which it otherwise would not play. Connectedly, it is plausible that the volitional act of willing that there be reason is precisely what is missing in the case of collecting corks. While my behavior in this case is superficially quite similar to my behavior with respect to the travelling project, I have not conferred any special normative significance onto the project of collecting corks. In other words, I have never willed that the cork-collecting project, as opposed to the travelling project, change the normative structure of my deliberative situation.

Besides explaining the difference between the two projects, Chang's proposal also promises to account for the difference between dropping my traveling project *on a whim* and abandoning it after *changing my mind* about it. I have noted that, *prima facie*, the former is problematic while the latter is not. Now, Chang's proposal promises to account for the problematic nature of the former because willing that there be reason to φ is different from deciding to φ . Although we *can* will that there be a reason to φ by deciding to do so, we can also do so without much explicit deliberation (Chang 2013b: 96). Connectedly, it seems natural to think that we can decide *not* to φ although we have willed—or, indeed, are still willing—that there be reason to φ . In this way, Chang's account promises to allow for the possibility of deciding *against* our existing commitments, which is precisely what happens when I abandon the travel project on a mere whim. Whether Chang's account actually allows for this possibility—and thus makes good on its promise to explain the problematic nature of dropping a commitment on a mere whim—will be the subject of my discussion in Section 3 below.

Chang's proposal also explains why it seems unproblematic to abandon my travelling project after changing my mind about it because, as she puts it, “[i]f commitment is a matter of will, then so too is ‘uncommitment’” (2013b: 94). In other words, Chang's proposal implies that, just as we can make commitments, thus creating new reasons, by an act of will, so we can withdraw our commitments or ‘uncommit’, thus terminating the corresponding reasons, by another act of will. This explains why abandoning my travelling project is unproblematic once I have changed my mind because ‘changing my mind’, in the relevant sense, can plausibly be interpreted as referring to the act of willing that my voluntarist reason to pursue the project be terminated. In sum, Chang's account states that, while it is problematic to drop a committed project on a whim because to do so is to decide not to φ despite having willed that there be a voluntarist reason to φ , it is unproblematic to abandon a project after changing one's mind about it because to do so is to decide not to φ after ‘uncommitting’ from the project, i.e., after ceasing to put one's will or agency behind there being a voluntarist reason to φ .

2.2 Chang's Brand of Voluntarism

Before I turn to the main argument of this chapter, it is important to highlight some central features of Chang's voluntarist account of commitments. To begin with, note that Chang is not suggesting that our commitments trigger some independent normative principle according to which committing to φ is a sufficient condition for our having a reason to φ (2013b: 97-103). While this suggestion would grant that our commitment plays a role in explaining the emergence of our reason to φ , the normativity of that reason

would ultimately be underpinned by the normative principle that assigns this role to our commitment in the first place. By contrast, Chang believes that the volitional act involved in our commitment to φ itself is *that in virtue of which* there is a normative reason for us to φ , i.e., the *source* of this reason's normativity.

Against the view that commitments give rise to reasons only by triggering an independent normative principle, Chang argues that no such principle could plausibly exist (2013b: 97-100). She starts by reminding us that the volition involved in committing to φ is a *willing that something be a reason to φ* . She then points out that anyone who engages in this volition thereby presupposes that she has the power to create a reason to φ by willing that there be such a reason (Chang 2013b: 97). This presupposition, however, would turn out to be false if the source of the normativity of commitments were located, not in our own will, but in an independent normative principle. But since this principle would still demand that you engage in this volition in order to create a reason by making a commitment, it would effectively demand that you engage in an activity which presupposes that the principle is *not* the source of commitments' normativity. In Chang's words, "the principle holds that in order to have a reason that is, by hypothesis, a reason in virtue of something other than your willing, you have to engage in an activity that falsely presupposes that your willing is that in virtue of which it is a reason" (2013b: 99). In short, the account on which the source of the normativity of commitments lies in an independent normative principle contradicts the presuppositions of the very volition that one would have to perform in order to trigger that principle. Accordingly, Chang objects (2013b: 99):

[O]n the suggested view, no clear-eyed, ideally rational agent could make commitments. For no ideally rational agent could go in for the irrational activity of creating reasons when, by hypothesis, reasons are not the sorts of things that can be created.

Therefore, Chang concludes, the volition involved in commitments must be the *source* of the normativity of our voluntarist reasons.

It is also important to note that Chang explicitly distinguishes internal commitments from *moral* commitments (2013b: 76-79). She argues that moral commitments typically involve other persons and require their *uptake* if they are to exert any normative force.⁶⁵ For example, the commitment involved in promising requires another person

⁶⁵ Of course, Chang does not say that all moral *requirements* involve uptake. She only says this of those moral requirements that are incurred by what we would ordinarily regard as a *commitment*.

who accepts the promise and relies upon its fulfillment. By contrast, internal commitments essentially only involve a single person, willing that something be a reason for her. Therefore, according to Chang, the normative force of internal commitments cannot be explained by our obligations to others.⁶⁶ I discuss Chang's way of drawing the distinction between moral and internal commitments below.

Chang contends that her account vindicates "what is arguably Kant's deepest insight: that the will can be a source of normativity" (2013b: 104, cf. *G* 4:431). As opposed to Kant, however, Chang does not think that there is a close connection between the normative powers of our will and morality (2013b: 76-79). In particular, she does not think that there is a purely *formal* principle of willing that places constraints on the content of our voluntarist reasons (2013b: 108). Accordingly, she thinks that we could *in principle* give ourselves voluntarist reasons for almost anything, including immoral actions (2013a: 184, 2013b: 109).⁶⁷

However, this does not mean that we will ever have all-things-considered reason to act immorally. This is because, on Chang's view, voluntarist reasons exist, and influence what we should do, all things considered, only in a very limited range of cases. To understand this, it is important to note that Chang does not think that *all* our reasons are voluntarist reasons; some of our reasons are "given" independently of what we will (2013b: 104). Thus, while adopting a non-Kantian form of source *voluntarism* about the reasons created by commitments, Chang is a source *externalist* with respect to all other reasons.

On Chang's "hybrid voluntarist" account, we can create *voluntarist* reasons, and in this way influence what we have all-things-considered reason to do, only if our *given* reasons have "run out" (*ibid.*). And our given reasons have run out if and only if, in a decision between two options, φ and ψ , either (1) φ and ψ are in "equipose" (i.e., the given reasons for φ do not outweigh the given reasons for ψ , the given reasons for ψ do not outweigh given reasons for φ , and neither are ψ and φ equally supported by given reasons) or (2) the given reasons favor one of the two options, but it is indeterminate how strongly they favor it (2013b: 103-110). Never will voluntarist reasons switch the "valence" of what we have all-things-considered reason to do (2013b: 105). That is,

⁶⁶ Note that internal commitment can nevertheless be a commitment *to* (one's relationship with) another person (Chang 2013b: 76). What distinguishes it from moral commitment is that it does not require uptake by that person.

⁶⁷ I say, 'almost anything', because Chang acknowledges "logical" constraints on commitment (2013b: note 39).

voluntarist reasons will never make it the case that we have all-things-considered reason to φ in a situation where the given reasons favor ψ -ing. On the assumption that our given reasons are valenced against acting immorally, then, voluntarist reasons cannot make it the case that we have all-things-considered reason to act immorally (Chang 2013a: 184).

My main objection to Chang's account will not directly be concerned with the hybrid character of her view. However, I want to note that this hybrid character leaves her view vulnerable to at least two distinct objections. It turns out that my proposed remedy to the main problem with Chang's account—that it does not have the conceptual resources to reconcile the normativity of commitments with their volitional source—also avoids these two objections. The two objections are as follows.

First, hybrid voluntarism is in line with our moral intuitions only if we think that positing the existence of voluntarist reasons for immoral actions does not impugn a metanormative view provided that these reasons do not determine what we should do, *all things considered*. After all, according to Chang, whenever the strength with which our given reasons favor a morally permissible option over a morally impermissible one is indeterminate, we can give ourselves a voluntarist reason for the impermissible option, thus narrowing the normative gap between it and the permissible one. Hence, if we find a metanormative view that renders us capable of giving ourselves reasons to perform immoral acts problematic—regardless of whether these reasons will ever be decisive in determining what we should do, all things considered—then Chang's decision to 'go hybrid' will not save her view. I shall refer to this as the *problem of voluntarist reasons for immoral actions*. Chang seems to anticipate this problem. She replies that "there might, arguably, not be much difference in there being a reason that cannot change the valence and there being no reason at all" (2013a: 184). As I shall point out below, this remark is unsatisfactory.

Second, although voluntarist reasons for morally *impermissible* actions cannot change the valence, on Chang's account, morally dubious voluntarist reasons for *permissible* actions still might. Let me explain. As an example of an agent whose given reasons have 'run out', Chang introduces the case of Jane, who faces a hard choice between a career in banking and one in painting (2013a: 171). Now, suppose that Jane wills the fact that, as a banker, she will be able to disappoint the hopes of poor people applying for loans to be a reason for her to become a banker. Since, by hypothesis, Jane's given reasons have run out, this reason will tip the scales in favor of the banking career. Yet, one might think that, although becoming a banker *per se* is not immoral, Jane would now be doing so for a morally dubious reason. I shall refer to this as *the problem of morally dubious voluntarist reasons*.

3. The Second-Personal Account of Commitment

Although I agree with Chang that commitment is essentially a volitional phenomenon, and that a commitment to φ involves a volitional act that aims to give ourselves a reason to φ , I want to argue, *pace* Chang, that this volitional act is to be understood as an act of holding ourselves *accountable*.

Let me begin by noting that the volition involved in commitment, as Chang describes it, takes a *reflexive structure*: in committing to φ , I will that a reason to φ be created *by this very act of will*. In other words, I will, not only that a certain effect be achieved, but that it be achieved *through* my willing that it occur. Chang's remark that creating a voluntarist reason for yourself is a matter of "placing your will—your very agency—behind its being a reason" is therefore not merely metaphorical (2013b: 93). After all, if commitment is to be essentially volitional, it must consist in willing that one's own will be a source of reasons for oneself.

Importantly, this means that the volition involved in commitment embodies the same reflexive structure which, as we saw in Chapter 5, is definitive of second-personal address (see Section 2.2). In particular, second-personal address is a matter addressing a reason which would not exist but for its possibility of being addressed in this way (Darwall 2006: 13). This aspect of second-personal address is connected to its *irreducibly* second-personal nature: any instance of second-personal address invokes the authority to address reasons that would not exist but for the authority to address them and, thus, the authority "to demand respect for this very authority" (Darwall 2006: 14). Analogously, Chang depicts commitment as a matter of willing that there be a reason that would not exist otherwise. And this is due to the fact that, according to Chang, commitment is *irreducibly* volitional: willing that something be a reason to φ invokes the normative power to create reasons that would not exist but for the normative power to create them. By acting on our voluntarist reasons, then, we appropriately respond, not to the reason-giving significance of a principle or value external to our will, but to our own will's power to create reasons for us. In this sense, at least, the normative power involved in commitment is a power 'to demand respect for this very power' and, thus, analogous to second-personal authority.

Now, one might respond that, by virtue of embodying this reflexive structure, the volition involved in commitment is 'authority-regarding' (or 'power-regarding'), but not in the same sense as second-personal address. More specifically, one might hold that commitment is authority-regarding in the sense that it invokes the agent's *authority* to create reasons that derive their normativity from this very authority, but not in the sense

that it involves a second party that is thereby being held *accountable*. Indeed, this seems to be Chang's view. After all, as we saw in Section 2, Chang argues that the commitments she is concerned with do not require uptake by an addressee.

However, this is where the limitations of Chang's account come to the fore: if voluntarist reasons are to be genuinely *normative*, then there must be such a thing as a failure appropriately to respond to these reasons. That is, it must be possible for an agent to be at fault for failing to φ simply because she is committed to φ . I do not claim that, for a standard to be normative, it must be psychologically or physically possible to violate it. But we must at least be able to *conceive* of something that would count as falling short of that standard. Therefore, if voluntarist reasons are to be genuinely normative, it must be possible to act *against* an existing commitment. Yet, Chang's account does not offer sufficient conceptual resources to allow for this possibility.

One might point out that, above, I said that Chang's account promises to explain the normative difference between dropping a committed project on a mere whim and 'uncommitting' precisely because it is natural to think that there is a difference between deciding (not) to φ and willing that there (not) be reason to φ (see Section 2.1 above). One might object that I am now contradicting this earlier statement by arguing that, in Chang's account, there is no room for the possibility of deciding or acting against an existing commitment. However, I am not arguing that there is no difference between deciding not to φ and willing that there be reason to φ . Accordingly, I am not suggesting that no account that, like Chang's, characterizes commitment as essentially involving the volitional act of willing that there be reason makes room for the possibility of acting against an existing commitment. Instead, my point is that Chang's account, as it stands, does not offer sufficient conceptual resources to *accommodate* this possibility.

Now, Chang suggests that our commitments often have "downstream effects" (2013b: 95). That is, we often incur new 'given' reasons to φ through actions that we perform as a result of having committed to φ (e.g., because these actions give rise to new expectations on the part of others). One might argue that these downstream effects account for the possibility of being at fault for failing to do what one is committed to. However, while this shows that we can be at fault for failing to φ when we are committed to φ , it does not show that we can be at fault for failing to φ *because* we are committed to φ , and thus independently of our given reasons to φ . Accordingly, it seems that Chang's account, by itself, does not have the resources to explain why a commitment to φ , construed as a mere act of will, should give rise to a *normative* reason to φ .

I suggest that the only way to remedy this problem is to maintain that, by the authority-regarding volition involved in making a commitment, we actually *bind*

ourselves. That is, if by committing to φ we are indeed to create a normative reason to φ , the act of will by which we invoke our authority to create such a reason must simultaneously be an act of holding ourselves *accountable* for recognizing this reason. Otherwise commitment would not amount to more than some kind of psychological force, and the authority involved in creating a commitment would be *de facto* rather than *de jure*.⁶⁸

Put another way, if commitment did not involve holding ourselves accountable, it would be a matter of asserting the superiority of φ -ing in such a way as to render it *difficult*, psychologically speaking, to choose not to φ . But it would not be a matter of erecting a *normative* barrier against doing so. After all, by choosing not to φ we would ‘uncommit’ from φ . Thus, we could not be at fault for failing to φ simply by virtue of our commitment to φ . Therefore, if our will’s authority to commit is to have the normative significance that Chang ascribes to it, it must be an authority *over* some person, i.e., it must make a difference to someone’s normative situation. And in the case of commitments, this person can only be ourselves.

In other words, the volition involved in commitment is not only authority-regarding but also an instance of addressing ourselves second-personally.⁶⁹ This means that there is a difference between *uncommitting* from φ , thus undoing one’s voluntarist

⁶⁸ One might argue that such *de facto*, psychological authority could still put us under normative pressure to φ by triggering a rational requirement demanding consistency among our attitudes. However, this proposal cannot save Chang’s account. First, on this proposal, the normativity of our commitments does not derive from the normative powers of our *will* but from an independently valid, general principle of rationality. Second, on the assumption that commitment is not based on *de jure*, normative authority, it is unclear why rationality should require that we act on them. Perhaps commitment involves the *belief* that we ought to φ , triggering the requirement against akrasia: if you believe that you ought to φ , you ought to intend to φ . However, this requirement does not give us a *reason* to φ that could serve as a basis for rational choice when given reasons have ‘run out’. After all, on the present proposal, the belief that we ought to φ , based on the commitment, is itself irrational; for, by hypothesis, commitment only involves *de facto* authority. Hence, if the requirement against akrasia gave us reason to φ , we would (irrationally) be ‘bootstrapping’ a reason into existence (Kolodny 2005). Therefore, the requirement is either not normative at all (see, e.g., Kolodny 2005), or it has “wide scope”, in which case it merely requires that we *either* intend to φ *or* drop the belief that we ought to φ (see, e.g., Broome 1999).

⁶⁹ My reasoning in the last two paragraphs mirrors Korsgaard’s argument for the shareability of reasons in *Sources* (1996: 136-137, see Chapter 4, Section 2.2, above). However, whereas Korsgaard takes this argument to establish that reasons are inherently public, I employ it to show that authority-regarding reasons, if they are to be normative, must be second-personal.

reason to φ , and simply *failing* to φ . This opens up the conceptual space that we need in order to accommodate the possibility of acting *against* an existing commitment to φ . After all, it makes a difference whether one is released from a relation of authority and accountability or simply neglects it. To use a familiar example, it matters if Arnold simply ignores Bella's legitimate demand that Arnold withdraw his foot from on top of Bella's foot or if Bella waives that demand.

My account implies that, at the conceptual level, commitment does not involve a single, unified person. Rather, when we make a commitment we 'step outside of ourselves' to make a demand of ourselves. But this should not move us to reject the second-personal account of commitment. On the contrary, it is congenial to our ordinary understanding of commitments. First, as we saw, committing to φ is different from deciding to φ —as, indeed, it must be if there is to be a possibility of deciding against one's commitments (cf. Chang 2013b: 69). Although this does not, by itself, imply that commitment is second-personal, it suggests that there is a sense in which the committing agent and the deciding agent are not identical. Second, failure to fulfill one's commitments usually elicits 'reactive attitudes' of self-reproach, such as guilt and blame. By forming these attitudes, we implicitly hold ourselves accountable, taking up both the role of the person holding accountable and that of the person being held accountable (Darwall 2006: chapter 5, cf. Strawson 1968).

My proposal furthermore implies that Chang is wrong when she suggests that the difference between moral commitments and internal commitments is that the latter type of commitment does not involve uptake (see Section 2.2). After all, what I just argued suggests that, while internal commitments do not involve uptake by *another* person, they do require uptake by *oneself*. In particular, I argued that commitments cannot have the normative significance that Chang ascribes to them unless they involve the kind of authority by which one party holds another accountable. But if we are to hold ourselves to our commitments, we must *accept* them as binding in a way that resembles a promisee's acceptance of a promise.

In fact, I think that Chang does not sufficiently motivate her claim that the difference between moral and internal commitments is that the former require uptake while the latter do not to begin with. To see this, note first that Chang acknowledges, if only in passing, that some moral commitments do not involve uptake by a *separate* person, but by oneself (2013b: 76).⁷⁰ Having made this concession, Chang then goes on to argue that

⁷⁰ Interestingly, Chang even seems to follow Connie Rosati's argument that self-promises require uptake by oneself—and are second-personal—because there is a difference between breaking a

internal commitments are not moral commitments (2013b: 76-79). However, her arguments for this claim do not even consider the possibility that internal commitments involve uptake by oneself. Instead, she merely points out differences between internal commitments and those moral commitments that involve uptake by *another* person. First, Chang argues that the commitments under discussion are “very much an *internal* affair” and therefore “not moral in the sense of essentially being relied upon by others” (2013b: 77). But this is beside the point since, as Chang acknowledges, moral commitments, too, can be entirely internal and independent of others’ reliance upon their fulfillment. Second, Chang argues that, when we make a moral commitment (e.g., a promise) to someone else, then additionally making a corresponding internal commitment adds further “significance” to the former (2013b: 77-78, cf. Shpall 2014: note 2). The idea seems to be that, if the internal commitment were a moral commitment, it would not constitute an *additional* act of commitment that contributes further significance. However, unless one neglects the possibility of uptake by oneself, it is difficult to see why there should not be two commitments involving uptake, one interpersonal and one intrapersonal. Chang’s arguments thus do not give us reason to doubt that internal commitments involve uptake by oneself.

This is not to deny that there are important differences between commitments that involve uptake by another person and internal commitments. However, although I cannot argue this at length here, I think that these differences are artifacts of the different ways in which these commitments *come about*—and in particular the epistemic asymmetries between the two parties in the interpersonal case—rather than being down to one kind requiring uptake and the other kind not so requiring.⁷¹

self-promise and releasing oneself from it (Chang 2013b: note 5, Rosati 2011: 142-144). One might wonder in how far my argument differs from Rosati’s, given that my characterization of commitments is very similar to her characterization of self-promises. Indeed, I take our respective analyses to be mutually supportive. Yet, my argumentative strategy differs importantly from Rosati’s: while she bases her case on the intuitive notion that there is a difference between breaking a self-promise and releasing oneself from it, my argument invokes the reflexive structure of authority-regarding reasoning and its link to accountability to *explain* this intuitive notion.

⁷¹ Shpall makes a parallel argument to the effect that rational commitment is not, fundamentally, a different kind of relation than moral commitment (2014: 161-171). This raises the question of how rational commitment relates to internal commitment (see next note). I argued that internal and moral commitment share the same (second-personal) structure. If Shpall is right and moral and rational commitment are fundamentally the same relation, does my account extend to rational commitment also? Unfortunately, I cannot answer this question within this essay.

4. Commitment and Morality

I have argued that commitments are instances of second-personal address, and thus voluntarist reasons are second-personal reasons. I now want to highlight two important lessons that this argument teaches us about commitments on the one hand, and morality, as construed by the second-personal account, on the other. It turns out that the second-personal account not only captures the difference between dropping a commitment on a whim and uncommitting but also avoids the two other problems that, as we saw Section 2.2, haunt Chang's account: the problem of voluntarist reasons for immoral actions and the problem of morally dubious voluntarist reasons.

4.1 Commitments as Obligations (to Oneself)

As I hope to have shown in Chapter 5, second-personal reasons are the type of reasons that are implicated in the concept of *obligation*. To say that an agent is obligated to φ is to say that she can legitimately be held accountable for failure to φ —and thus that she has a second-personal reason to φ . We might say, then, that a second-personal reason, insofar as it rests on a relation of legitimate authority and accountability, constitutes a *pro tanto* obligation. It follows that, if voluntarist reasons are second-personal reasons, then commitments give rise to *pro tanto* obligations.

Indeed, that commitments give rise to obligations, rather than merely the type of reasons that *recommend* some course of action, finds independent support in our ordinary understanding of commitments. Alida Liberman and Mark Schroeder identify three key features that distinguish obligations from what we might call 'recommending reasons' and argue that commitments share these features (2016: 107-108, 117-118).⁷² First, they point out that obligations are "strict" (Liberman and Schroeder 2016: 107, cf. Shpall 2013: 733, 2014: 158-160). More specifically, obligations present themselves as "demands" rather than "suggestions" and we are criticisable for failures to comply with them in a way in which we are not for failures to act on a recommending reason (*ibid.*, see Chapter 1, Section 4 above). Indeed, even if we are justified or excused in not fulfilling a certain (*pro tanto*) obligation, the grounds for criticism need not vanish completely (Liberman and Schroeder 2016: 107). According to Darwall, this feature of obligations is due to their second-personal nature: if I am obligated to φ , I can legitimately be held

⁷² Strictly speaking, Liberman and Schroeder restrict their discussion to rational commitment (2013, 2014). However, their arguments seem to apply to internal commitments as well. This makes the question of how these commitments relate to rational commitments all the more pressing (see previous note).

accountable for failing to φ (see Chapter 5). Since this relation of accountability does not wither away when my obligation to φ is outweighed by other considerations, I might still be obligated to account for my action—e.g., by explaining or apologizing for what I did (Darwall 2006: 11-12).

As Liberman and Schroeder correctly observe, commitments share this feature of obligations. After all, as long as I am committed to pursue a career in philosophy, my failure to pursue such a career will at least in principle be subject to legitimate criticism—as I pointed out in Section 3, I might form reactive attitudes of self-reproach—even if my commitment is outweighed by other considerations. As Shpall pointedly remarks, if the same were true of recommending reasons, “then pretty much every act we performed, and every new attitude we formed, would be an occasion for such negative emotions” since, arguably, we constantly have recommending reasons for mutually exclusive courses of actions (2014: 160).

Second, Liberman and Schroeder observe that obligations have a different structure—or, as they put it, a different “ontology”—than recommending reasons (2016: 107-108). In particular, recommending reasons essentially figure in a triadic relation: some fact, F , is a reason to perform an act, φ , for an agent, A . Obligations, by contrast, essentially figure in a dyadic relation: A is obligated to φ . Thus, while it makes sense to ask A what *fact* her recommending reason to φ consists in, it would be nonsensical to ask the same of her obligation. Of course, we can ask how an obligation arose, and what justifies it. But this is different from the question of what the obligation *is*. Again, the same is true of commitments. My commitment is *to* pursue a philosophy career, but it would be strange to ask what *fact* this commitment consists in.

Third, obligations, unlike recommending reasons, can be *bipolar* or *directed* (Liberman and Schroeder 2016: 108). That is, A 's obligation to φ can be *owed* to an individual or group, such that A 's failure to φ (lacking excuse or justification) *wrongs* this individual or group. Clearly, recommending reasons do not possess this feature. Commitments, on the other hand, seem to be the kind of thing that can be bipolar. This is clearest in the case of moral commitments: if A promises B to φ , her commitment is *to* B . Whether internal commitments can be bipolar as well is less obvious. Within this thesis, I cannot do justice to the vastness and complexity of the philosophical debate concerning the conditions under which an obligation is bipolar.⁷³ Nevertheless, I would argue that, since moral commitments can be bipolar, there is no obvious reason to think that internal commitments cannot. After all, like moral commitments, internal

⁷³ For an overview of the debate over the conditions of directedness, see May (2015).

commitments (a) are generated through uptake by a particular person, (b) are grounded in the relation of authority and accountability that is thereby established between the committing agent and the uptaking person, (c) often correspond to the uptaking person's discretionary authority to *release* the committed agent, and (d) tend to elicit second-personal reactive attitudes, such as blame, if not fulfilled. Therefore, I suggest, we *owe* fulfillment of our internal commitments to ourselves.⁷⁴

We can conclude that commitments give us *pro tanto* obligations rather than recommending reasons. This has three important implications for our understanding of commitments, which correspond to the three features of obligations that I just highlighted. First, since obligations continue to exert normative force and warrant criticism even when they are outweighed, our voluntarist reasons do not just fall by the wayside

⁷⁴ Following Kant and Hohfeld, it is commonly held that a duty is owed to someone if and only if that person has a (claim-) right to the duty's fulfillment (Hohfeld 1964, Kant 1996: VI 221-229). Since this view is typically conjoined with the notion that a duty corresponds to a claim-right only if it is appropriate to enforce its fulfillment, it is taken to imply that only enforceable duties can be bipolar (Hohfeld 1964, *MM* 6:231-233, also see Chapter 6, Section 2.2, in this thesis). This, in turn, implies that commitments must be either enforceable or non-bipolar—and I certainly don't mean to suggest they are enforceable. However, while enforceability is a plausible condition for right-hood, I do not think that it is necessary for bipolarity. Accordingly, I do not think that there is a one-to-one correspondence between bipolar obligations and claim-rights (cf. Cruft 2013: 204 and 209; Darwall 2013b: 28, note 10, Wenar 2013: 214, note 24). (Interestingly, this gives rise to the conceptual possibility that obligations arising from imperfect duties are also directed, suggesting an alternative response to the objection to the second-personal account discussed in Section 2.3 of Chapter 6.)

Another potential objection to the claim that we owe fulfillment of our commitments to ourselves is that, on Darwall's account, obligations are bipolar only if they involve the obligor's *individual* second-personal authority to hold the obligee accountable *personally* (as opposed to the *representative* authority of every member of the moral community to hold her accountable *impersonally*; Darwall 2013a, also see Chapter 6 above). It may seem that I have not shown that commitments involve the former rather than the latter kind of authority. However, as Simon May notes, Darwall's account does not provide a criterion to determine which kind of authority is involved in an obligation, other than whether it is bipolar or not (May 2015: 527-528). Accordingly, any argument for the bipolarity of commitments that I provide is simultaneously an argument for the thesis that commitments correspond to individual second-personal authority. Conversely, in order to show that commitments do *not* correspond to individual second-personal authority, one would have to provide an argument to the effect that commitments are not bipolar.

whenever they fail to make a difference as to what we have all-things-considered reason to do. Hence, Chang's remark that "there might, arguably, not be much difference in there being a reason that cannot change the valence and there being no reason at all" is not true in the case of voluntarist reasons (2013a: 184, also see Section 2.2). This might make Chang's claim that we can give ourselves voluntarist reasons to perform immoral actions harder to accept. I return to this topic in the next sub-section.

Second, since obligations—unlike recommending reasons—do not consist in a fact that counts in favor of an action, a further departure from Chang's account is in order. In particular, committing to φ cannot be a matter of willing *some fact about φ -ing* to be a reason to φ (*pace* Chang 2013b: 93). But facts about φ -ing might still play a role in *justifying* our commitments. Ultimately, of course, the normativity of our commitments is underwritten by the normative powers of our will and thus, as I argued, by our second-personal authority. Yet, this authority is not unbounded. To the contrary, our second-personal reasons must be justifiable from within the second-person standpoint (see Chapters 5 and 6 above, as well as the next sub-section). At this level, facts about φ -ing might be relevant. Moreover, facts about φ -ing might determine what we have to do to *fulfill* the obligation that results from our commitment to φ . For example, if I am committed to a career in philosophy and, as a matter of fact, a degree in philosophy is required for such a career, then I am committed to getting such a degree.

Third, since commitments give rise to second-personally grounded obligations whose fulfillment we owe to ourselves, Darwall's second-personal framework of morality accommodates *obligations to oneself*. Rather than being restricted to obligations that we owe to others or to no one in particular, the second-person standpoint also governs what we owe to ourselves, as a matter of self-respect. I say more about this below.

4.2 Constraints on Commitment

Since voluntarist reasons are second-personal reasons, commitment is subject to moral constraints, Chang's remarks to the contrary notwithstanding (2013b: 108-109). More specifically, since second-personal address necessarily presupposes, as part of its normative felicity conditions, that we are morally obligated to respect each other's equal second-personal standing, some immoral actions will never be supported by voluntarist reasons.

In particular, if an action is incompatible *in principle* with the requirement to act in ways that are justifiable to all members of the moral community from within the second-person standpoint, we cannot give ourselves voluntarist reasons to perform that action. Consider, for example, the action of enslaving a person (cf. Darwall 2006: 263-268).

Not only does this action fail to be justifiable from within the second-person standpoint. In addition, slavery—by its very nature—denies the equal second-personal authority of the slave and thus involves the enslaver’s refusal to even *enter* the hypothetical dialogue in which a second-personal justification of her conduct vis-à-vis the slave would have to take place. Hence, it seems, there cannot be any second-personal reasons, including voluntarist reasons, to perform actions of enslavement or slave-holding. Of course, one might still *try* to commit to such an action, but the attempt will not succeed at generating a normative reason.⁷⁵ It follows, *pace* Chang, that we do not have to rely on the priority of given reasons over voluntarist reasons to make sure that we will never have all-things-considered reason to perform such an action (see Section 2.2 above). In this way, the second-personal account of commitments solves the problem of voluntarist reasons for immoral actions.

Note that this does not rule out the possibility that we end up being committed to an all-things-considered immoral action. Some actions are compatible in principle with the necessary presuppositions of second-personal address. By committing to these, we obligate ourselves to perform these actions. But the resulting obligations are *pro tanto* and can be outweighed by other moral considerations. In this way, we can end up being committed to an action which—under the circumstances—is morally impermissible (cf. Shpall 2014: 167-169). But this implication of the second-personal account is not to its discredit. After all, we should expect agents frequently to have at least some reason to perform actions that, although they are wrong under the circumstances that obtain, would be permissible under different circumstances.

I suggest, however, that the moral constraints on commitment do rule out that we will ever create a voluntarist reason which is itself *morally dubious*. Recall the case of Jane, from Section 2.2, who wills the fact that, as a banker, she could disappoint the hopes of poor people as a reason for her to pursue a banking career. I noted that, while

⁷⁵ One might object that, even if φ -ing is utterly incompatible with the necessary presuppositions of second-personal address, the attempt to commit to φ puts us under *some* normative pressure to φ . However, I would argue that insofar as there is such pressure, this is not because we are committed to φ , but because we *believe* that we are. And the normative force of this belief might in turn be explained by appeal to a wide-scope rational requirement (cf. Broome 1999, see note 68 above). Alternatively, one might argue that the normative force of this belief is due to a “rational commitment” (Shpall 2014: 158). This would seem to imply that rational commitments are not second-personal (for otherwise they would be subject to moral constraints), making the question of how rational commitments relate to internal commitments all the more pressing (see notes 71 and 72 above).

becoming a banker as such might be morally permissible, Jane's reason for doing so would be morally dubious. In the previous subsection, I already observed that voluntarist reasons, *qua* obligations, do not consist in facts. This prevents Jane from making the *fact* that she could disappoint poor people into a voluntarist reason. Yet, as I also noted, this does not prevent facts from playing a role in the second-personal *justification* of voluntarist reasons. The moral constraints on commitment, however, do prevent *some* facts from playing such a role. After all, a second-personal reason is valid only if it is justifiable to all members of the moral community from within the second-person standpoint. This plausibly constrains the kinds of considerations that can be invoked against, and in favor of, particular instances of second-personal address. What exactly these constraints are is an issue that I cannot discuss at length here.⁷⁶ But whatever they turn out to be, I expect that they will rule out invocations of considerations which—like Jane's appeal to the fact that, in banking, she could disappoint the hopes of poor people—themselves exemplify a certain disregard for persons' second-personal standing. Thus, the second-personal account avoids the problem of morally dubious voluntarist reasons.⁷⁷

From the fact that commitments are subject to moral constraints in the ways that I described in this sub-section, it follows that voluntarist reasons not only compete with moral reasons but *are* moral reasons. After all, if a voluntarist reason is validated by the necessary presuppositions of second-personal address, this means that the person who creates it can legitimately address it by invoking her second-personal authority.

⁷⁶ Ultimately, the constraints on commitment are set by the Categorical Imperative. After all, as we saw in previous chapters, the Categorical Imperative is the constitutive principle of the second-person standpoint and the supreme principle of morality. However, the details of these constraints depend on how the Categorical Imperative is to be applied in practical judgment to yield substantive verdicts. On one promising proposal, a demand can intelligibly be raised from within the second-person standpoint only if its fulfillment would be conducive to the interests of its addresser, *qua* second-personally competent being (Darwall 2006: 309-310, cf. Hill 1989: 768, Kumar 2003). This idea is structurally analogous to Rawls's notion of "primary goods", which are conducive to the interests of every citizen, *qua* free and equal moral person (1980: 525-526).

⁷⁷ As far as I can see, while the second-personal account rules out that commitments be justified by morally dubious reasons, it does not rule out that commitments be made from a morally dubious *motivation*. However, I do not think that this is a problem for the second-personal account. It seems independently plausible that we might sometimes address legitimate demands, be it with each other or with ourselves, from a bad motivation. For example, a promisee's demand that a promise to her be fulfilled might be legitimate even if her sole motivation for addressing this demand is to make the promisor's life more difficult.

Accordingly, to recognize this reason—and to act as it directs if it goes undefeated—is a matter of respect for that person’s equal second-personal standing.

5. Second-Personal Reasoning and Intrapersonal Morality

At the end of the previous chapter, I noted that Korsgaard’s objection to the second-personal account, though it is explicitly formulated as an argument about the *optionality* of the second-person standpoint, might instead be based on the worry that second-personal address is a *social* practice and that the second-person standpoint is therefore *unavailable* to individuals who live in isolation from society. In this section, I want to show how my argument for the second-personal account of commitment resolves this worry.

Recall that the worry is that, insofar as second-personal address is a social practice, an encounter with the moral law is ruled out for agents who do not interact with others. In this chapter, however, I have argued that commitment constitutes a second-personal relation between oneself and one’s own will. When we commit to φ , willing that there be (additional) reason for us to φ , we address ourselves second-personally, holding ourselves accountable for φ -ing. On my account, commitments offer an avenue of access into second-personal reasoning that is symmetric to the demands that concrete others might make on us: just like someone else’s demands prompt us to take up the second-person standpoint, and thus facilitate an encounter with the moral law through the ‘fact of reason’, so do the commitments that we make, independently of our relation to others. It is thus not true, contrary to the complaint which I discussed at the end of Chapter 6, that interpersonal interaction constitutes the *paradigmatic* way of taking up the second-person standpoint.

Isolated individuals can therefore access the second-person standpoint, and encounter the moral law, not only by invoking their representative authority as members of the moral community, but also by invoking their individual second-personal authority over themselves, and thus putting themselves under an *obligation to themselves*. It follows that the second-personal account does not abandon the fundamental symmetry between the intrapersonal and the interpersonal realm which plays such a central role in Kantian moral theory.⁷⁸

⁷⁸ One might object that, from the fact that commitments are second-personal obligations to oneself, it does not follow that the second-personal account accommodates the whole range of Kantian duties to the self. For example, it is not obvious that the second-personal account accommodates the duty to develop one’s talents (cf. *G* 4:422-423). However, my aim in this chapter has not been to demonstrate how the particular duties to oneself that Kant endorses can be vindicated

Indeed, from what I have said in this and the previous chapter, it follows that second-personal address is not, first and foremost, a *social* practice at all. Instead, it is a mode of *reasoning* that can be applied to other-regarding as well as purely self-regarding contexts. What makes this mode of reasoning distinctly second-personal is its irreducibly authority-regarding nature, i.e., its exclusive concern with reasons that would not exist but for someone's authority to address them. While second-personal reasoning thus always takes an *addressive* and *relational* form in that it inevitably invokes our relation to a person or will who has the authority to make demands on us, it need not invoke our relation to a concrete, other person.

First, as I have argued in the previous chapter, an instance of second-personal address need not invoke anyone's individual authority over the addressee. Instead, it might invoke the representative second-personal authority of any member of the moral community. This will be the case both when the obligation that is thereby being addressed is owed to a person other than the addresser and when the obligation in question is not owed to anyone, i.e., when it is an instance of moral obligation *period*. As I have noted in the previous chapter, in either case the addresser need not be another person, separate from the addressee.

Cases in which we invoke our representative authority to address ourselves internally can be described as facilitating an immediate, direct encounter with the moral law within our own practical reason. Yet, it is important to note that such instances of second-personal address are still addressive and relational in character. After all, as I have argued in Chapters 5 and 6, the concept of moral obligation is tied to the "regulative ideal" of the kingdom of ends, the ideal of a "community of equal, mutually accountable, free and rational agents as such" (Darwall 2007: 64 and 2006: 101, respectively). In other words, when we encounter the law directly in our own practical reason we encounter our membership in a community of agents that are mutually accountable for complying with that law. Now, as I have also noted in Chapter 6, the law marks out persons as members

from within the second-person standpoint. Rather, my aim has been to establish that the second-personal account does not abandon the fundamental *structural* symmetry between the interpersonal and the intrapersonal realm that characterises Kantian moral theory. As I have argued, the Categorical Imperative is the constitutive principle of second-personal reasoning. Therefore, whether the second-personal account yields the particular duties to oneself—or for that matter, *any* of the duties—that Kant endorses depends on whether these can be derived from the Categorical Imperative, a question which I bracket in this thesis. In this respect, the second-personal account does not differ from most other Kantian moral theories.

in this community of co-legislating agents because they possess second-personal competence and, hence, autonomy and the capacity for acting from duty (see Section 2.1). As Kant puts it, “morality is the condition under which alone a rational being can be an end in itself; because it is possible only by this to be a legislating member in the kingdom of ends” (*G* 4:435, cf. Darwall 2006: 102). It follows that, by encountering the moral law within our own practical reason, and the representative authority of all members of the moral community that is conceptually linked to it, we encounter the dignity of the *person*, where ‘person’ is understood as the (second-personal) concept of a second-personally competent agent (cf. Darwall 2006: 80). Accordingly, we might say that internal address invoking representative authority, although it does not involve our accountability to a concrete person, is nevertheless addressive and relational in that it involves the authority of the very *idea* or *form* of a person.⁷⁹

That the second-personal account allows us to describe internal address invoking representative authority in this way should not surprise us. After all, as I have noted, this type of address can be viewed as facilitating a direct encounter with the moral law in our own practical reason. Now, the moral law is the formal principle of the autonomous will (see Chapter 1 above). Insofar as “persons” are distinguished from “things” and marked out as “ends in themselves” by their possession of autonomous wills, this allows us to say that the law is the formal principle of *the person* (cf. *G* 4:428). In addition, it is not uncommon, among Kantians, to characterise the law as *addressing* us (e.g., Sensen 2011: 104). Indeed, Kant himself frequently refers to the law as a “command” of reason (e.g., *G* 4:413, 4:416) and notes that we feel “compelled to revere the (moral) human being within [our] own person” (*MM* 6:436). About conscience, he writes (*MM* 6:438):

[Conscience] is peculiar in that, although its business is a business of a human being with himself, one constrained by his reason sees himself constrained to carry it on as at the bidding *of another person*.

These expressions, in conjunction with the fact that the law can be described as the formal principle of the person, suggest that, when we encounter the law, we are addressed by the form, or the very idea, of a person.

The second reason why second-personal address is not, fundamentally, a social practice is that, as I have argued in this chapter, we can also address ourselves second-personally by invoking our own individual authority over ourselves. In particular, we do

⁷⁹ Similarly, Sensen writes: “The proper moral motive is respect for the moral law or the idea of a moral human being” (2011: 172).

so whenever we make, demand compliance with, amend, or withdraw a commitment. While these instances of second-personal address, unlike those that merely involve representative authority, invoke our relation of authority and accountability to a *concrete* person, ourselves, they are not social in the sense of involving another person.

On the picture that emerges, second-personal address is not, fundamentally, a social practice but a distinct practice of reasoning. The hallmark of second-personal reasoning is its inevitable reference to the irreducibly second-personal authority to address demands. The central claim of the second-personal account is that, in order to elicit the Kantian analysis of obligation and the connected Kantian concepts of autonomy, etc., we need to emphasise the second-personal nature of obligation, i.e., the fact that the concept of obligation is inevitably bound up with, and implicated in, second-personal reasoning. But to maintain this claim, the second-personal account need not suppose that obligations depend on actual interactions between concrete individuals.

6. Practical Identities Revisited

The role that I have ascribed to commitments within the second-personal account is similar to the role that Korsgaard ascribes to our conception of practical identity within her first-personal account (see Chapter 4 above). First of all, both commitments and the adoption of a practical identity are, to an extent, ‘up to us’. That is, we can choose, within limits, what to commit to and what conception of practical identity to adopt.

Second, both commitments and practical identities, within their respective frameworks, serve as sources of normativity by creating normative reasons. Indeed, both commitments and the adoption of a practical identity involve *willing that there be reason* to act in certain ways. To see that Korsgaard thinks adopting a practical identity involves this type of volition, note that she contrasts the adoption of a practical identity with “particularistic willing” (2009: 72-76). Now, in Korsgaard’s framework, “particularistic willing would be a matter of willing a maxim for exactly this occasion without taking it to have any other implications of any kind for any other occasion” (2009: 75). In other words, particularistic willing is a type of willing which aims to have no implications for what we ought, or have reason, to do. To adopt a practical identity, by contrast, is to will that that identity—or the set of practical principles that constitute it—be a source of reasons and obligations (Korsgaard 1996: 103-104, 2009: 21).

Third, both the act of making a commitment and the adoption of practical identities are purely self-regarding volitions insofar as they do not essentially involve another, separate person.

Fourth, I have argued that commitment is an instance of second-personal address, facilitating an encounter with the moral law. Accordingly, commitment presupposes that we are autonomous and, thus, bound by the Categorical Imperative. As we have seen Chapter 4, Korsgaard ascribes a similar role to the adoption of practical identities. Connectedly, I have argued that commitments, because of their relation to autonomy and the Categorical Imperative, are subject to moral constraints. And, again, Korsgaard says the same, *mutatis mutandis*, about the adoption of practical identities. In particular, she argues that, since practical identities are sources of normativity only insofar as they solve the practical problem posed by our identity as rational human beings, i.e., beings who have a reflective consciousness, they must be compatible with the formal principle of that identity if they are to provide us with genuinely normative reasons (Korsgaard 1996: 120-123).

Finally, Korsgaard even depicts the way in which our conception of practical identity gives rise to obligations as involving a relation of authority and accountability (1996: 104). In particular, she contends that, when we adopt a conception of practical identity, there is a sense in which our practical deliberation involves two selves: an “acting self” and a “thinking self” (ibid.). The thinking self, representing the demands of our practical identity, has a certain authority over the acting self as “the acting self concedes to the thinking self its right to government” (ibid.). We might say that on Korsgaard’s account, by adopting a conception of practical identity we *address* ourselves, entering a relation of authority and accountability with ourselves.

In light of these similarities, one might wonder if my arguments in this chapter provide a vindication of Korsgaard’s account of practical identities from within the second-personal account.⁸⁰ I think that my arguments can indeed be interpreted in this way, albeit with some important qualifications. Most importantly, whereas Korsgaard thinks that the constraints on the first-person standpoint of agency as such require that we adopt a conception of our practical identity and, in this way, take up the second-person standpoint, I do not think that we inescapably make commitments. Again, as I have said in the previous chapter, there is nothing in the constraints of human nature or agency to render the second-person standpoint inescapable for us. While this does not mean that

⁸⁰ Relatedly, Herlinde Pauer-Studer points out that it is an open question whether Darwall is prepared to accept Korsgaard’s reflections on the connection between practical identities and the moral law if they are viewed in isolation from her attempt to vindicate autonomy from within the first-person standpoint (2010: note 8).

we can *choose* whether or not to take up this standpoint, it implies that human agents do not *necessarily* reason second-personally.

As I have argued in Chapters 5 and 6, this point is connected to the irreducibly second-personal nature of second-personal reasoning. Now, this irreducibly second-personal nature is, in turn, connected to a second important qualification: the second-personal account vindicates Korsgaard's notion of practical identities only insofar as practical identities are understood to give rise to a distinct *kind* of practical reason. In particular, practical identities must be understood as giving rise to *obligations*, where these are in turn analysed in terms of authority-regarding (i.e., second-personal) reasons. Accordingly, on my account, not *all* kinds of reasons are grounded in second-personal address, be it interpersonal or intrapersonal. Connectedly, we do not *need* to adopt a conception of practical identity, where this is understood as an instance of second-personal address, in order to have reasons for action (*pace* Korsgaard 1996: 101).

It is important to bear in mind, then, that although commitments provide an intrapersonal access to the second-person standpoint that is symmetrical to the access provided by interpersonal second-personal address, the second-personal account does not collapse into the first-personal account. For, to engage in second-personal address, be it with oneself or with other persons, is to engage in a distinct mode of reasoning that cannot be reduced to—and is not required by the constraints of—the first-personal mode of reasoning that we must engage in insofar as we are agents. This reinforces Darwall's point that, although the second-person standpoint is *a* first-person standpoint, it is never *merely* first-personal (2006: 10). While second-personal reasoning always proceeds from the point of view of a particular person, an 'I', it is distinct from other modes of first-personal reasoning because it invariably involves a relation of authority and accountability to a 'You'—be it the reasoner herself, another concrete person, or the idea of a person as such.

7. Conclusion

When we commit to φ , we will that a reason to φ be created through this very act of will. I have argued that if we are to succeed at creating a genuinely normative reason in this way, we must be addressing ourselves second-personally, holding ourselves accountable for φ -ing. It follows that, *pace* Chang, what we can successfully commit to is subject to moral constraints. Most notably, if a commitment is not to 'misfire' it must be at least in principle compatible with the necessary presuppositions of second-personal address. But if a commitment is within the moral constraints set by second-personal address, then it becomes an obligation to oneself and, thus, a moral obligation. This conclusion, in

conjunction with my remarks from the previous chapter, gives rise to an understanding of second-personal address according to which it is, not a social practice, but a distinct mode of reasoning that can be applied to self-regarding as well as to other-regarding contexts. Commitment provides an intrapersonal access to the second-person standpoint which is symmetrical to the access provided by interpersonal instances of second-personal address.

CONCLUSION

In this thesis, I have provided a restatement of Kantian constructivism that clears up some of the ambiguities and avoids some of the problems that have haunted previous statements of the view. My restatement of Kantian constructivism asserts that moral normativity has its source in the form of second-personal reasoning, a distinct mode of reasoning that already acknowledges, implicitly, the normativity of the moral law and the dignity of all persons as co-legislators in the kingdom of ends. By taking a stance on the source of moral normativity, Kantian constructivism addresses a metaethical question, albeit one that is distinct from the questions that the views it is usually contrasted with, such as moral realism, focus on. Unlike competing views of the source of moral normativity, Kantian constructivism provides an answer to the Normative Question. It characterises moral obligations as purporting to bind us by the form of our will, thereby providing their own *rationale*, so to speak, as to why we are rationally required to do what we believe we morally ought to do. While this answer to the Normative Question is not infallible, it suffices to defend our faith in morality's authority, rendering us immune to potentially debunking explanations of morality.

Bibliography

- Bagnoli, Carla (ed.). 2013a. *Constructivism in Ethics*. Cambridge, UK: Cambridge University Press.
- 2013b. “Introduction.” in Bagnoli 2013a.
 - 2014. “Starting Points: Kantian Constructivism Reassessed”, *Ratio Juris*, 27(3): 311-329.
 - 2016. “Kantian Constructivism and the Moral Problem”, *Philosophia* 44: 1229-1246.
 - 2017. “Constructivism in Metaethics.” *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/win2017/entries/constructivism-metaethics/>>.
- Blackburn, Simon. 2000. *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Clarendon Press.
- Broome, John. 1999. “Normative Requirements.” *Ratio* 12(4): 398-419.
- Chang, Ruth. 2009. “Voluntarist Reasons and the Sources of Normativity”, David Sobel and Steven Wall (eds.), *Reasons for Action*. Cambridge, UK: Cambridge University Press.
- 2013a. “Grounding Practical Normativity: Going Hybrid”, *Philosophical Studies* 164: 163–187.
 - 2013b. “Commitments, Reasons, and the Will”, in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics: Volume 8*. Oxford: Oxford University Press.
 - 2015. “Transformative Choices.” *Res Publica* 29(2): 237-282.
- Copp, David. 2005. “A Skeptical Challenge to Moral Non-Naturalism and a Defense of Constructivist Naturalism.” *Philosophical Studies* 126: 269-283.
- 2013. “Is Constructivism an Alternative to Moral Realism?”, in Bagnoli 2013a.
- Cruft, Rowan. 2013. “Why is it Disrespectful to Violate Rights?” *Proceedings of the Aristotelian Society* 113(2pt2): 201-224.
- Cumminskey, David. 2008. “Dignity, Contractualism and Consequentialism.” *Utilitas* 20(4): 383-408.
- Dancy, Jonathan. 2006. *Ethics Without Principles*. Oxford: Oxford University Press.
- Darwall, Stephen. 1977. “Two Kinds of Respect”, *Ethics* 88(1): 36-49.
- 1983. *Impartial Reason*. Ithaca: Cornell University Press.
 - 1990. “Autonomist Internalism and the Justification of Morals”, *Nous*, 24(3): 257-267.
 - 1992. “Internalism and Agency”, *Philosophical Perspectives*, 6: 155-174.

- 1995. *The British Moralists and the Internal 'Ought': 1640-1740*. Cambridge, UK: Cambridge University Press.
 - 2006. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press.
 - 2007. “Reply to Korsgaard, Wallace, and Watson.” *Ethics* 118(1):52–69.
 - 2013a. *Morality, Authority, and Law: Essays in Second-Personal Ethics I*. Oxford: Oxford University Press.
 - 2013b. “Bipolar Obligation”, in Darwall 2013a.
 - 2013c. “Moral Obligation: Form and Substance”, in Darwall 2013a.
 - 2013d. ““Because I Want It””, in Darwall 2013a.
 - 2013e. “The Value of Autonomy and Autonomy of the Will”, in Darwall 2013a.
 - 2016. “Making the ‘Hard’ Problem of Moral Normativity Easier”, in Lord and Maguire 2016.
 - 2017. “Morality, Blame, and Internal Reasons”, in Singer 2017.
- Darwall, Stephen and de Kenessey, Brendan. 2014. “Moral Psychology as Accountability”, in Justin D’Arms and Daniel Jacobson (eds.), *Moral Psychology and Human Agency: Philosophical Essays on the Science of Ethics*. Oxford: Oxford University Press.
- Darwall, Stephen, Gibbard, Allan and Railton, Peter. 1992. “Toward Fin de Siècle Ethics: Some Trends”, *The Philosophical Review*, 101(1): 115-189.
- Dreier, James. 2015. “Can Reasons Fundamentalism Answer the Normative Question?”, in Gunnar Björnsson, Caj Strandberg, Ragnar Francén Olinder, John Eriksson, and Fredrik Björklund (eds.), *Motivational Internalism*. Oxford: Oxford University Press.
- Driver, Julia. 2013. “Moral Sense and Sentimentalism,” in Roger Crisp (ed.), *The Oxford Handbook of the History of Ethics*. Oxford: Oxford University Press.
- Ebels-Duggan, Kyla. forthcoming. “Beyond Words: Inarticulable Reasons and Reasonable Commitments.” *Philosophy and Phenomenological Research*.
- Engstrom, Stephen. 2013. “Constructivism and Practical Knowledge”, in Bagnoli 2013a.
- Enoch, David. 2006. “Agency, Shmagency: Why Normativity Won’t Come from What is Constitutive of Action”, *Philosophical Review*, 15(2): 169-198.
- 2011. “Shmagency Revisited”, in M. Brady (ed.), *New Waves in Metaethics. New Waves in Philosophy*. London: Palgrave Macmillan.
- Fine, Kit. 2001. “The Question of Realism.” *Philosophers’ Imprint* 1(1): 1-30.
- FitzPatrick, William. 2005. “The Practical Turn in Ethical Theory: Korsgaard’s Constructivism, Realism, and the Nature of Normativity.” *Ethics* 115: 651-691.

- 2008. “Robust Ethical Realism, Non-Naturalism, and Normativity”, in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics, Volume 3*. Oxford: Oxford University Press.
 - 2009. “Recent Work on Ethical Realism.” *Analysis* 69(4): 746-760.
 - 2013. “How not to be an Ethical Constructivist”, in Bagnoli 2013a.
- Fitzpatrick, William and Watson, Eric. 2002. “O’Neill and Korsgaard on the Construction of Normativity.” *The Journal of Value Inquiry* 36: 349-367.
- Foot, Philippa. 1972. “Morality as a System of Hypothetical Imperatives.” *The Philosophical Review* 81(3): 305-316.
- Forst, Rainer. 2011. “Practical Reasons and Justifying Reasons: On the Foundations of Morality”, in Jeffrey Flynn (trans.), *The Right to Justification: Elements of a Constructivist Theory of Justice*. New York: Columbia University Press.
- Hanisch, Christoph. 2014. “Self-Constitution and Other-Constitution.” *Grazer Philosophische Studien* 90: 105–30.
- Harman, Gilbert. 2011. “Notes on Practical Reasoning.” *Cogency* 3(4): 127-145.
- Hart, H. L. A. 1955. “Are there any Natural Rights?” *The Philosophical Review* 64(2): 175–191.
- Hill, Thomas E. Jr. 1973. “Servility and Self-Respect.” *The Monist* 57(1): 87-104.
- 1989. “Kantian Constructivism in Ethics.” *Ethics* 99: 752-770.
 - 2007. “Kantian Normative Ethics”, in David Copp (ed.), *The Oxford Handbook of Ethical Theory*. Oxford: Oxford University Press.
- Hobbes, Thomas. 1651. *Leviathan, or, The Matter, Form, and Power of a Commonwealth Ecclesiastical and Civil*. London: Printed for Andrew Crooke, at the Green Dragon in St. Paul’s Church-yard.
- Hohfeld, W. N. 1964. *Fundamental Legal Conceptions as Applied in Judicial Reasoning*, W. W. Cook (ed.). New Haven: Yale University Press.
- Horgan, Terry and Timmons, Mark. 2006. “Morality Without Moral Facts”, in James Dreier (ed.), *Contemporary Debates in Moral Theory*. Oxford: Blackwell.
- Hussain, Nadeem J. Z. and Shah, Nishi. 2006. “Misunderstanding Meta-ethics: Korsgaard’s Rejection of Realism”, in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics: Volume 1*. Oxford: Oxford University Press.
- 2013. “Meta-Ethics and its Discontents: A Case-Study of Korsgaard”, in Bagnoli 2013a.
- Hume, David. 1896. *A Treatise of Human Nature*, L.A. Selby-Bigge (ed.). Oxford: Clarendon Press.
- Kagan, Shelly. 1991. *The Limits of Morality*. Oxford: Oxford University Press.

- Kain, Patrick. 2004. "Self-Legislation in Kant's Moral Philosophy." *Archiv für Geschichte der Philosophie* 86: 267-306.
- Kamm, Frances. 1992. "Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status." *Philosophy & Public Affairs* 22(4): 354-389.
- Kant, Immanuel. 1996a. *Practical Philosophy, The Cambridge Edition of the Works of Immanuel Kant*, Mary Gregor (ed. and trans.). Cambridge, UK: Cambridge University Press.
- 1996b. *Religion and Rational Theology, The Cambridge Edition of the Works of Immanuel Kant*, George Di Giovanni and Allen W. Wood (eds. and trans.). Cambridge, UK: Cambridge University Press.
- 1999. *Critique of Pure Reason, The Cambridge Edition of the Works of Immanuel Kant*, Paul Guyer and Allen W. Wood (eds. and trans.). Cambridge, UK: Cambridge University Press.
- 2011. *Groundwork of the Metaphysics of Morals: A German-English Edition*, Mary Gregor and Jens Timmermann (eds. and trans.). Cambridge, UK: Cambridge University Press.
- Kleingeld, Pauline. 2010. "Moral Consciousness and the 'Fact of Reason'", in Reath and Timmermann 2010.
- Kolodny, Niko. 2005. "Why Be Rational?" *Mind* 114(445): 509-563.
- Korsgaard, Christine. 1986. "Skepticism About Practical Reason." *The Journal of Philosophy* 83 (1): 5-25.
- 1989a. "Kant's Analysis of Obligation: The Argument from 'Foundations I'." *The Monist* 72(3): 311-340.
- 1989b. "Personal Identity and the Unity of Agency: A Kantian Response to Parfit." *Philosophy & Public Affairs* 18(2): 101-132.
- 1996. *The Sources of Normativity*. Cambridge, UK: Cambridge University Press.
- 2007. "Autonomy and the Second Person Within : A Commentary on Stephen Darwall's *The Second-Person Standpoint*." *Ethics* 118(1):8–23.
- 2008a. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press.
- 2008b. "The Normativity of Instrumental Reason", in Korsgaard 2008a.
- 2008c. "From Duty and for the Sake of the Noble", in Korsgaard 2008a.
- 2008d. "Acting for a Reason", in Korsgaard 2008a.
- 2008e. "Realism and Constructivism in 20th Century Moral Philosophy", in Korsgaard 2008a.

- 2009a. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.
 - 2009b. “The Activity of Reason.” *Proceedings and the Addresses of the APA* 83(2): 23-42.
 - manuscript. “Normativity, Necessity, and the Synthetic A Priori: A Response to Derek Parfit”. URL = <<http://www.people.fas.harvard.edu/~korsgaard/Korsgaard.on.Parfit.pdf>>.
- Kumar, Rahul. 2003. “Reasonable Reasons in Contractualist Moral Argument.” *Ethics* 114(1): 6-37.
- Lenman, James and Shemmer, Yonatan (eds.). 2012. *Constructivism in Practical Philosophy*. Oxford: Oxford University Press.
- Liberman, Alida and Schroeder, Mark. 2016. “Commitment: Worth the Weight”, in Lord and Maguire 2016.
- Lewis, David. 1996. “Desire as Belief II.” *Mind* 105(418): 303-313.
- Lord, Errol and Maguire, Barry (eds.). 2016. *Weighing Reasons*. Oxford: Oxford University.
- Mackie, J. L. 1990. *Inventing Right and Wrong*. London: Penguin.
- May, Simon Căbulea. 2015. “Directed Duties.” *Philosophy Compass* 10(8): 523-532.
- Miller, Christian. 2009. “The Conditions of Moral Realism.” *Journal of Philosophical Research* 34: 123-155.
- Nagel, Thomas. 1970. *The Possibility of Altruism*. Oxford: Clarendon Press.
- 1995. “Personal Rights and Public Space.” *Philosophy & Public Affairs* 24(2): 83-107.
- Neiman, Susan. 1994. *The Unity of Reason: Rereading Kant*. New York: Oxford University Press.
- O’Neill, Onora. 1989a. *Constructions of Reason: Explorations of Kant’s Practical Philosophy*. Cambridge, UK: Cambridge University Press.
- 1989b. “Preface”, in O’Neill 1989a.
 - 1989c. “Reason and Politics in the Kantian Enterprise”, in O’Neill 1989a.
 - 1989d. “The Public Use of Reason”, in O’Neill 1989a.
 - 1989e. “Constructivisms in Ethics”, in O’Neill 1989a.
 - 2015a. *Constructing Authorities: Reason, Politics and Interpretation in Kant’s Philosophy*. Cambridge, UK: Cambridge University Press.
 - 2015b. “Introduction”, in O’Neill 2015a.
 - 2015c. “Vindicating Reason”, in O’Neill 2015a.

- 2015d. “Orientation in Thinking: Geographical Problems, Political Solutions”, in O’Neill 2015a.
- Parfit, Derek. 1997. “Reasons and Motivation.” *Proceedings of the Aristotelian Society, Supplementary Volumes* 71: 99-130.
- 2006. “Normativity”, in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics: Volume 1*. Oxford: Oxford University Press.
 - 2011a. *On What Matters: Volume I*. Oxford: Oxford University Press.
 - 2011b. *On What Matters: Volume II*. Oxford: Oxford University Press.
- Pauer-Studer, Herlinde. 2010. “The Moral Standpoint: First or Second Personal?” *European Journal of Philosophy* 18(2): 296-310.
- Pettit, Philip and Smith, Michael. 1990. “Backgrounding Desire.” *Philosophical Review* 99: 565-592.
- Pistorius, Hermann Andreas. 2007. “Rezension der Grundlegung zur Metaphysik der Sitten”, in Bernward Gesang (ed.), *Kant’s vergessener Rezensent: Die Kritik der theoretischen und praktischen Philosophie Kants in fünf Rezensionen von Herrman Andreas Pistorius, Kant-Forschungen 18*. Hamburg: Felix Meiner Verlag.
- Plato. 2013. *Republic, Volume I: Books 1-5*, Christopher Emlyn-Jones and William Preddy (eds. and trans.). Cambridge, MA: Harvard University Press.
- 2017. *Euthypro, Apology, Crito, Phaedo*, Christopher Emlyn-Jones and William Preddy (eds. and trans.). Cambridge, MA: Harvard University Press.
- Prichard, H. A. 1912. “Does Moral Philosophy Rest on a Mistake?” *Mind* 21(81): 21-37.
- Rawls, John. 1980. “Kantian Constructivism in Moral Theory”, *Journal of Philosophy*, 77(9): 515-572.
- 1999. *A Theory of Justice*, rev. 2nd edn. Cambridge, MA: Harvard University Press.
 - 2000. *Lectures on the History of Moral Philosophy*, Barbara Herman (ed.). Cambridge, MA: Harvard University Press.
- Raz, Joseph. 1984. “On the Nature of Rights.” *Mind* 93: 194–214.
- Rosati, Connie. 2011. “The Importance of Self-Promises.” Hanoch Sheinman (ed.), *Promises and Agreements: Philosophical Essays*. Oxford: Oxford University Press.
- Rosen, Gideon. 2009. “Metaphysical Dependence: Grounding and Reduction,” in Bob Hale & Aviv Hoffmann (eds.), *Modality: Metaphysics, Logic, and Epistemology*. Oxford: Oxford University Press.

- Reath, Andrews. 2006a. *Agency & Autonomy in Kant's Moral Theory*. Oxford: Oxford University Press.
- 2006b. “Autonomy of the Will as the Foundation of Morality”, in Reath 2006a.
 - 2006c. “Legislating for a Realm of Ends: The Social Dimension of Autonomy”, in Reath 2006a.
 - 2010. “Formal Principles and the Form of Law”, in Reath and Timmermann 2010.
 - 2012. “Formal Approaches to Kant's Formula of Humanity”, in Sorin Baiasu and Mark Timmons (eds.), *Kant on Practical Justification*. Oxford: Oxford University Press.
 - 2013. “The Ground of Practical Laws”, in Stefano Bacin, Alfredo Ferrarin, Claudio La Rocca, and Margit Ruffing (eds.), *Kant und die Philosophie in weltbürgerlicher Absicht, Akten des XI Internationalen Kant-Kongresses, Band 3*. Berlin: DeGruyter.
- Reath, Andrews and Timmermann, Jens (eds.). 2010. *A Critical Guide to Kant's Critique of Practical Reason*. Cambridge, UK: Cambridge University Press.
- Rosati, Connie. 2011. “The Importance of Self-Promises”, in Hanoeh Sheinman (ed.), *Promises and Agreements: Philosophical Essays*. Oxford: Oxford University Press.
- Ross, William D. 2002. *The Right and the Good*, Philip Stratton-Lake (ed.). Oxford: Clarendon Press.
- Scanlon, Thomas M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- 2007. “Structural Irrationality” in Geoffrey Brennan, Robert Goodin, Frank Jackson and Michael Smith (eds.), *Common Minds: Themes From the Philosophy of Philip Pettit*. Oxford: Clarendon Press.
 - 2012. “The Appeal and Limits of Constructivism”, in Lenman and Shemmer 2012.
 - 2014. *Being Realistic About Reasons*. Oxford: Oxford University Press.
- Schaab, Janis D. 2018. “Why it is Disrespectful to Violate Rights: Contractualism and the Kind-Desire Theory.” *Philosophical Studies* 175(1): 97-116.
- Shafer-Landau, Russ. 2003. *Moral Realism: A Defence*. Oxford: Oxford University Press.
- Shoemaker, David. 2011. “Psychopathy, Responsibility, and the Moral/Conventional Distinction.” *Southern Journal of Philosophy* 49: 99-124
- Schapiro, Tamar. 2010. “Desires as Demands: How the Second-Personal Standpoint Might Be Internal to Reflective Agency.” *Philosophy and Phenomenological*

- Research LXXXI*(1): 229–36.
- Sensen, Oliver. 2012 (ed.). *Kant on Moral Autonomy*. Oxford: Oxford University Press.
- 2011. *Kant on Human Dignity*. Berlin: De Gruyter.
 - 2013. “Kant’s Constructivism”, in Bagnoli 2013a.
- Shemmer, Yonatan. 2012. “Constructing Coherence”, in Lenman and Shemmer 2012.
- Shpall, Sam. 2013. “Wide and Narrow Scope.” *Philosophical Studies* 163(3): 717-736.
- 2014. “Moral and Rational Commitment.” *Philosophy and Phenomenological Research* LXXXVIII(1): 146-172.
- Singer, Peter (ed.). 2017. *Does Anything Really Matter?: Essays on Parfit on Objectivity*. Oxford: Oxford University Press.
- Stern, Robert. 2010. “Moral Skepticism and Agency: Kant and Korsgaard.” *Ratio* 23(4): 453-474.
- 2013. “Moral Skepticism, Constructivism, and the Value of Humanity”, in Bagnoli 2013a.
 - 2014. “Divine Commands and Secular Demands: On Darwall on Anscombe on ‘Modern Moral Philosophy’.” *Mind* 123(492): 1095-1122.
- Strawson, P. F. 1968. “Freedom and Resentment.” *Studies in the Philosophy of Thought and Action*. London: Oxford University Press.
- Street, Sharon. 2006. “A Darwinian Dilemma for Realist Theories of Value.” *Philosophical Studies* 127: 109-166.
- 2010. “What is Constructivism in Ethics and Metaethics?” *Philosophy Compass* 5(5): 363-384.
 - 2012. “Coming to Terms with Contingency: Humean Constructivism about Practical Reason”, in Lenman and Shemmer 2012.
 - 2017. “Nothing Really Matters, but That’s Not What Matters”, in Singer 2017.
- Thompson, Michael. 2004. “What is it to Wrong Someone? A Puzzle About Justice”, in R. Jay Wallace, Philip Pettit, Samuel Scheffler, and Michael Smith (eds.), *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*. Oxford: Clarendon Press.
- Timmermann, Jens. 2005. “Good but Not Required?—Assessing the Demands of Kantian Ethics.” *Journal of Moral Philosophy* 2(1): 9-27.
- 2006a. “Value Without Regress: Kant’s ‘Formula of Humanity’ Revisited.” *European Journal of Philosophy* 14(1): 69-93.
 - 2006b. “Kantian Duties to the Self, Explained and Defended.” *Philosophy* 81(317): 505-530.

- 2007. *Kant's Groundwork of the Metaphysics of Morals: A Commentary*. Cambridge, UK: Cambridge University Press.
 - 2014. "Kant and the Second-Person Standpoint." *Grazer Philosophische Studien* 90: 131-147.
- Timmons, Mark. 2003. "The Limits of Moral Constructivism." *Ratio* 16(4): 391-423.
- Väyrynen, Pekka. 2013. "Grounding and Normative Explanation". *Proceedings of the Aristotelian Society, Supplementary Volumes* 87: 155-178.
- Waldron, Jeremy. 2012. *Dignity, Rank, and Rights*. Oxford: Oxford University Press.
- Wallace, R. Jay. 2001. "Normativity, Commitment, and Instrumental Reason." *Philosophers' Imprint* 1(3): 1-26.
- 2004. "Constructing Normativity." *Philosophical Topics* 32(1/2): 451-476.
 - 2007. "Reasons, Relations, and Commands: Reflections on Darwall." *Ethics* 118(1): 24-36.
 - 2012. "Constructivism About Normativity: Some Pitfalls", in Lenman and Shemer 2012.
 - 2013. "The Deontic Structure of Morality", in David Bakhurst, Brad Hooker, and Margaret Olivia Little (eds.), *Thinking About Reasons: Themes from the Philosophy of Jonathan Dancy*. Oxford: Oxford University Press.
- Wenar, Leif. 2013. "The Nature of Claim-Rights." *Ethics* 123(2): 202-229.
- Williams, Bernard. 1981. "Internal and External Reasons", in *Moral Luck: Philosophical Papers 1973-1980*. Cambridge, UK: Cambridge University Press.
- Wittgenstein, Ludwig. 2009. *Philosophical Investigations*, G. E. M. Anscombe (trans.), P. M. S. Hacker, and Joachim Schulte (ed. and trans.), 4th edition. Chichester: Wiley-Blackwell.
- Wolf, Susan. 2015. "Moral Obligations and Social Commands", in *The Variety of Values: Essays on Morality, Meaning, and Love*. Oxford: Oxford University Press.
- Wood, Allen W. 1999. *Kant's Ethical Thought*. Cambridge, UK: Cambridge University Press.
- 2008. *Kantian Ethics*. Cambridge, UK: Cambridge University Press.
- Zylberman, Ariel. 2014. "The Very Thought of (Wronging) You." *Philosophical Topics* 42(1): 153-175.
- 2017. "Two Second-Personal Conceptions of the Dignity of Persons." *European Journal of Philosophy* 25(4): 921-943.