

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/102839/>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Make No Exceptions *For Yourself* –
A Kantian Response to the Particularist Challenge

by
Irina Schumski

A thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy in Philosophy.

University of Warwick,
Department of Philosophy

July 2017

Table of Contents

Acknowledgments	iii
Declaration.....	iv
Abstract	v
Abbreviations	vi

Introduction	1
1. The Particularist Challenge to Kantian Ethics.....	2
1.1 Kantian Principles	2
1.2 A Trilemma for Kantians	4
1.3 Case-Scope vs. Agent-Scope: A Preview	9
2. A Remark on Methodology	10
3. Outline	12

PART ONE

Chapter 1: Three Responses to the Rigorism Objection	15
1. The Deontological Thesis	15
2. The Constitutive Deficiency Response.....	18
2.1 Defective Circumstances and Deficient Options.....	18
2.2 Different Games, Different Rules	22
3. The Conflicting Grounds of Obligation Response.....	26
3.1 A Kantian Theory of Pro Tanto Duties	26
3.2 Determinate Content and Indeterminate Import, or Vice Versa?.....	31
4. The Casuistical Questions Response.....	35
4.1 “ <i>Strictly Speaking</i> , This Is Not a Lie”	35
4.2 The Use and Abuse of Casuistry.....	40
5. Taking Stock: No Ideal World Without Exceptions	44

Chapter 2: Two Responses to the Particularist Challenge	47
1. The Particularist Challenge: A Trilemma.....	47
2. Korsgaard on Provisional Universality	52
2.1 Against Particularistic Willing: An Argument with Two Concessions.....	52
2.2 A <i>Reductio Ad Absurdum</i> of the <i>Two Standpoint View</i>	57
2.3 Two Responses on Korsgaard’s Behalf	59
2.3.1 The Voluntarist Response.....	60
2.3.2 The Implicit Awareness Response.....	62
3. Herman on Deliberative Presumptions	66
3.1 The Categorical Imperative as a Source of Deliberative Presumptions....	66
3.2 Two Independent Conditions of Defeat? An Inconsistency.....	70
3.3. Two Responses on Herman’s Behalf	75
3.3.1 The Competing Moral Concerns Response	76
3.3.2 The Deflationary Response.....	79
4. Taking Stock: A Comparison and a Note on Defeasibility	82

Chapter 3: The Inferentialist Response	85
1. Korsgaard’s Anti-Particularist Argument Revisited.....	85
2. The Agent-Scope Reading of <i>Universality</i>	88
2.1 Why Do We Need Concepts-Qua-Universal-Rules?	89
2.2 Two Kinds of Counterfactual Robustness	95
2.3 Defeasible Rules and Subsumption	100
2.4 The I That Judges and the I That Acts	105
2.5 Universal and Defeasible: The Agent-Scope Reading of <i>Universality</i>	108
3. Re-Interpreting Kant’s Ethics	111
3.1 Acting from Duty	112
3.2 Hypothetical and Categorical Imperatives.....	116
3.3 Applying the Formula of Universal Law	122
4. Taking Stock: The <i>Particularist Challenge</i> and Beyond	128

PART TWO

Chapter 4: Dissolving The Problem of Relevant Descriptions.....	132
1. The Problem of Relevant Descriptions	132
2. What Is the Problem of Relevant Descriptions? Three Misconceptions	135
2.1 First Misconception: A Puzzle Concerning Kant’s Notion of Maxims..	135
2.2 Second Misconception: The Opacity of the Human Heart.....	140
2.3 Upshot: Is There a Problem Left to Solve? The Subjectivizing Move ...	141
2.4 Third Misconception: Inadvertent Ignorance.....	144
2.5 The Correct Conception: A Flawed Moral Sensibility.....	146
3. Does Kant Need a Sensibility Theory? Clarifying the Task	147
4. A Kantian Standard of Moral Relevance? Two Unsuccessful Solutions.....	153
4.1 First Solution: Culturally Imparted Rules of Moral Salience	153
4.2 Second Solution: The Formula of Humanity.....	155
5. Taking Stock: A Dissolution of the Problem of Relevant Descriptions	160

Chapter 5: A Kantian Constructivist Response to Street	163
1. The Point of Contention between Humean and Kantian Constructivists.....	163
2. Street’s Objections to Kantian Constructivism.....	167
2.1 Korsgaard’s Argument for the Value of Humanity	168
2.2 Street’s Objection to Korsgaard’s Argument.....	171
3. Formal Metaethical Constructivism: A Kantian Variety	177
3.1 Street’s Taxonomy of Constructivist Positions	178
3.2 Kantian Formal Metaethical Constructivism: A Neglected Alternative .	183
4. Taking Stock: Conjuring Rabbits out of Hats.....	190

Bibliography	191
---------------------------	------------

Acknowledgments

Over the past years, I have been fortunate enough to enjoy the company and support of a group of exceptionally generous and talented mentors, colleagues, friends, and family members. I want to take this opportunity to say how grateful I am to have had them by my side.

First and foremost, I would like to express my heartfelt thanks to my supervisor, Fabienne Peter: for her very generous support, her ability to convey encouragement even when criticizing and challenging, for all her difficult questions and all her kind words. I could not have accomplished this work without her. I also want to thank several other members of the Philosophy Department at Warwick: Quassim Cassam for his support during my first year and for pushing me to branch out into unfamiliar philosophical territory, Johannes Roessler for his insightful comments on my work and his general encouragement and support throughout, Naomi Eilan for all her help and advice, and Felix Pinkert for keeping me sane when this thesis was nearing completion. I am also very grateful to Stephen Engstrom for his generous and insightful advice during my stay at the University of Pittsburgh and to Marcus Willaschek for inviting me to his Colloquium when I spent a few months as a visitor in Frankfurt.

I have been very lucky to be part of a supportive graduate community in which there was as much room for philosophical debate as there was for jokes and dreams and politics and all other important things in life. In particular, I would like to thank Alfonso Anaya, Adam Arnold, Matthew Dennis, Michele Giavazzi, Ivan Ivanov, Dino Jakušić, Tristan Kreetz, Richard Lambert, and especially Roberta Locatelli for their friendship and the many wonderful conversations, philosophical and otherwise. I have also received invaluable support from friends outside the Warwick Philosophy Department. Many thanks to Lara Choksey, Josh Eisenthal, Sophie Larnaudie, Timo Leidecker, Joe Saunders, Martin Schauss, Christoph Schuringa, and Garvan Walshe, among others.

Finally, I want to thank a few people whose influence on my development traces back much further: one of my very first philosophy teachers, Volker Dieringer, who opened my eyes to the intriguing depth of Kant's thought, my best friends, Leif Alter and Julia Neitzel, my whole family – and especially my parents – for their unwavering love, and patience and assistance, and Jochen Bojanowski for his invaluable support

and our countless conversations in which I learned so much, both about philosophy and about being a philosopher, the struggles it involves and the resilience it requires.

Declaration

This thesis is the result of original research that I carried out during my time as a doctoral candidate at the University of Warwick. This work is entirely my own. It has not been submitted for any other degree, neither at the University of Warwick nor elsewhere. With the exception of chapter 4, the material in it is unpublished.

Chapter 4 is a slightly modified version of my article “The Problem of Relevant Description and the Scope of Moral Principles”, which first appeared in the *European Journal of Philosophy* in May 2017. I would like to thank Wiley for the permission to reuse this material here.

Abstract

The primary aim of this thesis is to examine whether and how Kantian ethicists can accommodate the intuitions that motivate moral particularism: the intuition that the moral domain is very complex, that our moral obligations vary with circumstances in ways that are hard to codify, and that there are exceptions to most, if not all, moral principles that we can think of or formulate (Part One). The secondary aim of this thesis is to draw on the insights gained in the course of this investigation in order to contribute to the solution of two other problems that occupy contemporary Kantian ethicists (Part Two).

To begin with, I discuss and reject a number of existing attempts to account for the circumstance-dependence of our moral obligations within a Kantian framework. What all these attempts have in common is the assumption that, for Kant, a principle of duty is universally valid only if it is valid in all cases or situations. I call this the “Case-Scope Reading” of Kant’s conception of universal validity. When combined with the requirements that emerge from the challenge mounted by their particularist opponents, this reading throws Kantians on the horns of a trilemma. In response, I suggest that we should rethink this understanding of universal validity in light of the distinctive role and significance assigned to universal rules within Kant’s theory of objective knowledge. If we do, we are led to what I call the “Agent-Scope Reading” of Kant’s conception of universal validity: the view that a principle is universally valid if and only if it can be agreed to hold by all rational agents (qua judging subjects) and for all such agents (qua objects judged) *in the same circumstances*.

This reading has a number of advantages. Not only does it expose the trilemma mentioned above as merely apparent, it also helps Kantians to dissolve the so-called Problem of Relevant Descriptions and to defend Kantian Constructivism against its Humean critics.

Abbreviations

All references to Kant's writings are given by the following abbreviated titles, and all, except those to the *Critique of Pure Reason*, by the page numbers of the appropriate volume of the Prussian Academy Edition of Kant's *Gesammelte Schriften* (1902-). All references to the *Critique of Pure Reason* will be given by the page numbers of the first (A) and the second (B) editions. I will follow the Cambridge Edition translations of Kant's works (1991-) throughout.

CPR	Critique of Pure Reason (1781/87)
G	Groundwork of the Metaphysics of Morals (1785)
CPrR	Critique of Practical Reason (1788)
R	Religion Within the Boundaries of Mere Reason (1793)
MM	The Metaphysics of Morals (1797)
JL	Jäsche Logic (1800)

“So you are saying that human agreement decides what is true and what is false?” – It is what human beings *say* that is true and false; and they agree in the *language* they use. That is not agreement in opinions but in form of life.

(Wittgenstein 1953: §241)

Introduction

Moral life is full of subtleties. In most situations, we ought to be honest, but when we speak to someone on the verge of suicide, telling the truth might be a bad idea. The promises we make are generally binding, except if we were coerced into making them or if an emergency prevents us from following through. In a more or less just society, theft is wrong, but Robin Hood was a hero. In these and many other cases, it is of vital importance that we pay attention to the nuances and details of the moral situations that we are in, that we fully appreciate the ways in which they differ from one another, and that we do not shy away from the complexities that pervade the moral domain. Circumstances matter, and our moral thinking and moral theorizing should reflect their significance.

This is often thought to pose a problem for Kantian ethics, which is well known for its emphasis on universal laws and strict duties. According to Kant and Kantians, our moral obligations can be codified in universally valid principles of duty. But if the above examples are indeed indicative of the complexity of the moral domain and the subtle ways in which our obligations depend on the circumstances, then shouldn't we expect that any fruitful generalization would be riddled with exceptions? And should we not worry that insisting on principles that purport to cover *all* cases will make us blind to relevant details and differences? Consider one of the cases mentioned above.

Breaking a Promise to Help: You promised your friend to post their visa application, just in time for the deadline. But then, on the way to the post office, you see a child fall off their bike, hit their head, and pass out. You are pressed for time, so you must decide: either you carry on walking and keep your promise or you stop and help the child.

The duty to keep one's promises is a paradigmatic example of the kinds of duties that Kant is often thought to have regarded as absolutely strict and exceptionless. But in a case like the above, treating it as such would lead you to ignore the child's suffering, which certainly seems relevant. In a situation like this, we would think it permissible or even obligatory to break one's promise, and if you stuck to your principle, if you insisted on keeping your word, then we would accuse you of insensitivity, of ignoring a morally significant aspect of the situation. So perhaps there is no duty to keep one's promises *per se*. Perhaps the relevant duty is rather more complicated. But, again: if

the above cases are any indication, such duties would have to be very complicated indeed.

My primary aim in this thesis is to examine whether and how Kantian ethicists can accommodate the circumstance-dependence of moral obligations. My secondary aim is to explore what the interpretation that I present as a solution to this problem can contribute to the solution of other problems that occupy contemporary Kantian ethicists. In particular, I will discuss the so-called Problem of Relevant Descriptions (ch. 4), and the dispute between Humean and Kantian Constructivists (ch. 5). In the rest of this introduction, I want to first provide a more precise characterization of the challenge that the circumstance-dependence of our moral obligations poses for Kantians (sect. 1). Among other things, I will explain why it is a special case of the broader challenge that moral particularists mount against their generalist opponents, and indicate the outlines of my solution and its wider applications. Then I will briefly remark on my methodology (sect. 2) and outline the course of my argument (sect. 3).

1. The Particularist Challenge to Kantian Ethics

1.1 Kantian Principles

When critics argue that Kant's emphasis on principles leads him to disregard the role of circumstances (see e.g. Constant 1964: 63-71, Anscombe 1958), there are different targets that they might have in mind.

First, they might be objecting to either one of the two *kinds of principles* that Kant is known for advocating: his "*supreme principle of morality*" (G 4:392), the Categorical Imperative,¹ or the specific principles of duty that are thought to follow from it. Of course, to some extent, the two stand and fall together: the counterexamples devised to discredit the Categorical Imperative cast doubt on whether adequate principles of duty can be established on its basis, and the counterexamples developed to discredit specific principles of duty raise questions about how the Categorical Imperative is to

¹ Kant's aim in the *Groundwork of the Metaphysics of Morals* is to search for and establish this supreme principle. As is well known, he offers several formulations of the Categorical Imperative: the Formula of Universal Law (G 4:421), the Formula of Humanity (G 4:429), the Formula of Autonomy (G 4:432), and the Formula of the Kingdom of Ends (G 4:439). For our purposes, the most relevant of these is the Formula of Universal Law because this is where Kant's commitment to universally valid principles of duty is usually thought to originate. It will come up at various points, and its content and application will be discussed in ch. 3, sect. 3.3. The relationship between the formulas lies beyond the scope of this thesis, but I will briefly discuss the Formula of Humanity in ch. 4, sect. 4.2.

be applied.² So it would be ill-advised to focus on one to the exclusion of the other. However, if one's central concern is that Kant's theory overgeneralizes, that it treats cases that are importantly different as if they were alike, then it would make sense to either object to his allegedly exceptionless principles of duty directly or otherwise to the Categorical Imperative *as a source of* such principles. For if one didn't think that the Categorical Imperative yields any exceptionless principles of duty, then the overgeneralization worry would not arise, or at least not obviously.³ This is why the first part of this thesis will focus on specific principles of duty, such as the duty to keep one's promises, and counterexamples devised to undermine them, such as the above *Breaking a Promise to Help* case. But, of course, the Categorical Imperative and the notions associated with it (e.g. "universalisability", "maxim", etc.) will always be there in the background, informing our thinking about how Kant's specific principles of duty should be conceived. In the second part of the thesis, when we turn to other debates in Kantian ethics, the Categorical Imperative will take centre stage.

Second, they might be assuming that Kant's principles of duty are designed to play either one of two different *roles*.⁴ In the philosophical tradition, moral principles have been taken to play two distinct roles: the practical role of guiding action and the theoretical role of standards that explain why things have certain moral features or statuses, e.g. why an action is impermissible (Timmons 2001: 3-6). Accordingly, the concern might be, on the one hand, that Kant's principles of duty are bad guides: that using them is not conducive to good decision making. If a strict prohibition of promise-breaking was among his principles, for example, it would seem reasonable to complain that this principle lets us down when we need it most, namely in difficult situations, such as the *Breaking a Promise to Help* case. On the other hand, one could make a case that Kant's principles of duty fail as theoretical standards: that, far from helping us to understand and systematize the moral domain, they actually stand in the way. This, again, seems like a reasonable complaint, at least if the moral domain

² Tamar Schapiro is right when she notes that many Kantians have tried to exploit this link (2006: 35): many have tried to refute the objection that Kant's principles of duty are not sensitive to the role of circumstances by arguing that his critics have misapplied the Categorical Imperative. This is true of Barbara Herman's response, which I will discuss in ch. 2, sect. 3, but also of others, e.g. Hill 1991 and 2000: 33-55.

³ It could still arise, for example, if one thought that the Categorical Imperative is an algorithm for moral deliberation, which leads us to ignore the differences between situations. In ch. 3, sect 3.1, I will agree with Herman that, if the Categorical Imperative *was* such an algorithm, it would not do so.

⁴ The same question arises with regard to the role of the Categorical Imperative. I will come back to this issue in ch. 2, sect. 3.2, and ch. 4, sect. 2.3, 4.2.

is as complex as we imagine it to be. By the end of ch. 3, I hope to have shown that this distinction between roles is at best misleading. However, until then, I will focus on the theoretical role (if only because the literature on this role is more extensive),⁵ and leave aside whether Kant's principles of duty are suitable guides. In particular, following the widely shared view that Kant's principles of duty determine the deontic statuses of actions, I will examine whether they are fit to figure in explanations of why actions are permissible, impermissible, or obligatory.

Now that we have identified the target of the challenge, namely Kant's view that there are universally valid principles of duty which explain why our actions have the deontic statuses they have, we can turn to the challenge itself. What exactly is the nature of this circumstance-dependence of moral obligations, which Kantians should be able to account for?

1.2 A Trilemma for Kantians

In contemporary debates, it is often argued that there are broadly two ways in which circumstances can have a bearing on our moral obligations. We can refer to them as “undermining” and “outweighing”. Let us assume that we have found a case where a certain feature, e.g. being a promise-breaking, makes an action impermissible. This does not mean that every promise-breaking is impermissible because, in a different case, there might be circumstantial features that undermine the deontic import of promise-breaking (e.g. the fact that you were coerced into promising) or features that have a different deontic import that outweighs the import of promise-breaking (e.g. the fact that, by breaking your promise, you can help a child in need). Contemporary metaethicists associate these two modes of defeat⁶ with two distinct levels of import. Whether there are defeaters in the undermining sense is taken to determine whether a feature, e.g. being a promise-breaking, makes its usual contribution to the overall deontic status of an action or not. Whether there are defeaters in the outweighing

⁵ With a few exceptions (e.g. McKeever and Ridge 2006: 196-223, Väyrynen 2008), most contributions to the generalism-particularism debate seem to focus on the theoretical role.

⁶ As Garrett Cullity points out (2013: 9), it is common to use the term “defeat” to cover both of the modes mentioned above. However, sometimes philosophers use this term to refer to a version of the undermining mode only. I will stick to the common usage. It is also worth noting that our two modes are actually broader categories that encompass a range of more specific modes. For discussions of these categories and the more specific modes contained under them see e.g. Sinnott-Armstrong 1999: 1-12, Dancy 2004: 13-52, and Schroeder 2011: 328-44.

sense, by contrast, is taken to determine whether the contribution made by promise-breaking suffices to make your action *overall* impermissible.

In light of this, the challenge to Kantians *could be* framed as a twofold challenge. First, Kantians must be able to account for the possibility of defeat by features of the circumstances. Second, their account must be able to correctly discriminate between cases of undermining and cases of outweighing. The second challenge is a challenge that Kantians can hardly meet. For no matter how closely we look, there just doesn't seem to be any basis for a distinction between two modes of defeat in Kant. This explains why most Kantians deny the very existence of a contributory level and the possibility of outweighing.⁷ So, in order to give Kantian ethics a fair chance, we shall set the benchmark a bit lower: we shall limit ourselves to asking whether Kantians can account for cases of defeat by circumstantial features *at all*, and ignore whether they categorize them correctly. In practice, this will mean that we will not take issue with the fact that most of the Kantian accounts discussed below reduce outweighing cases to undermining cases.⁸

With this in mind, we can now turn to a thesis that contemporary metaethicists introduce to capture the possibility of defeat in the undermining sense and adapt it to our purpose by extending its scope so that it covers both modes of defeat. The thesis in question is “holism”. In its original form, it is a thesis about contributory reasons (Dancy 1993: 60-6, 2004: 73-8), but in line with what we decided, we shall consider a modified version of this thesis that concerns overall deontic statuses.

Holism about Deontic Statuses: A feature or set of features that makes an action have a certain (overall)⁹ deontic status in one situation may not play the same role in other situations.¹⁰

⁷ Thus, Dancy is right when he observes that Kantians are generally “suspicious of the very notion of a contributory reason, and ... try to capture the relevant phenomena in other ways” (2004: 67). There are some exceptions, however. Some Kantians (see ch. 1, sect. 3 and ch. 2, sect. 3.3.1) reduce cases of undermining to cases of outweighing, and not vice versa. But even they do not recognise more than one mode of defeat – they simply opt for the other one.

⁸ In this I include my own account. I give some reasons for thinking that we may not actually need the notion of outweighing or a contributory level in ch. 2, sect. 3.3.1. But the reasons given are far from conclusive.

⁹ For the sake of brevity, I will omit the qualification “overall” in what follows. Unless I say otherwise “deontic status” means “overall deontic status”, and the same goes for “impermissible”, “obligatory”, “wrong”, and so forth.

¹⁰ This formulation deviates from standard formulations (see e.g. Dancy 2004: 7) in that it omits the phrase “or may play the opposite role”. The reason for this omission is partly strategic, but only partly. In ch. 3, I will argue for a view that is in some respects similar to Robert Brandom's inferentialist view

For example, the fact that your action is a promise-breaking may make it impermissible in many cases, specifically, in standard cases where you weren't coerced and no weightier concern is at stake, and so forth. But it may not play the same role in other cases, specifically, in exceptional cases, e.g. when you were coerced or when a child needs your help. This, then, is the thesis that Kantians ought to accommodate. Now, why is this a problem?

According to so-called moral particularists, holism casts doubt on the viability of traditional moral theories which attach great importance to moral principles (see e.g. Dancy 1993: 60-2, 2004: 73-93, Little 2000). In fact, particularists believe that holism supports the following claim, which lies at the heart of their own position.

Moral Particularism: Even if there are some true moral principles, nothing in moral theory or moral practice hinges on them – we can explain the obtaining of moral statuses and act well without them.¹¹

It is nowadays widely agreed that the argument leading from holism to particularism is at best indirect (see e.g. Little 2000: 277, McKeever and Ridge 2006: 26-7). It is indirect because holism does not say how complex the moral domain *actually* is: how many circumstantial features actually do make a difference to various obligations, and how complicated the connections really are. All it says is that anything *could* make a difference. If there wasn't all that much complexity, then it might be possible to capture our moral obligations in principles that are both true and fruitful (principles with a few unless-clauses, say), but this, particularists say, would be a "cosmic accident", and it does not seem to be true of our morality as it is (Dancy 2004: 82). Let us grant that the moral domain is very complex indeed. In that case, all moral generalists,¹² including Kantians, face a challenge, which arises because the demand that principles should be valid in all situations, both actual and counterfactual, is in tension with the

(2000: 87-92). Like him, I appeal to non-monotonic inferences. As Dancy points out (2004: 8-9), this kind of view cannot explain how a feature that plays a certain role in one case could, *on its own*, play the opposite role in another. In response, I would want to dispute his claim that this is how it must be, intuitively speaking. I have no intuition to this effect.

¹¹ There are different versions of particularism, but this is the version that I shall focus on. It comes closest to the version advocated by Jonathan Dancy in *Ethics Without Principles*.

¹² For current purposes, we can characterize moral generalism quite broadly, namely as the view that *Moral Particularism* is false.

demand that they should be able to serve as guides and/or theoretical standards.¹³ What does this mean for Kantians, specifically?

I will argue that, for Kantians, this challenge takes the form of a trilemma. This is because, in order to meet the challenge to which the particularist critique gives rise, they have to accommodate three *prima facie* incompatible claims, which can be seen as desiderata on a satisfactory version of Kantian ethics.

Universality: The features in virtue of which actions have their deontic statuses can be captured in universally valid principles of duty.

Asymmetry: In standard circumstances, an action of a kind that is usually¹⁴ impermissible is impermissible solely in virtue of the features that make it count as an action of that kind (i.e. *not* in virtue of the fact that defeaters are absent).

Complexity: In exceptional circumstances, an action of a kind that is usually impermissible can be obligatory or permissible, and when it is, then this is partly in virtue of features other than the ones that make it count as an action of that kind (i.e. in virtue of the fact that defeaters are present).

It is generally thought that *Universality* is one of the foundational doctrines of Kantian ethics.¹⁵ *Asymmetry* captures the idea that an explanatorily fruitful principle should not make reference to features that are irrelevant in the case to be explained. *Complexity* is an articulation of the idea from which we began: the idea that our moral obligations depend on and vary with circumstances. These three desiderata seem to be in tension because it is, on the face of it, impossible to satisfy more than two of them at once. The tension can be summed up as follows.

The Particularist Challenge: Kantians have to explain how it is that our explanations of exceptional cases make reference to defeaters (*Complexity*) while our explanations of standard cases don't (*Asymmetry*), even though both explanations are meant to

¹³ For a discussion of these two demands see Väyrynen 2011: 248-9.

¹⁴ When I come back to these desiderata in ch. 2, sect. 1, I will use the more technical term "default" in order to refer to the deontic import that certain features usually have and the deontic statuses that they usually bestow. For discussions of default reasons see e.g. Lance and Little 2006, Dancy 2004: 69, 111-7, Väyrynen 2004, McKeever and Ridge 2006: 46-75.

¹⁵ Of course, as we will see, everything depends on how universal validity is understood. We will come back to the question of how *Universality* is thought to follow from what Kant says in ch. 1, and in ch. 2 and 3 we will consider different readings of it.

appeal to one and the same explanatory principle, that is, to the same principle of duty (*Universality*).

To illustrate, Kantians have to give an account of how it is that the deontic status of the relevant actions in both the *Breaking a Promise to Help* case and a standard case are meant to be explained by appeal to the same principle, if, in the former case, we have to refer to the fact that there is a passed out child, whereas in the latter case, we do not need to (and should not) refer to the fact that there is no passed out child. In ch. 2 I will argue that, in their attempts to deal with this tension, Kantians have found themselves thrown on the horns of a trilemma.¹⁶

It is worth noting that the combination of *Universality* and *Asymmetry* seems to lead to a view that resembles more “orthodox” forms of Kantianism.¹⁷ For if one combines the view that there is one principle covering all cases with the view that the relevant principle is the kind of simple principle that works in standard cases, then one seems forced to *treat* all cases as if they were standard cases. And this, it seems,¹⁸ is exactly what one is forced to do if one takes seriously the implications of Kant’s rigoristic¹⁹ texts and remarks – above all, his essay “On a Supposed Right to Lie from Philanthropy” (8:423-30), where he responds to an infamous example put forward by Benjamin Constant.²⁰

Murderer at the Door: A murderer knocks at your door and asks about the whereabouts of his victim, who is hiding in your house. If you are honest, you put the victim’s life at risk, but if you opt for a lie, then you violate the duty not to lie.

In his response to Constant, Kant argues that, even in such an extreme case, lying is strictly impermissible.²¹ The broader implication of his rigoristic assessment of cases

¹⁶ For now, I will limit myself to highlighting the tension. The horns of the trilemma will be discussed in ch. 2, sect. 1.

¹⁷ I say “resembles” because orthodox Kantians would reject our talk of action kinds that are “usually impermissible”. They would insist that they are always impermissible.

¹⁸ In ch. 1 I will consider three attempts to resist this pressure.

¹⁹ This is only one of many rigorisms that Kant’s critics have accused him of. Further views that have been dubbed rigoristic are his monism about value, his view that there are no moral conflicts and no moral residues, his view that the motive of duty is the only morally worthy motive there is, and finally, the two views that he himself calls rigoristic (though, of course, praisingly), namely the view that there are no morally neutral actions or characters (R 6:22). For a discussion of Kant’s various rigorisms see Timmermann 2001.

²⁰ Constant’s example occurs in a pamphlet entitled “On Political Reactions”, which was published in 1797. In the relevant section, Constant argues against excess in the application of principles.

²¹ In the eyes of his critics, this is proof of his outrageous rigorism. Kantians have spilled considerable amounts of ink trying to defend him. We will consider some of these attempts in ch. 1.

such as the above seems to be that all principles of duty take the form “Actions of kind A have the deontic status D”, e.g. “Lying is impermissible”. If this was indeed a commitment that Kantians could not flout, where would this leave them in relation to the task of accounting for the role of circumstances?

I think it would expose them to a more basic objection.

Rigorism Objection: Kant cannot account for the fact that there are exceptions to principles of the form “Actions of kind A have the deontic status D”.

This is the objection from which I want to begin. It will give me an opportunity to engage with the responses of Kantians who are more orthodox in the sense that they want to uphold the view that certain kinds of actions are strictly impermissible. From there I will move on to more unorthodox Kantian accounts, which allow that Kant’s principles of duty might be more complex in structure and content than the principles we considered above.

1.3 Case-Scope vs. Agent-Scope: A Preview

My response to the *Particularist Challenge* hinges on a distinction between two readings of *Universality*. According to the standard reading, which I call “Case-Scope Reading”, a principle of duty is universally valid only if it is valid in all cases or circumstances. If this is the sense in which Kant’s principles of duty are universal, then the worry that his moral theory is not sensitive to important differences between situations and that it overgeneralizes is appropriate indeed. However, in what follows I will try to show that there are in fact no good reasons for thinking that this is the notion of universal validity that Kant needs or wants.²² I will suggest that, if we pay attention to the kinds of considerations that lead Kant to insist on universal laws in the first *Critique*, we will find that *his* concern is not a concern with our ability to generalize across cases at all. What Kant is concerned with, instead, is the question of how it is possible for us to acquire objective knowledge of a certain domain. For this to be possible, he believes, we need to think of the relevant domain as being governed by universal laws. I argue that these findings point to a different reading of *Universality*, which I call the “Agent-Scope Reading”. According to this reading, a principle of duty is universally valid iff

²² It would be more accurate to say that this is not the notion that he *should* want, given his most basic commitments. One of the results of my discussion will be that, occasionally, Kant misunderstood the implications of his own views.

it can be agreed to hold by all rational agents (qua judging subjects) and for all such agents (qua objects judged) *in the same* circumstances. Since, on this reading, universal validity does not require validity in all cases or circumstances, since it is assessed relative to a particular set of circumstances, the overgeneralization worry does not arise.

This proposal has interesting implications for contemporary debates between Kantian ethicists and their opponents.²³ I will discuss two of these implications in the second part of the thesis. What I will argue, specifically, is, first, that the Agent-Scope Reading does away with the notorious problem of how to ensure that our maxims contain relevant descriptions of our actions and situations (see ch. 4). Second, I will argue that it allows us to understand the difference between Kantian and Humean Constructivism as a difference in what the form of reasons is taken to be, not merely a difference in what reasons we are taken to have, content-wise (see ch. 5).

2. A Remark on Methodology

This thesis is an attempt to reinterpret the arguments and ideas of a historical figure whose work has shaped the thinking of generations of moral philosophers, including our contemporaries, in a way that advances a current debate in metaethics. As such, its success depends on resisting two temptations that one is at risk of succumbing to when approaching a historical text with a philosophical mindset. On the one hand, there is the temptation to be so exegetically diligent, so rigorous and careful in one's attempt to arrive at a proper understanding of the text, that one is never quite ready to commit, never quite ready to settle on a reading from which a critical engagement could begin. This is the temptation of excessive reverence. On the other hand, there is the temptation of jumping to conclusions, of imposing contemporary distinctions, taxonomies, and argumentative patterns onto the text without questioning whether they fit, and thinking that one is ready to criticize and reject (or, indeed, defend) the views of the author, when one is not. This is the temptation of presumptuousness.

Both tendencies are well familiar and often remarked on, and the commonplace counsel is to aim for a healthy balance. This does seem right, but it is also somewhat vacuous. How does one determine when a healthy balance is achieved? Surely, this depends on one's objectives: on why one wants to engage with the historical figure in

²³ Its implications extend to discussions of other aspects of Kant's ethics, too. These will be explored in the final section of ch. 3, after I have presented my argument for the Agent-Scope Reading.

question and what one is trying to accomplish. If no aims are specified, the healthy balance counsel is vacuous. It seems to me that a similar vacuousness attaches to the now conventional distinction between *Kant's ethics* and *Kantian ethics*. These labels do not seem to be associated with any distinction in objectives. Rather, they are used to classify projects merely on the basis of how closely the author sticks to the letter of Kant's texts.²⁴ In my view, this is not a fruitful distinction. For this reason, I will not try to locate my project on the Kant's-ethics-Kantian-ethics spectrum.²⁵ Instead, I want to briefly explain how my reason for engaging with Kant's texts suggests a certain mode of engagement, and how a commitment to this mode will help me to strike a balance between faithful and liberal interpretation that is right *for my purposes*.

The larger goal to which the present thesis can hopefully make a contribution is to overcome some of the impasses that the generalism-particularism debate seems to have reached. Crudely speaking, one impasse seems to be the following. The debate as a whole has emerged from the particularists' complaint that principle-based moral theories tend to overgeneralize: that they tend to treat cases that are importantly different as if they were alike. This has shaped how both parties think of what is at stake. In many cases, both sides frame the dispute as a dispute about whether and to which extent generalization is possible and necessary. But with this being the implicit assumption, it seems that any attempt to defend generalism is almost set up for failure. After all, no one is prepared to deny that the moral domain is very complex and that, nonetheless, we manage to cope with this complexity. So the claim that we need to generalize in order to reduce complexity and facilitate understanding does not cut any ice. This is where I suggest we should turn to Kant and his reflections on why principles are needed.

However, in order to find what we are looking for, namely different reasons for thinking that moral principles are required, we must engage with the passages where these reasons are given on their own terms. We have to suspend our contemporary assumptions about the main point of contention that divides the friends and enemies of moral principles and be open to the possibility that terms like "universal validity" don't mean what we take them to mean. This is the approach that I will adopt in ch. 3, where I present my own response to the *Particularist Challenge*. In this chapter, I will

²⁴ Of course, the above assessment does not do justice to each and every account of the distinction that has been proposed. For a critical discussion of the distinction, see Wood 2008: 1-4.

²⁵ Accordingly, I shall use the terms "Kant's ethics" and "Kantian ethics" interchangeably.

turn to one of the centrepieces of Kant's first *Critique*, the Transcendental Deduction, and ask *why* Kant thought that universally valid laws were needed to acquire objective knowledge of nature. This investigation will bring to light that the kinds of principles that the possibility of objective knowledge presupposes don't have to generalize (and won't *overgeneralize*) across cases and situations. As such, it will point to a different reading of the *Universality* desideratum: the Agent-Scope Reading that I sketched above (sect. 1.3).

It will be a result of my discussion – and not one of its presuppositions – that Kant's rigoristic remarks, including his response to Constant, should be ignored, that, here, we should depart from his texts. This departure is justified because it allows us to appreciate the distinctive character of his arguments for why we need principles, thus helping us to overcome impasses that we have reached in our current debate.

3. Outline

In the first part of the thesis, I discuss and reject a number of existing Kantian responses to the *Rigorism Objection* and the *Particularist Challenge*. I then propose an alternative response.

In the first chapter, I look at three Kantian responses to the *Rigorism Objection* (see sect. 1.2 above). These responses are orthodox in the sense that they try to account for exceptions to Kant's principles of duty without abandoning the claim that certain kinds of actions are intrinsically impermissible (Tamar Schapiro refers to this claim as the "*Deontological Thesis*"). Each of these responses appeals to non-ideal conditions: to the idea that exceptional cases are in some sense corrupted instances of the principle of duty in question. The hope behind this appeal is that putting the blame on non-ideal conditions will allow Kantians to leave the principles themselves untainted. My discussion reveals that this hope is vain: that the cases in question are not corrupted, and hence that, even in an ideal world, there would still be exceptions. The appeal to non-ideal conditions is unsuccessful because, in each case, the claim that the relevant cases are corrupted is secured either by leaving the content of the relevant principles of duty unspecified or by blurring the lines between different action kinds. On this basis, I conclude that the content and structure of Kant's principles of duty have to be more complex, as unorthodox Kantians argue. However, in

thinking about what this more complex structure and content could be, they face the *Particularist Challenge* (see sect. 1.2 above).

In the second chapter, I discuss two quite unorthodox Kantian responses to the *Particularist Challenge*. I begin by introducing the three prima facie incompatible desiderata that a Kantian response to this challenge must satisfy: *Universality*, *Asymmetry*, and *Complexity*. I then explore the resulting trilemma and each of its horns. With the trilemma in mind, I turn to Christine Korsgaard's and Barbara Herman's responses, which can be understood as attempts to escape the trilemma. Both appeal to the idea that Kant's principles of duty are universal and nonetheless defeasible and, in both cases, the strategy for escaping the trilemma is a division of labour: their views consist of two distinct elements, each of which satisfies two of the three desiderata. This strategy fails, however, because, in both cases, the overall view is fraught with inconsistencies. In the end, Korsgaard and Herman have to choose from two problematic options: a view that flouts objectivity and a view according to which our principles of duty contain long lists of unless-clauses.

In the third chapter, I present my own response to the *Particularist Challenge*. In line with the methodological commitments outlined above (sect. 2), I first consider Kant's famous argument for the claim that we could not have any knowledge of the objects of experience if we did not synthesise the manifold of sense in accordance with concepts that serve as universal rules. With this in mind, I show that rules can fulfil the requisite function without holding in all cases or circumstances. I argue that it is enough if they are robust across a certain range of circumstantial variations, namely variations in what I call "subjectively relevant circumstances": aspects of the circumstances that merely *seem* relevant to a certain subject because they are under the influence of some distorting factor, but that aren't actually relevant to how things go with the object. When carried over to the moral domain, this distinction points to the Agent-Scope Reading of *Universality*, according to which a principle is universal iff it can be agreed to hold by all rational agents (qua subjects) and for all such agents (qua objects) in the same circumstances. I will argue that this novel reading has the advantage of dissolving the trilemma. In addition, it supplies an important distinction between legitimate exceptions that are called for by the circumstances, and illegitimate exceptions that agents might be inclined to make for themselves. I conclude the

chapter by considering how the Agent-Scope Reading bears on discussions of other elements of Kant's ethics.

In the second part of the thesis, I explore what, if anything, the Agent-Scope Reading can contribute to the solution of other problems that exercise contemporary Kantian ethicists.

The fourth chapter focuses on the so-called Problem of Relevant Descriptions. This problem goes back to G.E.M. Anscombe's famous complaint that Kant's "rule about universalizable maxims is useless without stipulations as to what shall count as a relevant description of an action with a view to constructing a maxim about it" (1958: 2). I begin by clearing up several misunderstandings of the problem. In doing so, I bring out that the problem consists in the fact that Kant's Formula of Universal Law seems to stand in need of an independent account of moral sensibility which does not render the formula itself superfluous. However, if we presuppose the Case-Scope Reading of *Universality*, there can be no such account. The only way in which the problem can be resolved, I argue, is by adopting the Agent-Scope Reading. For only then can we see the Formula of Universal Law itself as playing the role of a Kantian account of moral sensibility.

In the fifth chapter, I defend Kantian Constructivism against Sharon Street's Humean criticism. To begin with, I briefly outline and endorse Street's objections to Korsgaard's early argument for the value of humanity. I argue that this argument fails because Korsgaard is trying to combine a Humean view of what it is to be a reason with a Kantian view of the kind of strong objectivity that attaches to moral reasons. That she combines these inconsistent views, I suggest, can be explained by the fact that she misunderstands the concern that underlies Kant's Categorical Imperative. In her view, his concern is a concern for a certain stability of character – a concern that the Case-Scope Reading of his universalisability requirement is well suited to address. I argue that Kantian Constructivism can be rescued if we acknowledge that Kant's primary concern is a concern for objectivity. For once we see his project in this light, it becomes apparent that a properly Kantian view of what it is to be a reason has to introduce the point of view of other agents right from the start. This is exactly what the Agent-Scope Reading of his universalisability requirement does.

PART ONE

Chapter 1

An Ideal World Without Exceptions?

Three Responses to the Rigorism Objection

1. The Deontological Thesis

Kantians have had a tendency to downplay the significance of the *Rigorism Objection*, a tendency to treat it as a worry that does not really threaten the core of Kant’s moral theory. Among other explanations, they have suggested that the source of Kant’s perceived rigorism lies in matters as shallow as the bad temper of an old man (Paton 1953: 202), a lack of exegetical attention to the context in which Kant uses rigoristic language (Sedgwick 1991, Wood 2011), a failure to distinguish between two distinct notions of unconditionality on the part of his commentators (Singer 1954: 581), and misunderstandings of how to apply the Categorical Imperative (Herman 1993: 132-83, Korsgaard 1996a: 133-58, Hill 2000: 33-55).²⁶ However, according to Schapiro, there is an important point that all these downplaying explanations miss. In tracing the root of the problem to features that are *specific* to Kant’s work and the history of its reception, they forget that “the basic worry arises as readily in response to traditional deontological theories as it does in response to Kant’s theory” (2006: 35). In Schapiro’s view, this oversight has prevented Kantians from appreciating that the real source of the *Rigorism Objection* lies deeper, in a commitment that all deontological theories have in common: the view that impermissibility and obligatoriness are intrinsic features of certain kinds of actions. Focusing on impermissibility, she articulates this view as follows:²⁷

²⁶ As I noted in sect. 2 of the introduction, my main question is whether Kant’s most fundamental commitments expose him and his followers to the *Rigorism Objection* and to the trilemma to which the *Particularist Challenge* seems to give rise. For this reason, I am not concerned with any explanations of why he said the rigoristic things that he said *except* explanations that connect these rigoristic statements to his most basic commitments. This is true only of the third and fourth items on the above list. I will return to the third in ch. 3, sect. 3.1, and to the fourth in ch. 2, sect. 3, ch. 3, sect 3.3, and in ch. 4.

²⁷ The below formulation follows Schapiro (2006: 34), except in that she speaks of “categorical moral rules” and of “wrongness”. I have dropped the designation “categorical” because, as we will see in sect. 5, this has no bearing on whether circumstances can make a difference. The reason why I chose to speak of “impermissibility” is that the three-way distinction between permissibility, impermissibility, and obligatoriness is more nuanced than the two-way distinction between rightness and wrongness. Note that the *Deontological Thesis* could also be rephrased in terms of obligatoriness. This might raise

Deontological Thesis: The actions prohibited by Kant's principles of duty are impermissible in themselves, i.e. impermissible in virtue of the features that make them count as the kinds of actions they are. If an action is impermissible in itself, then variations in external circumstances cannot make it permissible or obligatory.²⁸

For example, if deception is inherently manipulative, and manipulating other people is the very essence of treating them as a mere means (as some have argued, see e.g. Herman 1993: 154-5, Korsgaard 1996a: 347), then Kantians should not regard the impermissibility of deception as in any way dependent on the circumstances. For, at least on the face of it, circumstances can make a difference only to the *consequences* of acts of deception, and, according to Kantians and other deontologists, these ought to be discounted.

It does not take much to see that this sort of view leads to rigoristic conclusions. If the very same features that make an action count as an act of deception also make it impermissible, and necessarily so, then the principle that acts of deception are impermissible must be exceptionless: it must be valid in all situations. Similar arguments can be made about other kinds of actions, e.g. lying, stealing, and breaking promises. Given this pressure towards rigorism, anti-rigoristic Kantian ethicists seem forced to either reject Schapiro's claim that Kantians are committed to the *Deontological Thesis*, or to show that the pressure can be resisted: that intrinsically impermissible action kinds can have permissible instances.²⁹ My aim in this chapter is to arrive at the conclusion that Kantians should go with the first option.

If I wanted to argue for this conclusion directly, then I would seek to refute the reasoning which leads Schapiro to attribute the *Deontological Thesis* to Kant. Her ar-

questions about positive and negative duties in Kantian ethics, however. We will touch on some of these questions in sect. 3 and 4 below.

²⁸ When discussing the possibility of defeated prohibitions in this and the next chapter, I will speak of the actions in question as "permissible or obligatory" and thus disregard the difference between cases in which the import of the circumstances is such that the action becomes permissible and cases where it is such that the action becomes obligatory. In contemporary debates, the difference between them is often captured by distinguishing between different modes of defeat: "outweighing" vs. "undermining" (see introduction, sect. 1.2).

²⁹ In fact, there is a third option: Kantians could also argue that, for the purpose of moral assessment, actions have to be picked out by more fine-grained descriptions, which take account of features that common parlance would usually classify as circumstances. This strategy is in some ways analogous to the strategy pursued by some contemporary generalists who reject holism about reasons by arguing that enablers and defeaters are, in fact, part of the reasons themselves (e.g. Raz 2006: 108-13). In both cases, the dispute is difficult to adjudicate because each party presupposes their own intuitions about what constitutes a fully specified action kind or a complete reason (Väyrynen 2006: 714, 716). For that reason, I shall put this third option to one side.

argument for this attribution, which is quite brief, takes as a starting point Kant's view that consequences cannot bear on the normative force of a categorical demand.³⁰ Features of the circumstances, such as others' wrongdoing, she says,

can only affect us by altering the causal nexus within which we act. That nexus determines the consequences of our own lawful action. But to think that the value of our own lawful action depends upon its consequences is ... to think of the law as a hypothetical imperative. Given that the moral law is *categorically* rather than hypothetically binding, others' misconduct cannot undermine its normative force. (Schapiro 2003: 330; my emphasis)

I will return to this argument in the final section. For now, I want to consider and criticize three responses to the *Rigorism Objection* that aim to reconcile the idea that Kant's principles of duty admit of exceptions with the *Deontological Thesis*. I want to proceed in this way because it gives me an opportunity to think about how more "orthodox" Kantians handle this issue. For whether or not we share their ambition to stick to the letter of Kant's texts, it seems that we should judge their responses to the *Rigorism Objection* on their own merits.

What all these responses have in common is that, in one way or another, they attribute the fact that principles of duty admit of exceptions to non-ideal conditions: to partial compliance, to natural and historical limitations,³¹ circumstantial contingencies such as the availability of means, or the phenomenon of loose speech. The point of arguing that, in an ideal world, where these non-ideal conditions would not obtain, Kant's principles of duty would be strictly exceptionless, is to show that there is no problem with the principles *themselves*, and that the problem arises only when we try to apply them. My objection is that these responses preserve the appearance of an ideal world without exceptions only at the cost of either leaving unspecified what a given duty demands or blurring the lines between different duties and action kinds. On this basis, I conclude that the *Rigorism Objection* does cast doubt on Kant's principles of duty, *at least* if they are taken to have the content and structure that orthodox

³⁰ For an explicit statement of this argument, we have to go back to an earlier paper of Schapiro's. In her 2006 paper, where the *Deontological Thesis* is first introduced, she alludes to this argument only indirectly, by sketching and endorsing Herman's and Korsgaard's accounts of the wrongness of deception as rooted in its inherently manipulative character (2006: 36-7).

³¹ The contemporary discussion of ideal and non-ideal theorizing is indebted to John Rawls. In *A Theory of Justice*, Rawls introduces the condition of partial compliance as the basis for his distinction between ideal and non-ideal theory (1999: 7-8), and natural and historical limitations as circumstances of justice (1999: 109-10, 215-6).

Kantians take them to have: if they present certain action kinds as impermissible or obligatory. From this it follows that Kantians aiming to defend these principles have to abandon the *Deontological Thesis*, that they have to be more open-minded – more “unorthodox” – about the structure and content of these principles.³²

2. The Constitutive Deficiency Response

2.1 Defective Circumstances and Deficient Options

The first “reconciliatory” response to the *Rigorism Objection* that I want to consider is Schapiro’s own. Her discussion is centred around the prohibition of deception. To put it crudely, Schapiro argues that in certain circumstances deception is permissible or even obligatory because being honest is impossible (2006: 45). Of course, it is always possible to stick to the letter of honesty: it is always possible to intentionally say what we believe to be true. However, when the circumstances in which we act are defective, then, Schapiro says, they can have a corrupting influence on our attempt to play by the rules: they can make it impossible for us to be honest in both letter and spirit. In such cases, the only form of honesty available to us is a deficient form of honesty, one that does not go beyond mere truth telling, and because this is so, because honesty proper is out of reach, it may be better – closer to the spirit of honesty, if not its letter – to intentionally say something that one believes to be false: to deceive in the spirit of honesty (Schapiro 2006: 47-8). When deception in letter, i.e. intentionally telling someone a falsehood, is the best surrogate for honesty, when it is the lesser of two evils, its intrinsic impermissibility is mitigated and we are by way of exception allowed to engage in it (Schapiro 2006: 49). Let’s consider the example that exercises Schapiro: Constant’s infamous example of the murderer at the door (see sect. 1.2 of the introduction).

Murderer at the Door: A murderer knocks at your door and asks about the whereabouts of his victim, who is hiding in your house. If you are honest, you put the victim’s life at risk, but if you opt for deceit, then you would seem to be doing something intrinsically impermissible.

³² This is what they have to do, if they don’t want to resort to the unpromising strategy mentioned in fn. 29 above.

In Schapiro's view, in such a situation, you cannot be honest in letter and spirit. You are thus allowed to tell the murderer what you believe to be false (deception in letter only) because, in this case, doing so comes closer to the spirit of honesty than telling him what you believe to be true (honesty in letter only).

Schapiro's response presupposes quite a rich conception of honesty and of its counterpart, deception, a conception of both as composites of letter and spirit. We have already mentioned the respective letter-components, but what does their spirit consist in? Schapiro agrees with many of her fellow Kantians that deception is essentially a form of interference with someone's ability to govern themselves (Korsgaard 1996a: 141, 347). Honesty is thus essentially "a way of acknowledging or respecting (by not interfering with) another's autonomy" (Schapiro 2006: 46). But, in Schapiro's view, this analysis is incomplete. The problem with deception, she notes, is not only that it interferes with people's autonomy, but also that it prevents people from taking up a shared practical standpoint, that it basically constitutes a refusal to reciprocate in a scheme of shared thought and action. Accordingly, being honest not only in letter but also in spirit means telling the truth *as a way of satisfying the other's* "(implied) demand to think and act from a shared standpoint" (Schapiro 2006: 50). Both analyses, the standard analysis and Schapiro's revised one, present honesty and deception as forms of interaction between members of a moral community, but they differ with respect to what they see as the governing norm of this community: on the standard analysis, it's a norm of non-interference; on Schapiro's analysis, it's a norm of shared agency or co-legislation.³³

With this in mind, we are in a position to specify what Schapiro means when she says that, in certain circumstances, being honest in letter and spirit is impossible. It is impossible in the sense that, in these circumstances, intentional truth telling is *not* a way of living up to the interlocutor's implicit demand to take up a shared practical standpoint. The two circumstantial conditions that she considers are, firstly, a lack of autonomy and, secondly, evil ends or intentions on the part of the interlocutor (2006: 50-1). If your interlocutor is unable or unwilling to take up a shared standpoint, then, whatever request they are making, it is not an implicit demand to take up a shared

³³ In sect. 4, I will suggest that, for Kant, these are two different perspectives on the moral community rather than two competing conceptions of it. The former is the perspective that Kant takes up in the Doctrine of Right (MM 6:229-371), the latter the perspective that he takes up in the Doctrine of Virtue (MM 6:375-493).

standpoint. So even if you say something that you believe to be true, saying it will not be a way of satisfying that demand. The murderer at the door, for example, chooses an end that is so blatantly immoral that he “makes it appropriate for you to regard him as having refused ... to play the co-legislation game,” and if he has refused to play that game, then, “whatever demand he is issuing, it is not the demand to share in thought and action” (Schapiro 2006: 52).

Schapiro is aware that talk of *mitigated obligations* towards wrongdoers threatens to lead us on a dangerous slippery slope, that this sort of talk can be thought to license problematic forms of vigilantism, for example. She therefore adds the following two provisos. First, when we are faced with less severe forms of wrongdoing, we are not usually inclined to interpret them as signs of a refusal to co-legislate: they are mere offences, not betrayals of the moral relationship (Schapiro 2006: 53-4). As such, they do not mitigate our obligations towards the wrongdoer at all. Second, even when our obligations are indeed mitigated, there are constraints on what we are allowed to do. For, since the moral relationship is not an optional relationship, since wrongdoers are still rational agents whom we owe respect, we are only allowed to deviate from the letter of a given duty in ways that help us come closer to its spirit, e.g. to deceive a wrongdoer in the spirit of honesty, but not to torture them gratuitously (Schapiro 2006: 54-5). Later on we will see that these two provisos do not suffice to restrict the field of deviations that are licensed by Schapiro’s account in a determinate manner. However, in order to get to the root of the problem, we need to first understand what Schapiro’s talk of defectiveness and deficiency amounts to.

Why are certain circumstances allegedly defective, leaving only deficient options for action, rather than simply being *different* from other circumstances and calling for different actions for *that* reason? In her 2006 paper on Kantian rigorism, in which the *Deontological Thesis* is first introduced, Schapiro presupposes this defectiveness claim without arguing for it (2006: 49), but we can find the missing argument in her 2003 paper on non-compliance as a non-ideal condition. In the latter paper, she discusses the constitutive dependency of actions that are part of a practice (so-called “practice actions”, such as promising, punishing or negotiating) on the compliance of others. Her discussion is indebted to John Rawls’ idea that practice rules define procedures which, if complied with, constitute participation in a new form of activity (Rawls 1955: 24). Taking up this idea, Schapiro notes that practices are subject to the threat

of subversion (2003: 344). If the degree of non-compliance with the rules of a practice crosses a certain threshold, then the practice can become a sham, meaning that attempts to perform practice actions on the part of compliant parties are not just *productively* unsuccessful in the sense that they fail to accomplish their respective ends, but also *constitutively* unsuccessful in the sense that it is not appropriate to characterize them in the terms of the practice (Schapiro 2003: 337). Let us consider an example.

Negotiation: It is constitutive of the practice of negotiating that both parties present and receive each other's claims and arguments in good faith, that they genuinely *try* to reach a mutually acceptable agreement. If the person you are negotiating with is merely pretending to do these things, while in fact stalling for time, then they are subverting the negotiation not only in the sense of making it impossible for you to come to an agreement, but also in the sense that they prevent you from engaging in an activity that deserves to be called "negotiating". Both of you are merely going through the motions, merely babbling on, as it were.

Now, of course, one *could* say that, in such a case, the fact of non-compliance, rather than introducing a defect, just creates a situation that is different in kind, a situation that calls for the "scolding game", say, rather than the "negotiation game",³⁴ but that would be an unusual way of describing what is going on. It is indeed more plausible to say that the circumstances are *defective* and that, as a result, the parties are engaged in a deficient negotiation. According to Schapiro, this talk of defectiveness and deficiency is appropriate whenever various forms of stage-setting, e.g. external signs or implicit communication, establish the relevant practice in such a way that the agents involved can be said to be occupying the relevant roles (2003: 339).³⁵ Then there is a game that they *ought to be* playing, whether they manage to or not.

These observations suggest that defective circumstances can be identified in the following way. First, we look for signs that indicate which "game" we are supposed to be playing on a given occasion. Once we know which rules we are supposed to play by, we can determine whether all players are a) capable of playing by these rules and b) sufficiently compliant for the game to be intact. If one of these conditions is

³⁴ I am assuming that scolding is the appropriate reaction to someone's refusal to negotiate properly.

³⁵ Schapiro adds that, "in these cases, the non-compliant party is tacitly or implicitly claiming the protections and prerogatives attached to his role while at the same time failing to live up to its demands" (2003: 339).

not fulfilled, then the circumstances are defective and the whole game is a sham, a series of deficient moves, a mere going through the motions. In such circumstances, the stringency with which the rules of the game apply to us, as compliant players, is mitigated, meaning that exceptions might be warranted. However, now that we have a set of criteria that allow us to distinguish between circumstances that are defective and circumstances that are merely different in the sense that different rules apply, we can see that Schapiro's story about the *Murderer at the Door* example is not entirely convincing.

2.2 Different Games, Different Rules

In order to determine whether the circumstances in the *Murderer at the Door* case are indeed defective, we need to settle on a "game". Which game is it that you and the murderer are supposed to be playing on this occasion? Is it the game of exchanging information? Or the moral game as such, the game of joint agency and co-legislation? The trouble is that neither answer is without problems.

On the one hand, if we say that it is the game of exchanging information, then, unless the murderer is trying to deceive you about his intentions,³⁶ there is no reason to assume that the relevant game or practice has been subverted. After all, his evil intentions notwithstanding, the murderer is a) an autonomous agent who is capable of distinguishing between truth and falsehood, and b) compliant, as far as the rules of the information exchange game are concerned. His evil intentions seem to have no bearing on the integrity of the information exchange game that he and you ought to be playing under the current supposition. This thought is confirmed by the fact that our intuitions about the *Murderer at the Door* case differ from our intuitions about the *Negotiation* case. Recall that, in the *Negotiation* example, we were inclined to agree with Schapiro's assessment: if one party is stalling for time, then it is counterintuitive to say that the other party is still negotiating. But arguably there is nothing wrong with saying that you, as the compliant party, are being honest with the murderer, despite his evil intentions. Compare this to the following scenario: someone in a Nazi uni-

³⁶ In Korsgaard's view, the murderer must disguise his evil intentions because otherwise he could not expect that you will cooperate (1996a: 136). Schapiro rejects this assumption, citing the following case as a counterexample: "Suppose that a Nazi came to the door under the mistaken impression that you are a good Aryan who believes strongly in the Nazi cause and that you are harbouring a Jew simply out of ignorance of his identity. This Nazi would have every reason to be completely up-front about his intentions on the assumption that you sympathize with them" (2006: 51-2).

form knocks on your door and asks whether you are hiding a Jew in your house. The stage for the information exchange game is set. Horrified you begin to stutter something, at which point your interlocutor suddenly bursts into laughter, revealing that it was all a bad joke. If, despite this revelation, you answered them truthfully, then it might be somewhat strange to say that you were being honest. It would be strange because, by bursting into laughter, your interlocutor indicates that they are not actually playing the information exchange game, and so any further attempt on your part to make a move in that game would be a mere going through the motions. This example differs from the *Murderer at the Door* example in that the rule being violated is a rule of the very game being played: the information exchange game.

On the other hand, it is equally problematic to say that you and the murderer are meant to be playing the moral game as such. *Prima facie*, this line might look more promising. Since the moral game is a game that does not need to be established via mechanisms of stage-setting, since it is a game that we ought to play all the time,³⁷ it seems appropriate to say that you and the murderer ought to be playing that game. And once we focus on the moral game, we can regard the murderer's evil intentions as a refusal to play by the rules. To use Schapiro's terms, we can see him as refusing to deliberate from a standpoint that all rational agents (including the innocent victim) can share. So why not focus on the moral game then? The problem is that we get the wrong result. The result that we were after is that the murderer's refusal mitigates the stringency of the principle of duty that prohibits deceiving others. But what we seem to have shown instead is that it mitigates the stringency of the supreme principle of morality that urges us to deliberate from a shared practical standpoint, that is, that it mitigates the stringency of the Categorical Imperative. Now, of course, as a Kantian one *could* argue (though I would say *against* Kant) that the murderer's unwillingness to play by the rules of the moral game makes it permissible to treat him as a mere means, at least if this is done in the spirit of morality.³⁸ But, once again, whatever the

³⁷ By this, I don't mean that every decision or practically relevant consideration is a moral decision or consideration – an idea famously criticized by Bernard Williams (1985: 179). In most situations, there is nothing moral at stake. What I mean, instead, is that we are always under the obligation to remain within the limits of morality, no matter what we are doing.

³⁸ A version of this view is advocated by Korsgaard in her essay "The Right to Lie". In this essay, she argues that, in circumstances of evil, we may deviate from the demand expressed by Kant's Formula of Humanity – from the demand to never treat humanity merely as a means, but to always treat it as an end (G 4:429). This is because, in such circumstances, trying to live up to the ideal that this formula articulates would make us into tools of evil (Korsgaard 1996a: 151-4).

general merits of such a Kantian twist on Rawls' double-level theory (Rawls 1999: 216), it does nothing to show that there are exceptions to the allegedly exceptionless prohibition of *deception*.

After all, it could be argued that it is already built into the content of our specific duties that people do bad things, that it is part of our shared understanding of what we owe to each other, part of our shared understanding of honesty, that there are situations in which honesty is not a reasonable demand. If this were true, then the murderer's evil plan would not result in a breakdown of the information exchange game. Instead, we would have to say that the game has a special rule for situations of this kind – a rule that is different from the rule that applies in standard cases, but a rule nonetheless. So from the point of view of the information exchange game, the circumstances would be different, but not defective.

Two points could be made on Schapiro's behalf. First, it is worth noting that the moral game is not only non-optional but also all-encompassing: all other games fall under its purview.³⁹ Thus, Schapiro could reply that, with the moral game having been subverted, the information exchange game cannot be perfectly intact either. This is a valid point, but it does not so much undermine the above criticism as prompt me to reframe it. Let us assume that, by subverting the moral game, the murderer has also subverted all other games that one might try to play with him. In that case, all bets are off. They are off not in the sense that we are allowed to treat him in any way we like (that is ruled out by the proviso that we must strive to realize the spirit of the relevant game), but rather in the sense that, for *any* activity involving him, we must ignore the rules that would usually regulate this activity and ask ourselves how to best approximate the spirit of morality. The problem with this very general precept is that it introduces a worrying degree of arbitrariness, that it puts agents on unregulated territory, leaving it up to every individual to decide what they think is right in a given case. In fact, the problem here is reminiscent of Kant's concerns about the state of nature. Unlike Hobbes, Kant does not believe that the state of nature is a state where anything goes, for, in his view, we are always bound by the demands of duty. But, in the state of nature, each of us is our own judge of whether

³⁹ See fn. 37 above.

a given act complies with these demands, and this why it is a very precarious state (Kant R 6:95).⁴⁰

There is a second point one could make on Schapiro's behalf. One could argue that my objection is question-begging in the sense that it presupposes an assumption that Schapiro rejects: the assumption that we have to choose between two alternative readings of the *Murderer at the Door* case, namely between the moral game reading and the information exchange game reading. Recall that, according to Schapiro's analysis of honesty as a composite of letter and spirit, being honest means telling the truth (that is, playing by the rules of the information exchange game) *as a way of* living up to the demand to think and act from a shared standpoint (that is, playing by the rules of the moral game). Once again, the point as such is well taken, but does not undermine my criticism. What I take issue with is, in a sense, precisely the expansion that comes with her analysis: the fact that, in order to account for the *Murderer at the Door* case, to explain how the murderer's evil plans make your honesty towards him constitutively deficient, she has to expand our ordinary understanding of honesty to a point where it becomes unrecognizable, to a point where it encompasses all things moral.⁴¹ Once we conceive of honesty in this very broad way, we lose our grip on the idea that the spirit of honesty will point you to the surrogate action that is best suited to address the *specific* defect that your situation exhibits. The precept to approximate the spirit of honesty becomes so general that we struggle to tie it back to the consideration that is presumably uppermost in your mind, namely the fact that someone's life is in danger.

The lesson to be learned, I think, is that the explanation of why we are allowed to deceive the murderer is not as generic as Schapiro makes it out to be. She is right: the murderer cannot expect an honest response. But the reason *why* this expectation is inappropriate is not simply that he refused to take up a shared practical standpoint, but rather, more specifically, that the duty to be honest, as we all understand it, ends where the duty to protect innocent victims from aggressors begins. This, of course, is just a different way of saying that principles of duty restrict one another internally or,

⁴⁰ In addition, we might worry that the precept to aim for or approximate the realization of the spirit of morality suggests that we have to engage in consequentialist reasoning, which is exactly the mode of reasoning that Schapiro is so determined to avoid (2003: 331-2, 344, 350 and 2006: 38, 45).

⁴¹ Let me clarify this point. The reason why her analysis of honesty amounts to such an expansion is that she is operating within the constitutivist framework outlined above. Within this framework, rules are constitutive of types of activities. If the activity of being honest is governed by the principle that one ought to say what one takes to be true as a way of living up to the demand to think and act from a shared standpoint, then *all* principles of duty are implicated in the constitution of this activity.

in other words, that the actions they prohibit are not intrinsically impermissible, i.e. that the *Deontological Thesis* is false. I have not done enough to establish this conclusion, but before we move on to the next attempt to resist it, I want to note that my explanation of why there are exceptions has another important advantage. It covers a wide range of cases, including cases that Schapiro's account does not cover because they do not seem to involve any constitutive failures.⁴²

3. The Conflicting Grounds of Obligation Response

3.1 A Kantian Theory of Pro Tanto Duties

Let us turn to another move that is open to those who strive to reconcile exceptions with the *Deontological Thesis*. Some commentators have argued that the below passage from the *Metaphysics of Morals* contains the key to a Kantian solution to the problems of moral conflict, moral residue, and moral regret.⁴³ This is because, in their view, it shows that Kant accepted the now popular idea that moral considerations can retain their moral import even when defeated (McCarty 1991, Timmermann 2013).

A conflict of duties ... would be a relation between them in which one of them would cancel the other ... But since duty and obligation are concepts that express the objective practical necessity of certain actions and two rules opposed to each other cannot be necessary at the same time, if it is a duty to act in accordance with one rule, to act in accordance with the opposite rule is not a duty but even contrary to duty; so a collision of duties and obligations is inconceivable. However, a subject may have, in a rule he prescribes to himself, two grounds of obligation ... one or the other of which is not sufficient to put him under obligation ..., so that one of them is not a duty. – When two such grounds conflict with each other, practical philosophy says, not that the stronger obligations takes precedence ... but that the stronger ground of obligation prevails. (Kant MM 6:224)

⁴² Schapiro acknowledges this limitation and mentions some such cases in a footnote (2006: 56-7).

⁴³ These problems have been raised by Williams and Martha Nussbaum, among others. In their view, if a moral theory leaves no room for genuine conflict between competing moral concerns, then it cannot explain why we feel regret even when we have done the right thing and were unable to do better. Williams argues that this is a problem for cognitivist moral theories in general (1973: 175-6), and Nussbaum maintains that it is a problem for Kant in particular (1986: 32).

According to the above commentators, Kant's claim that grounds of obligation can conflict suggests that defeated grounds of obligation retain their deontic import. As such, this claim is thought to point to a Kantian theory of *pro tanto* duties.⁴⁴

If they are right, then this passage may contain the solution to our problem, too. Think of the deontic import of the fact that an action is a lie. If the duty not to lie was a *pro tanto* duty, the deontic import of being a lie would be defeasible,⁴⁵ meaning that some lies could be overall permissible or obligatory. The principle of duty that prohibits lying would admit of exceptions. But, according to proponents of *pro tanto* duties, even permissible or obligatory lies are impermissible in at least one respect – namely *qua* lies. If this is true, if lies *qua* lies are always impermissible, and necessarily so, then it seems appropriate to say that lies, as such, are intrinsically impermissible: impermissible in virtue of the features that make them lies. This would mean that the *Deontological Thesis* could be upheld, even if the kinds of duties that it covers admitted of exceptions. Let us call this the “Conflicting Grounds of Obligation Response” to the *Rigorism Objection*. But why should we think that this is indeed what Kant has in mind when he speaks of conflicting grounds of obligation? Let us examine the details of this interpretation and consider its merits.

The *pro-tanto-duties* interpretation of the above passage hinges on a distinction between duties and obligations. According to Richard McCarty and Jens Timmermann, the term “duty” refers to a general principle of duty, i.e. to a duty *type*, whereas “obligation” refers to an instance of such a principle, to a particular case in which one is bound to perform a certain action, i.e. a duty *token* (McCarty 1991: 68, Timmermann 2013: 42-3). Duties, *qua* principles, determine which features constitute grounds of obligation or, in other words, which features have a *pro tanto* deontic import. When a ground is not defeated by a stronger competitor, then its import is decisive: it determines the overall deontic status of the relevant action and, as such, it

⁴⁴ The forefather of *pro tanto* duties is William David Ross (2002: 16-47). He speaks of “*prima facie*” duties, but notes that the phrase is somewhat unfortunate because it suggests that the duties in question are merely apparent, which is not what he has in mind (2002: 20). For this reason, his commentators agree that it is more appropriate to speak of “*pro tanto*” duties instead, as this phrase conveys that an action is impermissible or obligatory *in some respect*. Generally speaking, principles of *pro tanto* duty pick out features of actions or situations that make an action in some respect impermissible or obligatory.

⁴⁵ Recall that we are using the notion of defeat to cover both undermining and outweighing. In this context, “defeasible” means “outweighable”.

determines what we have an obligation to do.⁴⁶ Let us consider an example that we already mentioned in the beginning of the introduction.

Breaking a Promise to Help: You promised your friend to post their visa application, just in time for the deadline. But, on the way to the post office, you see a child fall off their bike, hit their head, and pass out. Two principles of duty apply here: the duty to keep one's promises and the duty to help people in emergencies. So there are two grounds of obligation or features with deontic import, favouring different actions: the fact that you promised makes it *pro tanto* impermissible for you to stop, the fact that the child needs help makes it *pro tanto* obligatory⁴⁷ to do so. If we assume that the child's need is the stronger ground, then it is overall obligatory for you to stop. This is what you have an obligation to do.

The advantage of distinguishing between duties and obligations is that it makes plain why Kant says what he says: that duties and obligations cannot conflict, while grounds of obligation can. Duties are principles considered in abstraction from the contingencies that attend their application.⁴⁸ The only way in which such principles could conflict would be by contradicting one another directly. A duty not to kill, for example, would contradict a duty to kill violent aggressors. Of course, such conflicts must be impossible if Kant's moral theory is to be consistent (Timmermann 2013: 42).⁴⁹ Grounds of obligation are features on which our duties bestow deontic import. Since actions can fall under more than one duty, there can be several grounds of obligation at once. As a result, the same action can be obligatory and impermissible, but it is so *in different respects*, and so the conflict here is not of a problematic kind: it is not a contradiction. Finally, there are obligations. Obligations cannot conflict because to say of one and the same action that it is overall permissible and overall obligatory is,

⁴⁶ Kant speaks of strength, not weight, and so do McCarty and Timmermann. But given their reading of the above passage, they *could* just as well speak of weight.

⁴⁷ This is perhaps a somewhat unusual way of speaking, as it would be more common to say that you have a *pro tanto* duty to stop. But what I want to convey is the idea that, on the present view, there are *pro tanto* deontic statuses and overall deontic statuses. Thus, to say that your action is *pro tanto* obligatory is to say that it is obligatory in one respect, not that one has an obligation to perform it.

⁴⁸ Since principles of duty *per se* do not say that this or that particular action has this or that deontic status, the fact that one and the same action can fall under several duties does not mean that *the duties themselves* conflict with one another.

⁴⁹ If they weren't, there could be *antinomies* of moral laws, which Kant rules out, e.g. in a draft of the *Metaphysics of Morals*, 23:389.

again, a contradiction. So, in each case, one ground of obligation must “prevail”.⁵⁰ To sum up, the pro-tanto-duties interpretation seems to make good sense of the key claims in the Kant passage cited above, and it has the additional benefit of countenancing moral residues, i.e. genuine deontic import that cannot be honoured.⁵¹

Now it is time to return to our own concern: the *Rigorism Objection*. If we attempt to appropriate the pro-tanto-duties reading for our purposes, we run into a problem. The problem has to do with the nature of the cases that we are trying to account for. Consider the *Breaking a Promise to Help* case, for example. What we want to say about this case is that a ground of obligation that obtains in virtue of what Kant would see as a *duty of right* (the duty not to break promises) is defeated by a ground that obtains in virtue of what he would call a *duty of virtue* (the duty to help).⁵² The worry is that this sort of case is ruled out from the start because, contrary to what we have been saying above, Kant’s theory of pro tanto duties is actually limited to duties of virtue, with duties of right being decisive whenever they apply. This is Timmermann’s view, among others (Timmermann 2013: 43-6, Aune 1979: 191-5).

Such a restriction suggests itself when we reflect on how pro tanto duties might fit into Kant’s overall project. According to Timmermann, the right place to look is Kant’s claim that duties of virtue possess *latitude*; that, because such duties prescribe ends, they leave it open how and when to take steps towards those ends (Kant MM 6:390). In Timmermann’s view, it makes sense to treat such duties as pro tanto duties because their latitude is a reflection of the fact that other duties can take precedence,

⁵⁰ It should make us suspicious that the present interpretation assumes that the two impossibilities, the impossibility of conflicts between duties and the impossibility of conflicts between obligations, have distinct sources. After all, Kant himself offers a single explanation and he seems to think that it covers both impossibilities. In the following section this suspicion will be confirmed. There I will argue for a reading according to which there is no such distinction between duties and obligations.

⁵¹ For McCarty’s and Timmermann’s purposes, it is important that the deontic import is *genuine*, and not merely *apparent*. Their opponents, the critics mentioned in fn. 43 above, believe that, for Kant, duties are binding whenever they apply, and thus that, in his view, all conflicts are merely apparent in the sense that they are based on mistakes of subsumption – on the mistaken view that a given case falls under two principles of duty. On this basis, these critics conclude that Kant cannot make sense of moral regret: the regret that we feel when we were unable to fulfil a genuine (albeit pro tanto) duty. McCarty and Timmermann aim to defend Kant against these critics.

⁵² Kant draws the distinction between duties of right and duties of virtue in MM 6:239, and explains it further in other passages, for example, 6:382-4, 394-5, 410. There are interesting questions about how this distinction relates to various other distinctions between kinds of duties that Kant introduces: wide vs. narrow, perfect vs. imperfect, mediate vs. immediate. Exploring these would take us beyond the scope of this thesis. So, apart from a brief remark in sect. 4.1 below, I will leave these questions aside. For detailed discussions of these and related questions see e.g. Trampota et al. 2013.

making the means required to pursue the prescribed ends “morally unavailable”.⁵³ In such cases, there is a *pro tanto* duty to pursue the end in question, but no overall duty or obligation to do so. This leads Timmermann to believe that duties of right, such as the duty to honour promises and contracts, are *not* *pro tanto* duties, because, as Kant says (MM 6:380), these duties do not prescribe ends but, instead, prohibit actions.⁵⁴

This reasoning rests on a mistake, however. There is indeed a sense in which duties of virtue prescribe the pursuit of obligatory ends, whereas duties of right restrict the range of permissible means, and in sect. 4.1 below we will see what exactly that sense is. But whatever the difference amounts to, when Kant says that duties of right do not prescribe ends, we should not take him to be saying that they cannot demand that we bring about certain states of affairs. For surely the duty to keep promises or the duty to fulfil contractual obligations can make such demands.⁵⁵ Conversely, when Kant says that duties of virtue do not prohibit actions, we should not take him to be saying that there are no specific actions that are blatantly incompatible with a genuine commitment to an obligatory end (McCarty 1991: 77-8, Walla 2015: 733-6). This suggests that the roles can be reversed: that there can be situations where the means that would be required to keep one’s promise, e.g. to keep walking to make it to the post office, are morally unavailable because taking these means would amount to a denial of assistance so harsh that it would betray a fundamental indifference to the needs of other people. For this reason, we should not rule out the possibility that a ground of obligation that obtains in virtue of a duty of right could be defeated by a

⁵³ Timmermann 2013: 48, 60. Note that this is not uncontroversial. It could be argued that the point of countenancing latitude is not to highlight the fact that duties restrict one another, but instead to avoid overdemandingness (Walla 2015: 737-41). However, Kant himself is quite explicit about endorsing the former view (MM 6:390).

⁵⁴ Timmermann 2013: 43-6. It is worth noting that, given Kant’s taxonomy, this would mean that the duty not to lie is a *pro tanto* duty, because it is classified as a duty of virtue (MM 6:429-31). If we accept Timmermann’s account of the distinction, this is an odd outcome. After all, it seems that the duty not to lie is as much a prohibition of actions as the duty not to break promises. That Timmermann ignores this problem has to do with the fact that he treats the distinction between duties of right and duties of virtue, on the one hand, and the distinction between perfect and imperfect duties, on the other, as co-extensive (2013: 44-6). In this, he deviates from Kant.

⁵⁵ By this, I do not mean that it is part of the *content* of the above duties that one ought to bring about certain specific states of affairs. All I mean is that, in order to fulfil such a duty on a given occasion, one might have to bring about a certain state of affairs. McCarty acknowledges this possibility when he says that some duties of right, e.g. the duty to keep one’s promises, “preclude the evasion of [a moral] commitment” (1991: 70). But his phrasing here betrays that he is reluctant to say that duties of right can prescribe the pursuit of ends.

ground that obtains in virtue of a duty of virtue. With this in mind, we can now move on to a critical discussion of the Grounds of Obligation Response.

3.2 Determinate Content and Indeterminate Import, or Vice Versa?

So far we have focussed on two issues: on whether the pro-tanto-duties reading can account for Kant's comments on conflicts and whether it works as a response to the *Rigorism Objection*. Now I want to raise a larger question: the question of whether this is indeed a good way of understanding Kant's ethics. The proponents of the pro-tanto-duties interpretation are generally in a difficult position. The textual basis for their reading is meagre, and even Kant's remarks on latitude do not actually support it – they merely leave room for it. After all, latitude could also mean that duties restrict one another internally, in the way suggested at the end of sect. 2.2.⁵⁶ So in order to adduce Kant's remarks on latitude as evidence for *their* view, pro-tanto-duties Kantians have to interpret them in a certain way, namely as expressing the idea that the means to an obligatory end can be morally unavailable. Let us scrutinize Timmermann's argument to this effect. In the following passage, he explains the notion of moral unavailability by using the example of someone with limited financial means who is forced to choose between supporting their poor parents and showing their gratitude to a benefactor. He argues that, given the strength of the respective grounds, the means for showing gratitude are morally unavailable, meaning that there is a pro tanto duty to show gratitude, but no overall duty or obligation.

Grounds of obligation depend on the precarious availability of means to generate [actual obligations] ... The reason for this follows directly from Kant's distinction between moral and instrumental reasoning. It is ... correct to say that it is my duty to give my savings to my parents, but strictly speaking, I have an obligation to help and giving them my savings is the way this is achieved, a mere *means*. The relevant description of the ground created by pure practical reason falls short of recommending that I transfer my savings to my parents' bank account. Nor does moral reasoning tell me that my benefactor's need can be relieved by giving him the same sum of money ... *Pure* practical reason tells me that benefactor and parents alike deserve my support and that my parents are morally more important. *Empirical* practical reason tells me that the only suitable means is a certain sum of money, which is insufficient to help out both. If it turns out that the money is re-

⁵⁶ In fact, that is exactly what Kant seems to be saying in MM 6:390.

quired to satisfy the stronger ground, this is no different in principle from other cases in which I am not in a position to help because the means are out of reach as a result of any other kind of impossibility: psychological, logical, physical, metaphysical etc., or indeed by the unrelenting direct ought of perfect duty. (Timmermann 2013: 57)

What is noteworthy about this passage is how Timmermann uses the division of labour between *pure* practical reason, as the capacity for moral reasoning, and *empirical* practical reason, as the capacity for instrumental reasoning, to explain how there can be grounds of obligation that don't translate into obligations proper. Since Kant rules out conflicts of obligation, it is important that pure practical reason does not go so far as to prescribe two incompatible courses of action, for example, that it does not tell me to both transfer my savings to my parents' bank account and to give them to the benefactor. The task of pure practical reason, in such a case, is limited to setting obligatory ends and determining their importance in relation to one another on an abstract level. The search for suitable means, which relies on knowledge about the world in general and the present situation in particular, is left to empirical practical reason, and sometimes its verdict is that there are no suitable means available. But what kind of fact is the fact that certain means, though physically and psychologically available, are morally unavailable?

Timmermann's claim that the moral unavailability of means is "no different in principle" from their physical and psychological unavailability suggests that the fact that certain means are morally unavailable is much like a non-moral fact. Specifically, it suggests that, from the standpoint of morality, it is *as contingent as* a non-moral fact. This, I think, is not an unintended connotation of Timmermann's remarks. In fact, this idea plays a crucial role in his argument for the conclusion that defeated grounds of obligation retain their deontic import. His argument is this: given that "Kantian ethics ... is a conscious attempt to protect moral life from the influence of fortune, accidental circumstance cannot render valid [grounds of obligation] null and void" (Timmermann 2013: 60). If this reasoning turned out to be faulty, the pro-tanto-duties interpretation would lose much of its plausibility.

So let us ask again: what kind of fact is the fact that the financial means required to show one's gratitude to a benefactor are morally unavailable? Surely it is different from the fact that one is too short to reach the branch with the nice red apple. It is

different because it is partly a matter of what or who *ought to be prioritized*. From the standpoint of morality, this fact is not contingent, it is not something that just is the way it is, not something to shrug one's shoulders at. Instead, it is the kind of fact that we know through pure practical reason. Let's think about what pure practical reason tells us, according to Timmermann. It tells us that both our benefactor and our parents deserve our support and that our parents are morally more important. How do we have to understand these commands such that, combined with knowledge of non-moral facts, they yield determinate instructions regarding what to do? Surely, to say that our parents are more important is not to say that they are more important *per se*: as rational agents endowed with dignity, the two parties have equal standing. Nor should we assume that the end of supporting our parents is more important in the sense that we should *always* prioritize it. After all, we can imagine extreme cases, cases where a small gesture could prevent a benefactor from losing their faith in humanity, for example. Timmermann's emphasis on the need to exercise judgment and his remark that urgency is an important consideration suggest that he would not rule out that, in exceptional cases, the duty of gratitude might take precedence (2013: 53). But if this is right, it means that the commands of pure practical reason, when stated in this abstract, formulaic fashion, are *indeterminate* in content. They command that we perform a *yet-to-be-determined* range of actions, whereby determining what those actions are is *not* just a matter of adding knowledge of non-moral facts.

Let us reflect on the implications of this insight. Recall why Timmermann insists on a division of labour between moral and instrumental reasoning. It is because pure practical reason issues its commands prior to and independently of the contingencies encountered in complying with them that such contingencies can never render these commands "null and void". Now, if the business of applying duties to the messy reality of everyday life was a matter *merely* of reviewing contingent facts to find the means to fully determinate obligatory ends, then Timmermann would be right: a case could be made that a shortage of means cannot take away from the deontic import of the features that our duties pick out. What our discussion has shown, however, is that moral and instrumental reasoning are in fact much more closely intertwined than Timmermann would have us believe. It has shown that it is *in* confronting the messy reality, *in* taking account of the circumstances, that we determine what our previously indeterminate duties entail, what actions they prescribe, and how they restrict each

other. In a nutshell, Kantian duties are not duties to perform a determinate range of actions with a yet-to-be-determined import – they are not pro tanto duties. Instead, they are duties to perform a yet-to-be-determined range of actions with a determinate (namely: decisive) import.

To illustrate, let us briefly revisit our example. The pro-tanto-duties reading says that, in transferring our money to our parents' bank account, we are failing to act on a pro tanto duty of gratitude, which urges us to transfer a particular sum of money to a particular benefactor on a particular occasion. We have found that this is not how it is. *Pace* Timmermann et al., there is no ground of obligation, no feature with deontic import that we are failing to honour. To say, in the abstract, that we have a duty of gratitude is not to say anything about what we ought to do in this particular instance. In this instance, the duty does not apply. There is therefore *no respect* in which our transferring the money to our parents is impermissible.

The conclusion that there is no residual deontic import brings us back to our main concern: to the Conflicting Grounds of Obligations Response to the *Rigorism Objection*. This response hinges on the premise that there *is* residual deontic import: that there is always still a respect in which a lie or promise-breaking is impermissible, even when it is permissible or obligatory overall. We have shown that this premise cannot be upheld. The Conflicting Grounds of Obligations Response fails.

However, before we can move on, we should briefly revisit Kant's comments on conflicting grounds of obligation. How are these to be understood, if not along pro-tanto-duties lines? Here it is important to pay attention to *where* such conflicts are supposed to occur. In the passage cited above, Kant claims that they occur in "rule[s] [that a subject] prescribes to himself". As many commentators have noted (Herman 1993: 167, O'Neill 2002: 342), this suggests that he is speaking of maxims, that is, of subjective principles of action. So why does Kant switch from talking about duties and obligations (objective principles) to talking about maxims (subjective principles)? I would argue that this is precisely because the kinds of conflicts that he goes on to countenance are conflicts that occur *only in the subject*. They are conflicts between two

perspectives on a given situation between which a subject is torn because they cannot decide which duty the present case falls under.⁵⁷

4. The Casuistical Questions Response

4.1 “*Strictly Speaking*, This Is Not a Lie”

The third and final response that I want to consider in this chapter takes as a starting point a long-neglected part of Kant’s discussion of duties in the Doctrine of Virtue: his so-called casuistical questions (MM 6:423, 426, 428, 431, 433, 437, 454, and 458). The twenty or so questions come in eight bundles, each associated with a particular duty.⁵⁸ Although they vary in their purport, ranging from quests for action guidance via empirical questions to questions concerning the value of certain states of affairs, they are all, in one way or another, geared towards settling questions of subsumption. In particular, it is clear that these questions are not supposed to cast doubt on the scope or bindingness of the principles of duty at issue: on whether they are strictly universal, say, or on whether they present the performance or omission of certain kinds of actions as necessary.⁵⁹ Rather, what they call into question is whether this or that particular act falls under the concept of the relevant action kind or end kind. Here are some examples:

Is it murdering oneself to hurl oneself to certain death (like Curtius) in order to save one’s country? (MM 6:423)

Can an untruth from mere politeness (e.g. the “your obedient servant” at the end of a letter) be considered a lie? (MM 6:431)

Does not all this [preferential tributes of respect in words and manners] prove that there is a widespread propensity to servility in human beings? (MM 6:437)

⁵⁷ This, of course, is the conclusion that McCarty and Timmermann are desperate to avoid because it would mean that conflicts between grounds of obligation are *merely apparent* after all. See fn. 51 above.

⁵⁸ The number of questions depends on how the blocks of questions in the Doctrine of Virtue are divided up. According to David James, there are twenty (1992: 68).

⁵⁹ This is certainly disputable. Many of the questions concern borderline cases of action kinds or states where the classification is a matter of degree, e.g. excess in the enjoyment of food and drink. These are clearly questions of subsumption. Some of the other questions are harder to interpret in this way, however. Consider the following example: Bitten by a mad dog, is it wrong for someone to destroy themselves so that, in their incurable hydrophobic madness, they don’t harm others? (MM 6:423-4). On the face of it, this is a quest for action guidance, not a question of subsumption. But one could argue that the incurable madness will extinguish this man’s humanity anyway or will have extinguished it already, so that one of the features that characterize suicide, according to Kant, is not present here.

In recent years, these questions and Kant's general comments on casuistry have attracted more and more attention, especially among commentators who are keen to defend Kant's ethics against the *Rigorism Objection* (e.g. Sherman 1997: 311-6, O'Neill 2002: 343-44). In an attempt to explain this recent surge of interest, Rudolf Schüssler observes that the only route open to Kantians who would like to grant "exceptions" to perfect duties is to resist subsuming the target cases under them (2012: 77). This remark is as illuminating as it is paradoxical. It is paradoxical because, *prima facie*, it does not make sense to regard a case that does not fall under a rule as an exception to that very rule.⁶⁰ Nevertheless, it is clear enough what Schüssler has in mind. When orthodox Kantians are presented with a lie or theft or promise-breaking that seems permissible or even obligatory, they can sidestep a counterintuitive verdict by saying that the act in question is not *actually* a lie or a theft or a promise-breaking. And, to justify such an ostensible deviation from the letter of Kant's rigoristic texts, they can point to his casuistical questions. Let us consider what they can say in more detail.

In raising and discussing these casuistical questions, Kant seems to acknowledge that there is a gap between principles of duty and particular acts and that bridging this gap is often far from straightforward, requiring judgment and attention to the details of the case. And if this is right, if he really does acknowledge that there is such a gap, then he can make room, if not for genuine exceptions to principles of duty, then at least for the appearance thereof. For the more intricate the subsumption of acts under action kind concepts is thought to be, the more reasonable it is to expect that ordinary language won't always follow suit and that, speaking loosely, people will often apply concepts of impermissible or obligatory action kinds, e.g. the concept of a lie, to acts that do not actually fall under these concepts. This loose speech will then create the illusion of exceptions, e.g. of permissible lies. This line of thought indicates how the appeal to Kant's casuistical questions can help Kantians answer the *Rigorism Objection*. By suggesting that the correct use of action kind concepts is a very intricate matter, the casuistical questions point to a distinction between ideal and non-ideal uses of such concepts and, in doing so, they pave the way to a story about exceptions that preserves the *Deontological Thesis*. *Strictly speaking*, so the story goes, certain action kinds are intrinsically impermissible and therefore impermissible in all situations, but, *loosely speaking*, there are instances of these kinds that are permissible or obligatory.

⁶⁰ This is why it is appropriate to put the term "exception" in scare quotes.

There are two obstacles to getting this response off the ground. First, there is a concern about the scope of casuistry in Kant's ethics. For although Kant does raise some casuistical questions about perfect duties to the self, his official position seems to be that the domain of casuistry is actually limited to imperfect ethical duties.⁶¹ This suggests that, in our target cases, i.e. cases of lying, promise-breaking, and so forth, there is no gap between principles and acts after all. The following passage from the *Metaphysics of Morals* hints at such a restriction in scope.

But ethics, because of the latitude it allows in its *imperfect* duties, unavoidably leads to questions that call upon judgment to decide how a maxim is to be applied in particular cases, and indeed in such a way that judgment provides another (subordinate) maxim (and one can always ask for yet another principle for applying this maxim to cases that may arise). So ethics falls into a casuistry, which has no place in the doctrine of right. (MM 6:411; my emphasis)

We need to carefully attend to the subject matter of this restriction, however. As Schüssler has argued (2012: 73-85), it is important to distinguish between *casuistry* in the narrow sense, which is a matter of identifying appropriate subordinate maxims to maxims of obligatory ends, and *casuistical questions*, which are exercises in casuistry only in the broad sense in which all forms of case-based reasoning in ethics and law are. In the above passage, Kant is speaking about casuistry in the narrow sense.⁶²

In order to understand this distinction, we need to take a little detour and revisit a topic touched on in sect. 3.1: the distinction between the ethical and the judicial domains, as reflected in Kant's division of the *Metaphysics of Morals* into a Doctrine of Right and a Doctrine of Virtue.⁶³ We were surprised to find that both duties of right and duties of virtue can demand that we bring about a certain state of affairs and that both can prohibit specific actions. This left us wondering why Kant maintained that prescribing ends was the distinguishing mark of duties of virtue and prohibiting ac-

⁶¹ This apparent tension has puzzled Kant's commentators (e.g. James 1992: 70, Unna 2003: 455-8), but, as we will see below, there is a resolution.

⁶² While I agree with this distinction, I do not agree with Schüssler's claim that the latitude involved in identifying subordinate maxims means that it is not a matter of getting it right or wrong (2012: 72). As I argued in sect. 3.1 and fn. 53, the latitude of duties of virtue should be understood as a reflection of the way in which duties restrict one another.

⁶³ The following account is inspired by Kant's remarks in MM 6:382-3. To simplify matters, I here ignore the distinction between ethical duties and duties of virtue. The former category is broader than the latter in that it includes the indirect duty to seek the incentives for the fulfilment of one's duties of right in the moral law (MM 6:220, 383). I think that everything I say is compatible with this additional point, however.

tions the mark of duties of right. I think that we can make sense of this claim if we regard the distinction between the ethical and the judicial domain as a distinction between *perspectives on moral life*, broadly construed. When we look at moral life from the judicial perspective, we treat people's ends as contingent and regard morality as limiting the means that they can take to advance their interests. This is the sense in which duties of right are concerned with actions only: they abstract from whether people's ultimate ends are good or not. When we look at the same moral life from the ethical perspective, we think of morality as prescribing necessary ends (the end of perfecting ourselves and the end of promoting the happiness of others), to which we have to seek suitable means. This is the sense in which duties of virtue are concerned with ends only: they do not assess actions one by one, but consider them only as parts of a larger life plan. Some actions are such that they are blatantly incompatible with a good life plan, e.g. torture and murder, and perhaps extreme cases of denying assistance. The duties that tell us to abstain from these actions are on the "perfect" side of the spectrum. But for most actions there is room for doubt and disagreement about whether they can be part of a good life plan. It is not obvious, e.g., how to best go about helping people in need. The duties that regulate these kinds of actions are on the "imperfect" side of the spectrum.

With this in mind, we can see why casuistry in the narrow sense is limited to imperfect ethical duties: it is only in the ethical domain that we "move down" from maxims of obligatory ends to subordinate maxims, and it is only in the case of imperfect duties that difficult questions "that call upon judgment" can arise. Casuistry in the broad sense, by contrast, matters in both spheres and for all kinds of duties, for we can always ask casuistical questions about whether a given concept, e.g. an action kind concept or a virtue concept, applies to a given case.⁶⁴

The second obstacle concerns the standpoint from which a claim to the effect that, "*strictly speaking*, this act is not a ϕ -ing" could be made. In her paper "Instituting

⁶⁴ It is an interesting question why Kant does not raise casuistical questions in the Doctrine of Right. Of course, in practice, the task of answering such questions falls to judges. Perhaps discussing precedents would have taken Kant too far afield. It is also worth noting that many of the questions of subsumption that arise in the ethical sphere *are* casuistical questions in the narrow sense, i.e. questions of whether a given subordinate maxim specifies suitable means to the realization of some obligatory end. So, as one would expect, there is an overlap between casuistry in the narrow sense and casuistry in the broad sense. For a similar point, see O'Neill 2002: 332.

Principles”, Onora O’Neill insists that the appeal to the faculty of judgment⁶⁵ is “wholly unhelpful in showing how the gap between principle and act is to be bridged” (2002: 332). In her view, the broadly Neo-Aristotelian accounts that focus on this faculty are accounts not of practical but of theoretical judgment, of a kind of judgment that is third-personal rather than first-personal, and that concerns a particular that is already present (e.g. a situation), not a particular that is to be produced (an act). The task of practical judgment with its world-to-mind direction of fit, she argues, is to “satisfy or contribute to satisfying a maxim [or principle of duty]” (2002: 335), not to subsume given act tokens under such principles. This is why she would argue that we cannot defend Kant’s ethics against the *Rigorism Objection* by claiming that some of the act tokens that we tend to subsume under concepts of intrinsically impermissible action kinds do not actually fall under these concepts. She would say that, in the context of such a defence, it is irrelevant whether we *speak* loosely. What matters is, so to say, whether we “*act* loosely”, and the worry is that, once we shift to this wording, it is all the more obvious that we have fallen into casuistry in the pejorative sense.⁶⁶

It seems to me that O’Neill is wrong when she suggests that the problems that we encounter in moral deliberation from the first-person standpoint are completely different from the ones we face when judging situations, characters or other people’s actions (2002: 81-2). The question of whether a certain principle of duty applies in a certain case or not arises from both standpoints, and in both cases there is a danger of answering it in a biased way.⁶⁷ This means, on the one hand, that the Casuistical Questions Response should not be rejected out of hand: that the way we *speak*, both as ordinary people and as philosophers, is not irrelevant in the context of the present dialectic. But, on the other hand, it also means that we cannot *simply claim* that a real or imaginary act, such as the false statement that is at stake in the *Murderer at the Door* case, does or does not fall under a certain principle. Instead, we have to corroborate

⁶⁵ In Kant’s terminology, the faculty of judgment is the faculty of subsumption (CPR A132/B171). This is not to be confused with the activity of judging, which he attributes to the understanding.

⁶⁶ O’Neill raises similar concerns about Anscombe’s idea that maxims contain descriptions (O’Neill 2004: 312-4). I discuss these concerns in ch. 4., sect. 3, where I expand on the response given here.

⁶⁷ One of the problems that O’Neill takes to be distinctive of practical judgment, for example, is that, in judging practically, we have to balance the demands of many different principles (2002: 85). But surely we can say the same of third-personal judgments about situations, characters and other people’s actions: that it is difficult to determine how to describe a case or person in morally appropriate terms and which principles to invoke in doing so.

such a claim in a way that dispels the worry that we are *abusing* casuistry. Let us see if this can be done.

4.2 The Use and Abuse of Casuistry

The pejorative notion of casuistry is a good starting point for our critical discussion.⁶⁸ It is generally difficult to shake off the impression that in casuistical arguments terms are being stretched and bent to people's wills.⁶⁹ The suspicion is well familiar. It is felt whenever people seeking to justify exceptions start quibbling over language, and, on reflection, there is no reason why it should not extend to moral philosophers – especially those who are anxious to reject a rigorism charge that critics have brought against their favourite thinker. This, then, is the objection we need to consider: that the Casuistical Questions Response is an *abuse* of casuistry. To clarify: the objection here is not to the claim that, in day-to-day conversation, we tend to use our terms loosely. *That* claim is not likely to cause much controversy. What is problematic is the use of this claim in attempts to justify exceptions, given the obvious temptation of abuse. Before we can assess the reasoning that leads the proponents of the present response to the conclusion that, strictly speaking, this or that act is not a ϕ -ing, we need criteria. How does one fend off the charge of abuse? How does one back up the claim that, strictly speaking, a given act is not a lie without begging the question? Let us begin from the standpoint of an ordinary agent, rather than the standpoint of a moral philosopher, and consider two strategies that are most certainly unsuccessful.

First, one cannot simply shrug one's shoulders. This might seem obvious, but it is not, for one might think that this is exactly what the term “judgment call” conveys: that there is nothing more to say, nothing to convince an interlocutor who just does not “see it”. Think of disagreements about the applicability of colour concepts and the utter pointlessness of pushing someone to back up their claim that the sweater is blue, not green. This is not how it is with moral disagreements, however.⁷⁰ Here we

⁶⁸ *Merriam Webster* defines casuistry, among other things, as a “specious argument” or “rationalization”, and one of the definitions given by the *Oxford Dictionaries* is “the use of clever but unsound reasoning, especially in relation to moral questions; sophistry”.

⁶⁹ This is an objection that both early modern thinkers *and* conservative catholic theologians mounted against the casuistical practice of Jesuits, when casuistry fell into general disrepute in the 17th century. For a discussion of how this debate influenced Kant, see Schüssler 2012: 90-4.

⁷⁰ It seems to me that this difference is not always sufficiently appreciated. In a much-discussed paper, entitled “Values and Secondary Qualities”, John McDowell draws an analogy between values and colours (1998: 131-50). The point of his analogy is to argue for a no-priority view: for the view that

do expect people to say more. Of course, it is possible that the subsumption of acts under principles of duty is a judgment call in a broader sense: in the sense that it is impossible to specify a second-order rule for how to go about the task of subsuming cases under a given first-order rule. This seems very plausible.⁷¹ The truth, I think, must lie somewhere in between the two extremes. If you were challenged to back up your claim that your telling the murderer a falsehood was, strictly speaking, not a lie, you would have to be able to say *something* to corroborate this assertion, but you would not be required to give a foolproof algorithm or full list of necessary and sufficient conditions, which would settle the question beyond doubt.

Second, one cannot fend off the charge of abuse by pointing to the deontic status of the act in question. In the *Murderer at the Door* case, for instance, you cannot substantiate your claim that your statement was not a lie by saying that making it was permissible or obligatory. This would be question-begging, given your opponent's suspicion that you are quibbling over terms precisely because you are biased and thus inclined to treat the act in question as permissible or obligatory.⁷² The considerations that you cite must provide *independent* support for the intended conclusion.⁷³

values and colours are both essentially related to human sensibilities and objective, and that neither of these characteristics is to be given explanatory priority over the other. However, when particularists draw on his views and arguments as a source of support for theirs, they sometimes seem to take the analogy further, assuming that evaluative concepts share a further characteristic of colour concepts, which McDowell highlights elsewhere. In *Mind and World*, McDowell says that colour concepts are only "minimally integrated into possible views of the world" (1994: 30). Part of what he means by this, I think, is that the possession of a colour concept is a conceptual capacity with relatively little content, that, when we apply the concept "red", for example, there is not much we can say in support of this application. It seems to me that something like this must be presupposed in Dancy's claim that "there is nothing that one brings to the new [moral] situation other than a contentless ability to discern what matters where it matters" (Dancy 1993: 50). I doubt that McDowell would agree with this extension of his analogy, and I think that, on reflection, it is not very convincing. I would argue that our evaluative concepts are *unlike* colour concepts in that they are deeply integrated into our views of the world, and that we do expect each other to give reasons when such concepts are applied.

⁷¹ Indeed, according to Kant, we have to assume that this is impossible and unnecessary if we are to steer clear of an infinite regress. He says: "The power of judgment is the faculty of subsuming under rules, i.e. of determining whether something stands under a given rule ... or not ... Now if we wanted to show generally how one ought to subsume under these rules, i.e., distinguish whether something stands under them or not, this could not happen except once again through a rule. But just because this is a rule, it would demand another instruction for the power of judgment, and so it becomes clear that ... the power of judgment is a special talent that cannot be taught but only practiced ... in the absence of [this] natural gift no rule that one might prescribe to him [to a student] for this aim is safe from misuse" (CPR B 171-3). Of course, this is nothing but a gesture at a topic that would have to be discussed in much more detail.

⁷² The worry is, in other words, that one is trying to make an exception *for oneself*. I will say more about the distinction between exceptions for oneself and exceptions that are called for by the circumstances in ch. 3, sect. 2.5.

⁷³ This is what Schüssler has in mind when he cautions against an inference that some Kantian have been tempted by, namely the move from the claim that questions of subsumption are often open and

Having said what fending off the charge of abuse *cannot* look like, we have come closer to understanding what it must look like. To fend off such a charge, we have to cite features of the act other than its deontic status that warrant its exclusion from the sphere of the relevant action kind concept. What is interesting is that this kind of justification is strikingly similar to the kind of justification we would expect from someone trying to back up their claim that a given lie is permissible or obligatory, *except* that the latter justification would not take the detour via the question of subsumption. Let us consider an example.

Lie to a Friend: I overheard how you lied to a common friend and now I am scolding you. One way of justifying what you did would be to say that it was not strictly speaking a lie. If I responded with a rebuking look, you might try to allay my suspicion by saying that you had no intention to deceive. But note that there is another, more direct route that you could take: you could accept my claim that it was a lie and appeal to the very same consideration, that is, to the fact that you had no intention to deceive, to show that it was a permissible lie.⁷⁴

There is no doubt that real life disputes about alleged exceptions to perfect duties come in both these forms. Yet, according to proponents of the Casuistical Questions Response, justifications that proceed via the latter, more direct route are at best loose talk, and at worst misguided. If you and I agree that lying is impermissible, then, in their view, the *real* point of contention between us must be whether your act was a lie, no matter which direction our conversation takes on the surface level. (After all, if lying is indeed impermissible, then, following the *Deontological Thesis*, it must have that status across the board.) But why should we accept that, at a deeper level, *all* disputes about alleged exceptions to perfect duties are disputes over semantics?⁷⁵ Why should we accept that, once it is settled that the action falls under the concept of a lie or a theft or a promise-breaking, it is also settled what its deontic status is?

intricate to the claim that settling such questions is a matter of focusing *directly* on the deontic status of the relevant acts (2012: 94-5). Schüssler is right to note that this conclusion leaves us wondering how it is possible to have intersubjective insight into or share judgments about whether a particular case falls under a principle of duty or not. After all, if I think that some act is a lie and you think that it is not, then my pointing out that it is impermissible won't help us to make any progress.

⁷⁴ This example is interesting precisely because an intention to deceive is sometimes thought to be a necessary condition for lying. I will come back to this below, both in the main text and in fn. 77.

⁷⁵ In fact, in ch. 3, I will argue that *in a sense* they are, but by that I won't mean that they are all disputes about the *applicability* of an action kind concept.

It seems that we have run into a difficulty. On the one hand, we have managed to get an idea of what a successful attempt to substantiate claims of the form “*Strictly speaking*, this is not a ϕ -ing” might look like. But, on the other hand, we have realized that this is not necessarily how such substantiations would be presented. So now that we know what a good substantiation might look like, we no longer know how to tell whether we have found one. The following thought might help to get us out of this conundrum: if the difference between the two justificatory strategies open to you in the *Lie to a Friend* case was indeed *merely* a difference between strict and loose uses of language, as proponents of the Casuistical Questions Response maintain, then they should sound equally intuitive. And, arguably, in this case, they do. Given that you intend to justify yourself by appeal to the fact that you had no intention to deceive, it makes at least as much sense for you to reject my characterization of your act as a lie as to proceed on the assumption that it was one. But what about our target cases? What about the *Murderer at the Door* case, for example? Presumably, here, you would justify your making a false statement by appeal to the fact that the addressee of your statement was planning to use the information provided to murder someone. Now, in this case, it would seem much more appropriate to accept your opponent’s claim that you lied and argue that, in this situation, lying was permissible. For why would the fact that someone’s life is in danger bear on whether your statement is a lie?

Let us take a step back and reflect on what we have said. What the comparison between these two cases reveals, I believe, is that our intuitions about the felicity of justifying oneself directly or indirectly, via a refusal to subsume, turns on the extent to which the consideration appealed to is part of what we regard as the semantic core of the relevant term.⁷⁶ In this respect, the *Lie to a Friend* case is a bit of a special case: the reason why, in this case, the two justificatory routes seem equally intuitive is that it is genuinely controversial whether an intention to deceive is necessary for lying or not.⁷⁷ But many of our target cases, including the *Murderer at the Door* case, fall clearly on the direct-route side of the spectrum. These cases differ from the cases presented

⁷⁶ It could be argued that I am using a bad criterion because our felicity intuitions might be influenced by pragmatic considerations. Perhaps it is for merely pragmatic reasons that the shorter route sounds more intuitive. I think a comparative analysis would show that this is not what is going on. The more important point, in any case, is the idea of a semantic core and the question why the fact that a life is in danger would bear on whether your statement is a lie. The appeal to justificatory routes and felicity intuitions is meant to serve as an illustration, not as an argument that stands on its own.

⁷⁷ Deceptionists argue that an intention to deceive is necessary for lying (e.g. Mahon 2008b, Lackey 2013); non-deceptionists argue that it is not (e.g. Saul 2012).

in Kant's casuistical questions in that the considerations appealed to are clearly *not* part of the semantic core of the relevant action kind concepts. Thus, in these cases, claims to the effect that a given act is, *strictly speaking*, not a lie, or a promise-breaking, or what have you, would be unfounded: not proper *uses* but *abuses* of casuistry.

With this in mind, let us return to the standpoint of the orthodox Kantian moral philosopher. He or she wants to defend Kant's ethics by arguing that any apparent exception to principles of perfect duty, e.g. the principle that lying is impermissible, isn't *actually* an exception because the act in question isn't *actually* of the relevant kind, e.g. not a lie, despite the fact that, speaking loosely, we might call it that. The above argument has shown that that might work in some cases, but certainly not in all, and, crucially, not in the cases that Kantians and their critics tend to argue about. So even though there is a gap between principles or concepts, on the one hand, and particular cases, on the other, and even though Kant shows awareness of this gap, none of this helps to reconcile the idea that there are exceptions with the *Deontological Thesis*. The Casuistical Questions Response is ultimately a second-order abuse of casuistry.⁷⁸ But when we say this, it is important to keep in mind that *our* objection to this abuse of casuistry is *not* the objection of 17th century catholic theologians who were anxious to restore the strictness of divine commands. Most of *us* share the intuition that cases such as the *Murderer at the Door* case call for exceptions, that lying to the murderer can be justified, although only via the direct route: by arguing that it is a permissible lie because an innocent victim needed protection. So, from our point of view, the above discussion suggests that we need to reject the *Deontological Thesis* and to admit that it is part of our understanding of the duty not to lie that, in certain circumstances, it has to give way to the duty to protect.

5. Taking Stock: No Ideal World Without Exceptions

In this chapter, we looked at three orthodox responses to the *Rigorism Objection*: three attempts to show that a commitment to the *Deontological Thesis* would not preclude Kant from acknowledging exceptions. In each case, the general strategy was to argue

⁷⁸ This point is very similar to the point that I made when criticizing Schapiro's response in sect. 2.2. There I said that, in order to claim that the murderer's evil plan makes being honest with him impossible, Schapiro has to expand our ordinary understanding of honesty to the point where it includes all moral matters. Now my point is that, to deny that the lie to the murderer is a genuine lie, proponents of the Casuistical Questions Response must expand what we would usually see as the semantic core of the term "lie" or "lying" to the point where one of the necessary conditions of applying this term is that the interlocutor isn't planning to murder anyone.

that, whenever we think that we have found an exception to a principle of duty, what we have *actually* found is an instance of the principle that is in some way corrupted: a constitutively deficient case or a case where the principle cannot be followed because the means are unavailable or a case that common parlance usually misclassifies. The hope was that putting the blame on these non-ideal conditions would allow Kantians to leave the *principles themselves* untainted. Unfortunately, our discussion has brought out that this hope is vain: that the cases in question are not actually corrupted, and hence that even in an ideal world there would still be exceptions.

We arrived at this conclusion by showing that, in each case, in order to secure the claim that the relevant cases (such as the *Murderer at the Door* case) are corrupted, it was necessary to either stick to a very abstract understanding of the duties at issue and avoid specifying their content (sect. 3), or to expand our ordinary understanding of the relevant action kinds to the point of unrecognizability (sect. 2 and 4). These are actually but two sides of the same problem: once we specify what the relevant principles of duty demand *in concreto*, and do so without blurring the lines between different kinds of actions, we realize that these principles are internally intertwined. Grasping such a principle involves understanding in which situations it is valid, and when it has to give way to another principle and the moral concern that animates it. It is part of our understanding of the duty not to deceive, for example, that, in a case like the *Murderer at the Door*, this duty has to give way to the duty to protect people from violent aggressors.

Of course, if we accept this point, we thereby abandon the *Deontological Thesis*. According to Schapiro, this is problematic because, in allowing that circumstances can make a difference, we go against a fundamental Kantian doctrine: the idea that morally worthy conduct is a matter of *principles*, not *consequences*. In sect. 1, we mentioned why she has this concern: in her view, circumstances “can only affect us by altering the causal nexus within which we act” (Schapiro 2003: 330). But why should we accept this assumption? Why should we rule out the possibility that circumstances might affect us by making a difference to how we *conceive* our actions? Why should we deny that they can figure *in our principles*, that they can matter in the space of reasons and not just in the space of causes? This is the possibility to which we shall turn in the next chapter.

In acknowledging this possibility, we move beyond orthodox readings of Kant's ethics towards more unorthodox readings that treat the *Rigorism Objection* as a concern about Kant's principles of duty *themselves* – about their structure and content, rather than just their application to a world of contingencies. As such, the concern becomes an essentially *particularist* concern. It becomes a concern about how we can and why we should uphold the idea that principles play a fundamental role in moral thinking, given the complexities of the moral domain. As we will see, this concern comes with its own challenges, challenges that unorthodox Kantians have to rise to if they want to say that features such as being a lie or a promise-breaking have their deontic import only *by default*⁷⁹ and not across the board.

⁷⁹ In the introduction I used the less technical term “usually”. In what follows I will use the term “by default”. The idea that some features might have a certain moral status *by default* (e.g. be a reason by default, be wrong by default etc.) has been adduced to defend moral particularism against the charge that it “flattens the moral landscape” because it draws no distinction at all between the moral import of features such as killing and lying and the moral import of shoelace colour. In this context, default statuses have been introduced as a solution. For more detailed discussions of default reasons see e.g. Lance and Little 2006, Dancy 2004: 69, 111-7, Väyrynen 2004, McKeever and Ridge 2006: 46-75.

Chapter 2

Universal and Nonetheless Defeasible?

Two Responses to the Particularist Challenge

1. The Particularist Challenge: A Trilemma

Our discussion in the previous chapter has brought to light that the *Rigorism Objection* calls for radical measures: that, if Kantians want to provide an adequate response, they have to abandon the orthodox assumption that there is only a small number of features that can make a difference to the deontic status of actions (namely those that make them count as instances of action kinds that are by default obligatory or impermissible, e.g. as lies), and that these are the only features that figure in Kant's principles of duty. As a result, Kantians are under pressure to deviate from the letter of Kant's texts and to admit that the content and internal structure of such principles is, in fact, more complex. However, in going down this route, they get caught up in disputes that are in some ways very similar to those that exercise contemporary generalists in their attempt to face up to their particularist opponents. My hypothesis in this chapter is that we can understand why unorthodox Kantians deal with rigorism concerns the way they do and, more importantly, why they fail, if we appreciate that the various demands that are being negotiated in the generalism-particularism debate leave them stuck in a trilemma. Let us first recall what these demands are.

In the introduction, we defined particularism in the following way.

Moral Particularism: Even if there are some true moral principles, nothing in moral theory or moral practice hinges on them – we can explain the obtaining of moral statuses and act well without them.

As we noted, the main argument cited in support of this view is the argument from holism (see Dancy 1993: 60-2, 2004: 7, 78-85). Holism, recall, is a view about how certain things (in our case, actions) come to have certain moral statuses or properties (in our case, overall deontic statuses). More specifically, it is the view that, strictly speaking, *anything* can make a difference to the obtaining of a given moral status, and that this is why our explanations (in our case, of why a certain action is permissible, impermissible or obligatory) vary so much from case to case. The version of this view that, we said, matters for our purposes is *Holism about Deontic Statuses*.

Holism about Deontic Statutes: A feature or set of features which make it the case that an action has a certain (overall) deontic status in one situation may not play the same role in other situations.⁸⁰

Particularists maintain that holism substantiates their misgivings about principles, if not directly, then at least indirectly, because, in their view, it is difficult to see how the variability with circumstances that it posits could be captured in moral principles that are not entirely useless (Väyrynen 2011: 255). We will return to the specifics of this worry below. For now, we must register how this argument sets up the dialectic between generalists and particularists: if generalists accept that holism is true, they are faced with the task of accommodating the variability with circumstances that holists insist on.⁸¹ Taking our clue from the foregoing discussion, we can frame this as leading to two challenges. First, to account for the *complexity* of the moral domain, that is, for the fact that even features with a default import (e.g. being a lie) do not have this import invariably (i.e., that their import can be defeated). And, second, to resist the temptation to say that the defeaters were part of the explanation all along and hence to deny that the explanation actually varies, i.e. that there is any *asymmetry* in explanatory burden between standard and exceptional cases (sect. 2.3.2, 3.3.2). The challenge, in short, is to accommodate the following two desiderata.

Asymmetry: In standard circumstances, an action of a by default impermissible kind is impermissible solely in virtue of the features that make it count as an action of that kind (i.e. *not* in virtue of the fact that defeaters are absent).

Complexity: In exceptional circumstances, an action of a by default impermissible kind can be obligatory or permissible, and when it is, this is partly in virtue of features other than the ones that make it count as an action of that kind (i.e. in virtue of the fact that defeaters are present).

For Kantians, this is a challenge indeed, not least because they tend to defend a radical form of generalism:

⁸⁰ As before, “deontic status” means “overall deontic status”, unless otherwise indicated. I will omit the “overall” in what follows.

⁸¹ Of course, generalists are also free to reject holism. Some have argued, for example, that holists rely on misleading examples because the reasons in these examples aren’t fully specified. See e.g. Stratton-Lake 2000, Raz 2006, and Hooker 2008.

Universality: The features in virtue of which actions have their deontic statuses can be captured in universally valid principles of duty.

The commitment to the third desideratum, the *Universality* desideratum, is thought to follow from a centrepiece of Kant’s ethics: the universal law formulation of the Categorical Imperative (G 4:421). If it does, then it is surely a commitment that even the most unorthodox Kantians should be unwilling to flout – and *that* it does follow is by and large beyond dispute.⁸² However, *how* it follows is something that is rarely made explicit. Crudely speaking, the idea seems to be this: if the “universalisability test” has shown that an action with a certain set of features (the ones that the maxim specifies) cannot be willed or even thought to be performed across the board, then it follows that *any* action with the same features is equally problematic. In ch. 3, we will see that this line of reasoning is flawed, but for now its merits and demerits are beside the point. In this chapter, all that matters is the opinion of those whose views we are trying to make sense of: that *these thinkers* aim to accommodate *Universality*, and that they aim to do so for the reasons just cited, is enough.

Taken together, these three prima facie incompatible desiderata give rise to a trilemma. Let us begin with an overview, followed by an illustration.

The Trilemma

- | | | |
|-------------|---|--------------------------------------|
| First Horn | <i>Universality</i> and <i>Complexity</i> rule out <i>Asymmetry</i> | (Fine Print Reading) |
| Second Horn | <i>Universality</i> and <i>Asymmetry</i> rule out <i>Complexity</i> | (Rigourism) |
| Third Horn | <i>Complexity</i> and <i>Asymmetry</i> rule out <i>Universality</i> | (Particularism,
Pro Tanto Duties) |

Recall the *Breaking a Promise to Help* case. If Kantians want to combine the view that breaking a promise in this exceptional case is permissible in virtue of the fact that the action is not only a promise-breaking but also a helping (*Complexity*) with the view that the deontic import of promise-breaking can be captured in a universal principle of duty covering all cases (*Universality*), then they must say that, in standard cases, where breaking a promise is impermissible, this is partly because there is no child

⁸² Sometimes Kantians seem to be disputing it, but in many cases their attempts to reinterpret Kant’s Categorical Imperative in a way that involves no commitment to *Universality* don’t go very deep. As we will see below, this is true of Herman, who distances herself from the derivation-of-duties model of the role of the Categorical Imperative, but actually does not give up on *Universality* completely (see sect. 3 below).

that needs help (*vs. Asymmetry*). On this view, the principle of duty that captures the deontic import of promise-breaking has a fine print: breaking promises is impermissible *unless* there is a child that needs help. Of course, the list of defeaters is bound to be much longer, and so is the list of corresponding unless-clauses. Some Kantians bite the bullet and endorse this so-called Fine Print Reading (see e.g. Wood 2008: 68, Cholbi 2013: 447). But this sort of view has its problems.⁸³ According to Jonathan Dancy, there is no reason to think that such fully expanded principles will contain anything less than a description of the exact state of the world in which the case to be accounted for is embedded (a “supervenience base”). This, he thinks, is a *reductio*.

First, these principles specify complexes of such a size and in such detail that there is ... no chance that they should be capable of recurring. A principle that has only one instance is worse than useless, for no such principle could ever be a guide for judgement. Second, these things do not really have the form of a principle ... For they contain all sorts of irrelevancies. Principles are in the business of telling us which actions are wrong and why they are wrong. But the vast propositions generated by supervenience fail utterly to do this ... They do no doubt contain the relevant information, but they do not reveal it. (Dancy 2004: 87)

However, despite these issues, the second horn of the trilemma is often seen as worse. For if Kantians want to combine the view that breaking a promise in a standard case is impermissible solely in virtue of the fact that it is a promise-breaking (*Asymmetry*) with the view that the deontic import of promise-breaking can be captured in a principle of duty covering all cases (*Universality*), then they have to belie our moral intuitions and claim that in the allegedly exceptional *Breaking a Promise to Help* case it is in fact impermissible to break the promise after all (*vs. Complexity*). This, of course, is the rigorist position that we were hoping to get away from. Since we have shown that orthodox Kantians cannot account for exceptions *without* touching the content and structure of Kant’s principles of duty, the only choice they have left is to stick to their guns (Timmermann 2001: 65-7, Schüssler 2012: 3).

Finally, there is a third option. If Kantians want to combine the view that breaking a promise in the *Breaking a Promise to Help* case is permissible in virtue of the fact that the action is not only a promise-breaking but also a helping (*Complexity*) with the

⁸³ The disadvantages of expanding Kant’s principles of duty in this manner will be addressed in sect. 2.3.2 and 3.3.2 below. In ch. 4, we will discuss a related issue, namely the question of how much detail a principle of action (or maxim) should contain.

view that breaking a promise in a standard case is impermissible solely in virtue of the fact that it is a promise-breaking (*Asymmetry*), then they would have to admit that there is, in fact, no universal principle of duty that captures the deontic import of promise-breaking in all situations (*vs. Universality*). Instead, the import of promise-breaking would have to be judged on a case-by-case basis, as particularists insist it must. However, as we shall see in sect. 3.3.1, particularism is not the only view that the combination of *Complexity* and *Asymmetry* leaves open – an account of *pro tanto* duties would be a further option. That these two views are grouped together follows from how we set things up: from the very beginning we assumed that, for Kant and Kantians, there is only one level, the overall, and that the phenomena for which contemporary metaethicists introduce a further level, namely the contributory, are “absorbed” in and handled via the overall. This assumption is reflected in the three desiderata listed above. Thus, if Kantians were to argue (as Herman seems to do at times, see sect. 3.3.1 below) that the *contributory* import of promise-breaking is stable, so that it could be codified in universal principles of *pro tanto* duty, and that the explanatory asymmetry is due to the fact that, in the exceptional case, its import is outweighed by the import of the child’s needs, then, in our taxonomy, these Kantians would still end up on the third horn. This is because they would still deny that there are principles of *overall* duty that cover all cases, which is one of the claims that we are trying to accommodate.

This, then, is what I meant when I said that Kantians encounter the particularist challenge in the form of a trilemma. We can summarize their specific challenge as follows:

Particularist Challenge: Kantians have to explain how it is that our explanations of exceptional cases make reference to defeaters (*Complexity*) while our explanations of standard cases don’t (*Asymmetry*), even though both explanations are meant to appeal to one and the same explanatory principle, that is, to the same principle of duty (*Universality*).

In the previous chapter we discussed the responses of Kantians who are impaled on the second horn, and we will come back to those impaled on the first (see sect. 2.3.2 and 3.3.2) and the third (see sect. 3.3.1). For the most part, however, this chapter will look at two Kantian views that, I think, are best understood as attempts to

satisfy all three desiderata at once and to escape the trilemma altogether: Korsgaard's and Herman's views. Both appeal to a version of the idea that principles of duty can be both universal and defeasible. In relating their views to the trilemma, I do not mean to suggest that they actually *intend* to escape this trilemma; all I want to say is that, given their commitments, the trilemma is a device that can help us to shed light on the structure of these two views and their flaws. In each case, I will outline the view (sect. 2.1, 3.1), highlight the aspects that indicate a commitment to our desiderata, and then show that there is a tension between two elements of the view (sect. 2.2, 3.2). These tensions, I will conjecture, reflect the struggle to satisfy all three desiderata at once and the pressure to then introduce a "division of labour". Subsequently, I will consider whether either of the two views could be saved by giving up one of the two irreconcilable elements and conclude that they cannot (sect. 2.3, 3.3). I will finish off with a brief discussion of the conceptions of defeasibility that they appeal to and the reasons why these are inadequate (sect. 4).

2. Korsgaard on Provisional Universality

In a footnote in ch. 1, sect. 2, we briefly mentioned Korsgaard's early response to the *Rigorism Objection*: her claim that, in circumstances of evil, we may have to deviate from the demand expressed by Kant's Formula of Humanity, that we may have to lie and deceive, for instance, because, in such circumstances, living up to the ideal can mean becoming a tool of evil. Her more recent take on the problem of rigorism in Kantian ethics, as outlined in the fourth chapter of *Self-Constitution*, differs from her take in the early essay in that it is, quite explicitly, an attempt to meet the *Particularist Challenge*: an attempt to show that we must appeal to universal principles if we are to will at all, but also that we can appeal to such principles without being rigoristic because *provisionally universal* principles are enough.

2.1 Against Particularistic Willing: An Argument with Two Concessions

Korsgaard's response to the *Particularist Challenge* consists of two parts, namely of a direct attack on particularism and of two concessions, which are supposed to indicate how her Kantian brand of generalism can accommodate the particularist's concerns. Her project in the relevant chapter is to rule out competing accounts of the role and normativity of principles of practical reason on the basis that they cannot make any

sense of the unity of the will. One of the accounts rejected along the way is particularism, defined as the view that it is “possible to have a reason that applies only to the case before you, and has no implications for any other case” (Korsgaard 2009: 72-3). Korsgaard’s concern at this point is with acting for reasons or from principles of action in general, but we can relate her claims to the subject matter of our investigation if we bear in mind that acting from moral reasons or principles of duty is but one specific instance of the broader kind that she is interested in.

Korsgaard’s argument against particularism is the following (2009: 75-6).⁸⁴

1. Willing means determining yourself to be a cause, not being moved by the causal operations of incentives within you.
2. Determining yourself to be a cause means following an incentive on the condition that doing so is representative of your practical identity.
3. Willing particularistically, if possible, would mean following an incentive in its full particularity, i.e.
 - a. as not “representative of any sort of type” of policy or standing commitment,
 - b. or as “in no way [...] further describable” (2009: 76).
4. Following an incentive in its full particularity means *not* making your following it conditional on whether doing so is representative of your practical identity.
5. Thus, it is impossible to will particularistically.

An example might help to illustrate the key transitions.

White Lie: A thin-skinned friend asks you for your opinion on a sensitive topic. You know that the truth will hurt their feelings unnecessarily and the very thought of it makes you cringe. After some toing and froing, you decide to tell a white lie.

Since, in doing so, you exercise your will, your recoiling impulse does not simply make you spit out a lie (premise 1). If you do tell the lie, then this is because, having stepped back from the impulse and considered it in light of your practical identity,

⁸⁴ For now, I will use Korsgaard’s terminology, which reflects both her constitutivist reading of Kant’s ethics (her view that action is self-constitution through the reflective endorsement of incentives), and the specific aim that she pursues in the relevant chapter (demonstrating the importance of principles for securing the unity of the will). Later, when I articulate my criticism, I will rephrase some of her ideas in my own terms.

you have chosen to treat the prospect of hurting your friend, which made you recoil, as a defeater of the default deontic import of lying (premise 2). Of course, particularists do not want to say that your recoiling impulse simply makes you spit out a lie, and they might even agree that, in a sense, willing to lie or lying for a reason differs from compulsive lying in that it is conditional on your regarding the lie as representative of your practical identity (Dancy 1993: 55-6). However, according to Korsgaard, they cannot make good on this claim, because, in denying that a certain feature's (your friend's feelings) being a defeater here and now has any implications for how things stand in other cases (with further defeaters, contributory reasons, etc. present), they also deny that, in treating it as such, you have to appeal to any principle or policy or standing commitment to treat a certain feature as making a certain difference across the board (premise 3a).⁸⁵ As a result, particularists rob themselves of the resources needed to make sense of the very idea that your white lie represents your practical identity (premise 4). For what could such a practical identity possibly consist in, if not principles or policies or standing commitments that are robust across varying circumstances, and that make you the kind of person you are, e.g. someone who generally pays her debts, and keeps her promises, and lies to spare her friends' feelings? This is why, according to Korsgaard, if you did not "will[] [your maxim] as a universal law" (2009: 73), the distinction between you and the contingent play of incentives within you would collapse, leaving a "*mere heap* of unrelated impulses" (2009: 76).

In arguing against the possibility of particularistic willing, Korsgaard aims to defend Kant's Formula of Universal Law and the idea that the principles on which we act, including principles of duty, must be universally valid. So part of her aim at this stage is to accommodate *Universality*.

Of course, as it stands, this argument does nothing to address the problematic implications of Kantian ethics that particularists find fault with, and even though addressing these does not advance the dialectic of Korsgaard's fourth chapter, she seems to think that they are sufficiently pressing to justify a diversion. As I noted at the outset, the diversion comes in the form of two concessions. Korsgaard notes

⁸⁵ As the above overview indicates, Korsgaard offers two versions of premise 3. In ch. 3, sect. 1, I will argue that there is an important difference between them, which escapes Korsgaard's attention. But, for now, I shall focus on 3a, which suggests a reading of the overall argument that is quite common. See e.g. Stern 2015: 58.

that, in arguing against particularism, she is neither arguing that we cannot “will a new maxim for each new occasion”, nor that we cannot “will[] a maxim that [we] might have to change on another occasion” (2009: 75). These two points are connected. The first one concerns the circumstance-sensitivity of our principles of action.⁸⁶ It is an acknowledgment that our principles have to be sensitive to the specific details of situations, including those that are not taken to have a deontic import by default. The second concerns the universality of our principles. Although the above argument shows that our principles of duty cannot be particular, Korsgaard allows that their universality might be merely *provisional* or, in other words, that they might be defeasible.⁸⁷ The second concession is related to the first, for if it turned out that a given principle was not absolutely universal after all, then this would be because it was not fully sensitive to all the features whose presence would make a difference to what we ought to do.

Let me illustrate the distinction between provisional and absolute universality by revisiting the *White Lie* case. Imagine that until today you held a strict principle of truthfulness, but now, faced with the prospect of unnecessarily hurting your friend’s feelings, you decide that the case at hand is an exception. It is an exception in the sense that it reveals that your principle of truthfulness, insensitive as it is to a morally relevant feature of the situation, is not absolutely universal. Does this mean that until today your willing was particularistic or that your principle was a mere rule of thumb? Korsgaard insists that it does not, because whenever we “encounter an exceptional case, we ... go back and revise [our principle], bringing it a little closer to the absolute universality to which provisional universality essentially aspires” (Korsgaard 2009: 74). Since she assumes that we treat exceptions as signs that our principles stand in

⁸⁶ Korsgaard speaks of “generality” and “specificity”, saying that our principles might have to be quite specific. This is somewhat misleading, however. The specific principles she has in mind are principles with a range of unless-clauses. Such principles are indeed more specific than their counterparts without unless-clauses in the sense that they cover a smaller number of cases. However, in addition to the number of cases covered, she seems to think of the difference between relatively specific and relatively general principles in terms of being more or less tailored to a given situation, and that is misleading. Because, if circumstance-sensitivity is secured by adding a long list of all possible defeaters, then the resulting principle will be *anything but* tailored to a particular situation. This is an issue to which we will return in sect. 2.3.2. (She also uses “general” in the Kantian sense of “mere generality”, meaning non-strict, empirical universality, but this need not concern us any further).

⁸⁷ When we say that a principle is defeasible, there are two things that we could have in mind. First, we could be saying that our credence in it is not absolute. This is something that a proponent of the reflective equilibrium method would say about any principle (Rawls 1999: 42-5). Second, we could be saying that there are, in fact, defeaters. In sect. 2.2, 2.3, and 4, we will see that Korsgaard’s argument runs into problems precisely because these two notions aren’t clearly distinguished.

need of revision, that they need to be supplemented with unless-clauses that render them more sensitive to the morally significant differences between situations, she can insist that we do not treat exceptional cases as if they had *no implications* for other cases.⁸⁸ And this, she believes, shows that provisional universality is different in kind from particularity.

Together these two concessions hold out the prospect of accomplishing the twofold task that generalists inherit from their particularist opponents. The first concession can be read as an attempt to satisfy *Complexity*. In admitting that “every occasion may be different in relevant ways from the ones we have previously encountered” (2009: 73), and that, therefore, our principles of duty have to be sensitive to the specifics of different situations, Korsgaard acknowledges that situational factors can affect whether a certain kind of action is permissible or impermissible or obligatory in a given case. The second concession, in turn, can be construed as an attempt to satisfy *Asymmetry*. For the most plausible explanation of why Korsgaard says that provisional universality is enough, I think, is that she is trying to block a conclusion that her reader would otherwise feel entitled to draw from her first concession. This is the conclusion that our principles of duty, being circumstance-sensitive, come with a host of unless-clauses (with a fine print), which clutter our deliberation about and explanations of standard cases with a lot of unnecessary detail. The second concession goes some way towards blocking this conclusion because, if we incorporate defeaters as we go along, then the content of our principles is at least to some extent controlled by the specifics of the exceptional situations that we actually encounter.⁸⁹

These two concessions provide a clue to Korsgaard’s brand of Kantian generalism. It is an account of principled agency that trades on the contrast between two standpoints: the standpoint of finite agents “on their way” who, over the course of their lives, learn more and more about how they ought to conduct themselves in different sorts of circumstances, and the standpoint of an omniscient agent, who has arrived at the ideal endpoint of this epistemic journey.⁹⁰ Even at the end of our lives

⁸⁸ For a discussion of the common saying “The exception proves the rule” see Holton 2010.

⁸⁹ It does not go all the way because, on this account, our principles grow more and more complex over time. So when we return to standard cases after having experienced exceptional cases, the defeaters built in on these occasions will show up, although there is no need to rule them out here and now.

⁹⁰ Note that this is not Kant’s notion of moral progress. For Kant, moral progress consists in progressively aligning our willing with the demands of the Categorical Imperative, not gaining a more extensive understanding of what the Categorical Imperative demands. In fact, one may doubt that his view leaves room for the latter kind of progress.

we will only have encountered and considered a tiny proportion of all possible situations, and so, depending on how complex the moral domain really is, we can assume, more or less safely, that we will never actually reach the endpoint: that, taken together,⁹¹ our principles of duty will always remain provisionally universal, lacking the circumstance-sensitivity required to handle all possible scenarios appropriately. Korsgaard points out that we are well aware of these limitations of ours. We acknowledge them whenever we qualify our principles, saying that they hold “everything else equal”, or that it is impermissible to do x “unless there is some good reason why not”, etc., without being able to list all the conditions that have to be in place (2009: 74).⁹² (This remark suggests that she thinks of the list as *open-ended*. We will come back to this point in sect. 2.3.2 below). Because of its appeal to two standpoints, I will refer to this generalist account as the “*Two Standpoint View*”.

In the following section I will argue that the *Two Standpoint View* is untenable, but even if it wasn’t, it is in any case a significant deviation from Kant’s texts. So why does Korsgaard opt for this sort of view? If we bear in mind that her argument and her concessions can be construed as attempts to satisfy the three desiderata that give rise to the above *Trilemma*, then the distinction between two standpoints is important because it allows her to *divide labour*. The standpoint of finite agents allows her to satisfy *Universality* (via “provisional universality”) and *Asymmetry*, while the standpoint of the infinite agent allows her to satisfy *Universality* and *Complexity*. By keeping these standpoints apart, she can hope to avoid clashes and thus to escape the *Trilemma*.

2.2 A *Reductio Ad Absurdum* of the *Two Standpoint View*

The clashes cannot be avoided, however. Korsgaard’s *Two Standpoint View* implies, rather absurdly, that finite agents usually get it wrong: that, most of the time, when such agents exercise their will, they fall short of one of the norms constitutive of this

⁹¹ Maybe for *some* principles, we could manage to reach absolute universality. This depends, in part, on the thickness of the concept(s) by which the feature(s) with deontic import are picked out. Think of the principle “Murder is impermissible”. If it is part of our understanding of the term “murder” that the killings to which this term applies are impermissible, then this principle might have very few, if any, exceptions. Whether and to which extent the evaluative content conveyed by thick terms is variable, and how this bears on whether this content is part of their meaning is a complicated question, which I cannot address here. For a discussion see Väyrynen 2013: 215-232.

⁹² Actually, Korsgaard goes back and forth between saying a) that we are unaware of certain background conditions, and b) that, although we are aware of them, we do not explicitly think about or mention them in each and every case. This is an important ambiguity, which I will come back to in sect. 2.3.2.

faculty. This problem is easily overlooked, however, and that is due to the dialectic of the passage in which the view is first introduced. As we saw, the *Two Standpoint View* emerges in the context of an argument against particularism. The argument aims to show that a certain theory of the role of principles in practical reasoning, namely particularism, does not achieve what it purports to achieve. The theory purports to provide an account of what it means to act for reasons but, by denying principles their place in such an account, it misrepresents the nature of its explanandum. If we focus on the norms of *philosophical enquiry* appealed to in this argument, on the question whether a given theory gets at the nature of its explanandum, then Korsgaard's *Two Standpoint View* seems to have the upper hand: it seems to satisfy the demands to which such a theory is subject.

However, if we abstract from the dialectical context in which the *Two Standpoint View* is introduced and focus, instead, on the norms that guide the activity that the relevant philosophical enquiry takes as its object, i.e. on willing, then we find that this dialectical advantage comes at a cost. For, among other norms, this activity is supposed to be subject to the norm expressed by Kant's Formula of Universal Law, which, according to Korsgaard, demands that we act from universally valid principles – from principles that hold in *all* cases or situations (see premises 3a to 5). However, if this is what it takes, then the *Two Standpoint View* implies that, at least most of the time, our day-to-day *moral enquiry* fails to achieve what it purports to achieve because, as finite agents, we tend to act on principles that fall short of the absolute universality to which we “essentially aspire”. More specifically, even when we do perform right actions and abstain from wrong ones (when we keep our promises, tell the truth, and so on), we do so for the wrong reasons and so our actions have no moral worth (Kant G 4:399-400).⁹³ This suggests that Korsgaard avoids a philosophical error at the cost of accusing finite agents of constant moral error.

Think of the *White Lie* case, for example. If this case is an exception to the strict principle of truthfulness that you held until today, then all the true statements that you made in the past may have been right, but they cannot have had any moral worth. They must have been unworthy because they were based on a principle that

⁹³ Kant's idea that our actions have no moral worth unless we act from duty will be discussed in detail in ch. 3, sect. 3.1. We will see that he has a very demanding view of what it means to act for the right reasons.

was only provisionally universal, on a principle that, upon reflection, you *cannot* will to become a universal law.

Let us pause and ask why the *Two Standpoint View* has this absurd implication. Prima facie, it is not in any way absurd to distinguish between two epistemic standpoints. Of course, we might say, our quests for knowledge, for example, for scientific knowledge, are never completed, we never reach the state of omniscience. But surely that does not mean that most of our beliefs are false. They are *fallible*, but that is very different from saying that most are *de facto false*. This is true, but it does not undermine what we said above. Instead, it serves to highlight a distinctive feature of the *Two Standpoint View* and its conception of the validity of moral judgments (which many Kantians share). According to this view, the mark of the ideal endpoint of our epistemic journey, namely having absolutely universal principles, is at the same time the norm on the basis of which every single act of cognition, i.e. every derivation of a moral judgment from a principle “on the way”, is assessed. It is treated as both an ideal endpoint that we strive to reach eventually *and* a norm that we aim to comply with in every single exercise of the will. As a result, falling short of omniscience amounts to making a false (not merely fallible) judgment.⁹⁴ Korsgaard’s *Two Standpoint View* has this absurd implication because it saddles absolute universality with two incompatible roles. However, *that* it has this implication is not immediately obvious, and that is partly because the distinction between two epistemic standpoint seems unproblematic and is unproblematic in many other contexts.

2.3 Two Responses on Korsgaard’s Behalf

The above *reductio argument* has brought to light that absolute universality can only play *one* of the two roles that Korsgaard assigns to it, not both. Accordingly, there are two responses that might be given on Korsgaard’s behalf. One response (to be considered in sect. 2.3.1) would be to insist that absolute universality was never meant to play the role of a norm that regulates every exercise of the will; the other (to be considered in sect. 2.3.2) would be to insist that it is not a distant ideal but rather an aim that is well within reach. Let us see if one of these responses can be made to work.

⁹⁴ This is not, strictly speaking, correct. A judgment about the deontic status of a particular action that was derived from a non-universal principle could be correct, but it would be based on a false premise. As such, it would be based on bad reasons and would lead to an action with no moral worth.

2.3.1 The Voluntarist Response

It might be thought that my objection rests on a misunderstanding of Korsgaard's constructivist reading of the Formula of Universal Law. Perhaps I was wrong to assume that her reading of this formula amounts to a commitment to *Universality*. In order to pursue this suggestion, we need to extend the scope of our discussion beyond principles of duty and consider maxims in general.

Recall Korsgaard's objection to the particularist. According to her, particularists fail to make sense of the idea that willing is a matter of following incentives on the condition that doing so is representative of our practical identity or, in other words, on the condition that it accords with the principles that *we hold* as universal. Now, if part of her aim in presenting this argument is to defend Kant's Formula of Universal Law, then it seems that her reading of this formula is rather weak. On this reading, it expresses the demand that we ought to treat a given feature as having a certain normative import only if, at that time, we are committed to treating this kind of feature as having this kind of import across the board. Let us assume that this is how she sees things. In that case, Korsgaard's being content with provisional universality would simply reflect her view, defended on other occasions (e.g. 1996b: 120), that most parts of our practical identity (most of our specific practical identities) are contingent, optional and alterable, parts that we may adopt, develop, expand, and revise – in particular, when faced with new and unfamiliar circumstances, such as the ones encountered in the *White Lie* case. Let us call this reading of Korsgaard's project the Voluntarist Reading.⁹⁵ On the Voluntarist Reading, Korsgaard's view does not imply that the judgments of finite agents are mostly wrong. For although new and unfamiliar circumstances can prompt us to decide that a principle of ours does not hold across the board after all, and thereby prompt us to revise it (to add an unless-clause, for example), this does not mean that we were not committed to its absolute universality before. We have changed our practical identity, and so it is only appropriate, the voluntarist will say, that the range of actions that count as representative of that identity should change as well.

I doubt that this is how Korsgaard would like to be understood because, in the end, it would render her view incoherent. The Voluntarist Reading derives its initial

⁹⁵ The first in a long line of commentators who have read Korsgaard's argument in this way is G.A. Cohen, in Korsgaard 1996b: 167-88.

plausibility from how well it seems to work in the realm of the merely permissible.⁹⁶ Consider, to begin with, how it fails to account for the *White Lie* case, where something moral is at stake. We said that the prospect of hurting your friend's feelings makes you doubt your strict principle of truthfulness and eventually prompts you to revise it. Let us assume that you made the right choice: that telling this thin-skinned friend a white lie was the morally appropriate thing to do. According to the voluntarist, Korsgaard would argue that, in this case, seeing your friend vulnerable effects a change of heart: it makes you reconsider who you are, makes you modify parts of your practical identity, makes you want to become a more considerate person.⁹⁷ Now, of course, this kind of thing can happen. But it better not be an example of the supposed right to revise our principles as we please that was meant to help us resist the above *reductio argument*. After all, Korsgaard insists that our moral identity is a necessary, non-optional, unalterable part of our practical identity (1996b: 120-5). In light of this, we would expect her to oppose the idea that someone might align their principles with their moral identity without acknowledging that what they did before was wrong. Maybe there is a chance that your policy of truth-telling never actually hurt anyone, but if morality really demands that you revise your principle in the way you do, then your original principle *must have* fallen short of the norm expressed by the Formula of Universal Law.⁹⁸ This is to say, we cannot re-interpret the *Universality* desideratum as saying that the deontic import of features can be captured in principles of duty that *we treat as* universally valid, for then moral requirements would not be objective – and that is a conclusion that Kantians are very anxious to avoid.⁹⁹

Prima facie, it might seem that, in the realm of the merely permissible, things are different. Consider the following example of a non-moral principle of action.

Taste for Theatre: You've never had much of a taste for the arts, but one day your friends convince you to see a play. The play affects you so deeply that it brings about a change of heart. You start questioning your attitudes and modify parts of your practical identity. Among other changes, you revise your principle to al-

⁹⁶ We will come back to this realm and the demands within it in ch. 3, sect. 3.2.

⁹⁷ Let us use the expression "change of heart" to distinguish the voluntarist account of principle revision from the account discussed in sect. 2.1 and 2.2, on which principle revision amounts to a "change of mind".

⁹⁸ For now, I am ignoring Korsgaard's distinction between the Categorical Imperative and the moral law (1996b: 98-100). I will come back to this distinction and its problems in ch. 5, sect. 2.1 and 3.1.

⁹⁹ We will come back to the question of what it would mean to "subjectivize" Kant's ethics in ch. 4, sect. 2.3.

ways reject invitations to the theatre, deciding that, from now on, you intend to reject such invitations only if they clash with your professional obligations.

Here it may seem that the original rejection policy not only succeeded in representing the person you were in the past, but also that it was perfectly legitimate. Thus, in the realm of the merely permissible, the Voluntarist Reading of Korsgaard's account seems to accord with our intuitions. However, for all their contingency, optionality and alterability, our non-moral practical identities are not adopted at random. We adopt them for reasons, and these, Korsgaard argues, trace back to our moral identity (1996b: 121).¹⁰⁰ If we realize that there are good reasons for adopting a certain specific practical identity, or for thinking about it in a certain way, we should admit that we would be failing to appreciate these reasons, if we did not do so. And if we had these reasons all along, then there is a sense in which we *did* go wrong. Arguably, your philistine self, for example, missed out on many opportunities to enjoy themselves and to develop their sensibilities.

These considerations, and especially the objectivity worry, suggest that it is more charitable to treat Korsgaard's *Two Standpoint View* as a more or less self-contained element of her overall position, and to abstain from interpreting her contentment with provisional universality as an implication of her alleged voluntarism. But, in that case, the above *reductio argument* retains its force: the *Two Standpoint View* implies that most of our moral judgments are flawed and most of our actions devoid of moral worth.

2.3.2 The Implicit Awareness Response

The Voluntarist Response was an attempt to resolve the tension within Korsgaard's account by denying that absolute universality was ever meant to play the role of a norm that applies to every single exercise of the will. According to the voluntarist, *that* role is played by a weaker norm, a norm that amounts to little more than the demand to be consistent. The response that we will consider now tackles the tension from the opposite direction. Perhaps we misunderstood Korsgaard's talk of encountering exceptions and making revisions, perhaps this talk is not intended to characterize an ongoing epistemic journey towards absolute universality after all.

¹⁰⁰ This point is also emphasized by Thomas Scanlon in his discussion of constructivism (2012: 238).

In sect. 2.1 we assumed that, when Korsgaard speaks of an agent discovering an exception to their principle, she is thinking of someone who has discovered that they made a mistake, and that when she speaks of revisions and addendums, she is thinking of someone who has changed their mind (as they should, given the mistake). I have to admit, however, that the textual evidence here is ambiguous. On some occasions, Korsgaard's phrasing suggests a rather different picture, a picture on which the "discovery" of an exception is really more of a recollection or a bringing to attention of something that one was implicitly aware of all along. At one point, for instance, Korsgaard discusses the example of "someone who decides to become a doctor in the full light of reflection" (2009: 74). What she seems to have in mind is that, through reflection, this person could concentrate their attention on something that they knew all along, namely that a principle according to which one is allowed to enter a profession when one desires to do so is subject to a range of background conditions, e.g. the condition that there is a social need. To encounter an exception, on this account, is not to discover a counterexample to a principle that is insufficiently circumstance-sensitive, but to be prompted to foreground (in deliberation or in a conversation) part of the hidden structure of an adequately sensitive principle, to be prompted to attend to one of its unless-clauses because, for once, the defeater that it specifies is present. Accordingly, to "revise" a principle is not to change one's mind, but to think about or cite something that was in the back of one's mind all along, and to say that a principle holds "everything else equal" is not to acknowledge one's epistemic limitations or, worse, one's fallibility, but rather to simply indicate that one has not spelled out which conditions the principle is subject to. If this is indeed what Korsgaard's talk of exceptions and revisions amounts to, then finite agents are not necessarily "on their way". They may already know everything they need to know, and thus there is no reason to think that they are constantly making mistakes.

In ch. 3, I will argue that we can and should regard our principles of duty as defeasible in the sense intended here, so, in my criticism of the Implicit Awareness Response, I want to focus on why Korsgaard can't avail herself of the resources needed to maintain this view. In particular, I will argue that she cannot avail herself of the distinction between foreground and background. But first we need to think about the point of this distinction: how would it help if it *could be* upheld? The answer to this question becomes apparent once we rephrase it: why is Korsgaard unwilling to say

that we have to act from absolutely universal principles without any hidden structure? Why does she introduce the notion of provisional universality in the first place?

In sect. 2.1, we explained this unwillingness by appeal to *Asymmetry*. Opting for such a view, we said, would allow her to satisfy *Universality* and *Complexity*, but leave her unable to explain why it is that, in standard cases, our explanations and justifications seem to make no mention of defeaters at all. She would be unable to explain, for instance, why it is that, in a standard situation, I can justify my decision to tell the truth *without* mentioning that the addressee of my utterance has a thick skin or that they are not a murderer, or some such. That I do not need to mention these things is important for several reasons. Some have already come up when we cited Dancy in sect. 1 above. First, depending on how much variability there is, the list of defeaters that one would have to rule out might be extremely long or even open-ended. In this case, complete explanations and justifications would be a matter of impossibility. Second, even if the list was “merely” long, the resulting explanations and justifications would not be very good ones. After all, it is their job to *reveal* why something is or is taken to be the case. As such, they should not be cluttered with irrelevancies: with considerations that could make a difference on some other occasion, but are not actually pertinent here and now.¹⁰¹

The problem of irrelevancies is particularly pressing when it comes to justifying our actions to others. Normally, we would think that standard cases of truth telling should not leave much room for disagreement: I ask why you said x, you respond that x is true, and that is it. But now imagine your response had to feature a long (if not open-ended) list of defeaters. The chances of disagreement would increase enormously. Is the fact that some chatty acquaintance might tell your thin-skinned friend what you really think a defeater of the defeating import of the prospect of hurting your friend with respect to whether you should lie? Who knows. You might think it is, I might think it is not. If we had to agree on these sorts of details, then successful justificatory exchanges would be hard to come by. (Note that this interpersonal dimension of the problem of irrelevancies is more of a threat to Kantian Constructivists than to their realist opponents, because constructivists maintain that a moral judgment is correct precisely in virtue of being justifiable in light of an inter-

¹⁰¹ If principles contained long lists of unless-clauses, this would affect their ability to serve as guides as well (Dancy 2004: 87), but, as I said in the introduction, I am focusing on their role as explanatory standards, at least for now.

subjectively intelligible procedure of construction or scrutiny (Korsgaard 2008: 321-4).¹⁰² That might be why this dimension does not come up in *Dancy's* critique; after all, he is a moral realist.)

Here the foreground-background distinction might seem to come in useful. For if there is such a distinction, then Korsgaard can argue that there are indeed absolutely universal principles of duty that cover all cases, standard as well as exceptional, but that, in explaining and justifying ourselves, we only have to explicitly appeal to the pertinent elements of these principles, with the other elements remaining implicit. On the face of it, this is a viable response to the *Trilemma*, as summed up in the *Particularist Challenge* (sect. 1), and it seems all the more convincing when the argument for it draws on an analogy with the inexact sciences (Korsgaard 2009: 74). It is a commonplace that the inexact sciences, e.g. clinical medicine or economics, trade in *ceteris paribus* laws, whose validity is subject to a wide range of conditions (Brandom 2000: 88).¹⁰³ In these sciences, to say that a given generalization is a genuine law is *not* to say that one could fully specify the whole range of conditions to which it is subject. In fact, the list of conditions can be treated as open-ended. Rather, to say that a generalization is a law, and hence a source of genuine explanations, is (at least, i.a.) to say that it is counterfactually robust or, in other words, that its truth is not merely accidental. To put it crudely: there must be a causal connection. Similarly, Korsgaard could argue that what we take to be a principle of duty can be a genuine “practical law” (Kant CPrR 5:19), a law that we can appeal to when justifying our actions or explaining the obtaining of a deontic status, even if we are not in a position to fully specify the whole range of conditions on which its validity depends. (We will return to this comparison of principles of duty to laws of nature in ch. 3).

Unfortunately, the analogy does not hold up. It does not hold up because, according to Korsgaard’s reading of the Categorical Imperative, a principle of duty is valid *precisely* in virtue of being robust across not only an unspecifiable range of cases but absolutely all cases (anticipating a distinction to be drawn below, we could say: in virtue of being Case-Scope universal). That is what “absolute universality” means in the context of her account, and that is the role that it plays if we treat it as a norm of

¹⁰² In ch. 5, I will argue that this is how Kantian Constructivism should be understood.

¹⁰³ Marc Lange argues that all sciences appeal to laws that are only robust across a certain range of possible worlds (2000). For a discussion of how his view bears on the generalism-particularism debate see Lance and Little 2007.

good willing rather than an ideal endpoint of an epistemic journey – as we chose to in setting up the present defence. The norm of correctness in the inexact sciences is a different one. There a generalization counts as lawful (and the explanations it yields as correct) if it captures a necessary connection rather than an accidental one or, again, if it captures a causal nexus rather than a mere correlation. The explanatory import of an alleged cause is not constituted by and doesn't depend on its being embedded in an absolutely universal law, but, for Korsgaard and many other Kantians, the deontic import of a feature is and does. This is why ruling out all potential defeaters, and hence explicating what is supposed to remain implicit, is precisely what it would take for you to justify telling the truth or to explain to me why telling the truth was obligatory. And if this is right, if Korsgaard cannot avail herself of the distinction between foreground and background,¹⁰⁴ then the Implicit Awareness Response fails. It fails in the sense that it leads to the Fine Print Reading, according to which principles of duty are infeasible and cumbersome.

3. Herman on Deliberative Presumptions

Over the course of her engagement with Kant's ethics, Herman has made a number of proposals for how to defend Kant against the charge of rigorism. In this section, I want to focus on a very influential proposal developed in her 1993 paper "Moral Deliberation and the Derivation of Duties". Of course, this paper does not address itself to the *Particularist Challenge* (after all, it was published in the very same year as Dancy's *Moral Reasons*), but, like Korsgaard's view, her view can also be *understood as* an attempt to meet this challenge, and to meet it by appeal to defeasible principles. Herman's central idea in this piece is that Kant's Categorical Imperative yields so-called deliberative presumptions that can be defeated by competing moral concerns.

3.1 The Categorical Imperative as a Source of Deliberative Presumptions

Herman's declared aim in the above paper is to develop a new account of the role that Kant's Categorical Imperative plays in moral judgment and deliberation. In her view, there are problems with both received accounts: with the view that it is a pro-

¹⁰⁴ Even if there was no "deep" distinction between foreground and background, something along those lines could still exist as a pragmatic phenomenon. Imagine you justified your saying *x* simply by appeal to the fact that *x* is true. Strictly speaking, you would have failed to satisfy my demand. If I accepted your justification nonetheless, this could be because I chose to treat the claim that none of the potential defeaters are present as an *implicature of your utterance*.

cedure for the derivation of duties and with the view that it is an algorithm for moral deliberation. These are the problems to which her account is meant to provide a solution and, on closer consideration, they pose a challenge that is strikingly similar to the *Particularist Challenge*.

The view that the Categorical Imperative is a source of universal principles of duty that prescribe or prohibit certain kinds of actions (e.g. promise-keeping or lying) exposes Kant's ethics to the *Rigorism Objection* (Herman 1993: 133). Part of the task that Herman sets herself is therefore to develop an account of principles of duty that acknowledges that circumstances can be relevant, i.e. to satisfy *Complexity*. The view that the Categorical Imperative is an algorithm for moral deliberation that takes us from principles of action (maxims) to judgments about the deontic status of particular actions is immune to the *Rigorism Objection* (after all, maxims can refer to all sorts of details and specifics), but faces another problem instead. Given that, on this view, moral deliberation starts from maxims, i.e. from subjective principles of action, there seems to be no way of ensuring that the resulting judgments take account of all and only the morally relevant features of the case (Herman 1993: 135).¹⁰⁵ The other part of Herman's task is thus to develop an account of moral deliberation on which the classification of circumstantial features as relevant is not just presupposed but regulated. In order to accomplish this twofold task, Herman needs to distinguish between default principles from which moral deliberation can begin, and which we can draw on when judging standard cases, and the more complex principles in which deliberation concludes in exceptional cases. This is to say, she needs to accommodate *Asymmetry*. Let us consider the details of her account.

According to Herman, the role of the Categorical Imperative is to supply default principles. These so-called "deliberative presumptions" make up our pre-deliberative moral knowledge and inform the formation of our actual maxims without necessarily determining their content.¹⁰⁶ More specifically, her proposal is this:

The idea is that we are to think of the Categorical Imperative procedure as applying not to actual maxims of action but to a type of action-justification pair: to do

¹⁰⁵ This raises a host of worries, such as the worry that, on this view, Kant's ethics looks suspiciously subjectivist, and the worry that his so-called universalisability test won't be reliable unless the maxims fed into it have the right content (Herman 1993: 135-43). I will return to these issues in ch. 4, when discussing the so-called Problem of Relevant Descriptions.

¹⁰⁶ Actually, Herman's comments on maxims (1993: 144) suggest that it would be more accurate to speak of maxim explication than of maxim formation.

x-type action for y-type reason. I call these pairs “generic maxims”... The rejection of a generic maxim by the Categorical Imperative procedure shows that a certain kind of action may not be done for a certain kind of reason ... The deliberative presumption can be rebutted by reasons ... of a different sort. (Herman 1993: 147-8)

What should we make of this talk of different kinds of reasons? Herman notes that “the standard form of a deliberative principle is derived from a generic maxim of self-interest” (1993: 148).¹⁰⁷ This suggests that we can replace the phrase “for y-type reason[s]” with “for reason[s] of self-interest” or better still: “on self-interested grounds” (a phrase that Herman uses as well).¹⁰⁸ If this is right, then the rejection of a generic maxim will yield a principle of duty of the following form: “Performing an x-type action (or simply: ϕ -ing) on self-interested grounds is impermissible”, for example, “Deceiving on self-interested grounds is impermissible”. It is important to note that this principle itself is universally valid: it holds in all circumstances. So in positing these kinds of grounds-relative principles of duty, Herman satisfies *Universality*. But, according to her, such principles contain defeasible presumptions of the following form: “*Ceteris paribus*, ϕ -ing is impermissible”. Herman argues that, in standard cases, we can draw on these presumptions to *simply judge*, while, in exceptional cases, we can use them as starting points for *moral deliberation* (1993: 145-7).

In the above passage, Herman explains that a “deliberative presumption can be rebutted by reasons ... of a different sort” or, again, by *grounds* of a different sort. According to Kant, there are two kinds of grounds on which our interest in an action or end may be based: self-interested and moral grounds (see e.g. CPrR 5: 79). Kant’s distinction is meant to be exhaustive, so if we assume that Herman is following his lead, then any pursuit that is not based on self-interested grounds is *ipso facto* based on moral grounds. This means that a deliberative presumption (e.g. against lies or acts of deception) can be rebutted or defeated if the agent’s interest in the action or

¹⁰⁷ What Herman says in other places (e.g. 1993: 148, at the bottom) suggests that this is not just the “standard form” of a generic maxim, but rather the only form there is. In light of this, one might wonder why actions on self-interested grounds are being treated as standard cases. We will come back to this question in sect. 3.3.2.

¹⁰⁸ When Herman speaks of *kinds of* reasons she is usually referring to Kant’s distinction between two kinds of grounds. When she speaks of reasons *per se*, her usage of the term is not quite as consistent. Sometimes she uses “reason” to refer to grounds of interest, sometimes to ends, and sometimes to circumstantial features that are included in the maxim.

end is based on moral grounds.¹⁰⁹ However, according to Herman, having moral grounds is not sufficient. It is only one of two conditions of defeat. In addition, the agent's end or concern must "outweigh"¹¹⁰ the end or concern behind the relevant presumption. In short, she argues that a deliberative presumption is defeated if and only if

- (i) the agent's interest in the action or end is based on moral grounds, and
- (ii) the agent's end or concern "outweighs" the end or concern that lies behind the presumption.

Consider the following example.

Deceit for Aid: You deceive A in order to save B's life. Your interest in the action is based on moral grounds and your end, the end to save B's life, "outweighs" the concern for truthfulness that underlies the presumption against deceit. So, in this case, the presumption against deceit is defeated and it is therefore permissible or even obligatory for you to deceive.

Taken together, the two conditions of defeat regulate moral deliberation. The first condition determines whether a feature of the circumstances should be taken into consideration at all, that is, whether it should be included in the actual maxim that we consider acting on.¹¹¹ A circumstantial feature is worthy of consideration and inclusion only if it brings to bear some competing moral end or concern. In the *Deceit for Aid* case, for example, you might have to consider and include the fact that B's life is at stake and the fact that you are in a position to help. The second condition determines whether the feature thus included is genuinely relevant, with its relevance being a function of whether the moral end or concern that it brings to bear is "weighty" enough to defeat the presumption. But how do we decide whether it is?

¹⁰⁹ Herman speaks of presumptions being "rebutted", but since I am using the notion of defeat very broadly (see introduction, sect. 1.2), I take whatever particular mode of defeat she has in mind to fall under it.

¹¹⁰ In my exposition of Herman's view, I will put all terms that are associated with weight and weighing in inverted commas in order to indicate that Herman is anxious to distance herself from this sort of talk (1993: 53). However, as we shall see in sect. 3.3.1, it is doubtful that she can steer clear of it.

¹¹¹ Commenting on this point, Herman says: "The maxim relevant to deliberation must include the special justificatory features of the end or circumstances ... that the agent takes to warrant acting against the presumption. Inclusion of detail in the maxim description is controlled by the structure of deliberation. [A feature] ... may not be introduced for purposes of deliberation unless it is supported by reasons not already excluded by the deliberative presumption" (1993: 150).

Herman acknowledges that we need some “way to rank or compare or weigh different moral considerations” (1993: 153). Her suggestion is that we should turn to Kant’s theory of value: his view that rational agency is a value constraint on willing. Using this as a basis, we can compare the “weights” of competing ends or concerns by asking ourselves whether a given moral end or concern is more or less essential to respecting rational agency than the concern behind the relevant presumption. In sect. 3.3.1, I will suggest that this criterion is empty, but for now the main point to keep in mind is this: Herman specifies two *independent* conditions of defeat, namely that the agent’s interest in the action or end must be based on moral grounds *and* that their end or concern must “outweigh” the concern underlying the relevant presumption. This point will play a crucial role in my objection, which I will now present.

3.2 Two Independent Conditions of Defeat? An Inconsistency

The above presentation of Herman’s account glosses over two ambiguities. The first ambiguity has to do with the fact that Herman speaks of the Categorical Imperative and the two conditions of defeat in largely procedural terms,¹¹² associating them with stages of decision making. First, we recognize the action that we are contemplating as an instance of a presumptively impermissible kind, then we ask ourselves whether our grounds are self-interested or moral, and then we either simply judge or enter moral deliberation in order to weigh our end or concern against the end or concern that lies behind the presumption.

Although very common, it is notoriously unclear how this talk of the Categorical Imperative (in particular, the Formula of Universal Law) as a “test” or “procedure” is to be understood. If taken literally, it suggests a decision tool reading: the view that the Categorical Imperative is primarily a decision tool that agents can use to navigate an independent moral reality (Timmons 1997). If taken as a metaphor or illustration, then it is compatible with the more common view that the Categorical Imperative is primarily a *norm* that determines the deontic status of actions,¹¹³ and only secondarily,

¹¹² Among other procedurally-sounding expressions, she refers to the Categorical Imperative as the “CI procedure” (see e.g. Herman 1993: 132).

¹¹³ Rawls, whose John Dewey Lectures “Kantian Constructivism in Moral Theory” and other writings have played a major role in entrenching this procedural jargon in the debate, is explicit about using the procedural terminology metaphorically, as a way of illustrating a norm (see 1980: 571, where he speaks of his “model-conception”). Contemporary Kantian Constructivists follow him in this regard. I will come back to these two ways of understanding the talk of the Categorical Imperative as a procedure in ch. 4, sect. 2.2 and 4.2. Kantian Constructivism will be the topic of ch. 5.

if at all, also a decision tool.¹¹⁴ In my view, the latter reading is more plausible because it takes seriously Kant's idea that the Categorical Imperative is the *supreme principle* of morality and that it has to be sought out in a *metaphysical* investigation. In light of this, it seems more charitable to resolve the first ambiguity (call it the "norm-vs.-decision-tool ambiguity") in Herman's discussion in favour of the view that the Categorical Imperative is primarily a norm and to also extend this view to the two conditions of defeat with which she supplements Kant's highest principle. However, it is important to keep in mind that, occasionally, Herman's comments and remarks *may well* pertain to their secondary function, their function as decision tools. We will return to this point in sect. 3.3.2.

There is a second, more important ambiguity. In commenting on the distinction between self-interested and moral grounds, Herman is anxious to point out that we should not confuse the *ends* being pursued with the *grounds* of the agent's interest in these ends (1993: 148). Pursuing an end on self-interested grounds, be it a morally praiseworthy end or a self-serving end, is a matter of pursuing it merely because it is something that we happen to want; pursuing an end on moral grounds is a matter of making sure that, in our pursuit, be it what it may, we are appropriately sensitive to moral constraints. Herman notes that this distinction between ends and grounds of interest entails a conception of "morality ... [as] the regulative norm for our interests" (1993: 148). On this conception (we can call it "*Morality as a Limit*"), the distinctive feature of actions on moral grounds is that they reflect an appropriate sensitivity to the limits of morality, not that they aim at certain ends rather than others.¹¹⁵ Here is an example.

¹¹⁴ Note that this question arises for Herman's own account as much as it does for the two received accounts that she considers. In her case, the question is, strictly speaking, whether the Categorical Imperative is primarily an element of a decision making tool or a norm that determines the default deontic status of certain action kinds.

¹¹⁵ Herman changes her mind on this point. In her recent writings (e.g. her 2006, 2007: 254-275, 2011, 2013), she argues that acting on moral grounds is a matter of formally good reasoning (reasoning in accordance with the Categorical Imperative) on the basis of good premises (obligatory ends). If this is how acting on moral grounds is understood, then pursuing an obligatory end is a necessary condition. There are two aspects of this more recent account of Herman's that are worth highlighting. First, in her recent work, she speaks of obligatory ends rather than morally good or morally praiseworthy or other-regarding ends. This suggests that she is thinking of Kant's notion of an obligatory end (MM 6:382-7). If this is right, then the ends in question are not *morally contingent* ends (like the ones we are discussing in the main text above), but rather *morally necessary* ends: ends whose pursuit is per definition morally worthy, ends which one wouldn't count as pursuing if one wasn't sensitive to moral demands. This brings me to the second point. My main objection to Herman's more recent account is that there is a problematic duplication of norms. As I said, actions in pursuit of obligatory or morally necessary

Sick Aunt: You take time off from work in order to go and visit your sick aunt. You do this in order to cheer her up and get her back on her feet.

The ends you pursue in visiting your aunt are certainly the kinds of ends we tend to consider morally laudable. After all, you are concerned with her well-being. But from what we know, we cannot rest assured that your action is based on moral grounds. Perhaps you would not have cared about her well-being and would not have visited, if she was not your favourite aunt, or perhaps you would have dropped everything at work and left your colleagues in the lurch, even if all your aunt had was a harmless cold. If some such counterfactual was true, then your visit would *not* exhibit a proper sensitivity to moral constraints.¹¹⁶ According to the conception of *Morality as a Limit*, the moral character of our actions cannot be determined on the basis of a description of our ends,¹¹⁷ it cannot be read off the label, as it were.¹¹⁸ Instead, it hinges on what we would be disposed to do in various kinds of circumstances. Your end of getting your aunt back on her feet, for instance, could be the manifestation of a healthy concern for a relative or of an anxiety disorder, depending on what you would do if all she had was a harmless cold.¹¹⁹

ends are *ipso facto* morally worthy, so when we say that someone acted in pursuit of an obligatory end, it is no longer an open question whether they complied with the Categorical Imperative. The two notions represent different perspectives on one and the same norm (ch. 1, sect. 4.1), not two norms that work together. Herman fails to acknowledge this when she says that we need to *both* start from obligatory ends *and* reason in accordance with the Categorical Imperative. Since Herman's latest response to the *Rigorism Objection* (2013) draws on the same duplication, I do not find it very convincing. In ch. 4, sect. 3, I will explain why this kind of duplication must undermine any attempt to defend the Categorical Imperative against its critics.

¹¹⁶ Of course, we may not know what we would have done. This is the whole point of Kant's so-called opacity thesis, his view that we are intransparent to ourselves in the sense that we cannot be certain of the grounds of our maxims. See e.g. G 4:407 and R 6:63.

¹¹⁷ The reverse is true as well: we cannot determine whether an action is directed at a morally laudable end, in the ordinary sense of that term, merely on the basis of whether it is based on moral grounds. After all, it is possible to pursue one's own happiness on moral grounds. More generally, this suggests that the two are more or less independent: that we can pursue our own well-being on moral grounds and other people's well-being on self-interested grounds. I say "more or less" because I want to leave open the possibility that some of the terms or labels that we use to pick out ends pick them out *as* morally good or evil. To describe someone's end in this way would be to imply that their action is based on moral grounds or self-interested grounds, respectively.

¹¹⁸ This is true if the labels do not stand for necessary or obligatory ends (see fn. 115), so when I speak of ends as labels, I am assuming that the commitments that lie behind these labels are not necessarily in line with moral demands.

¹¹⁹ One could opt for a weaker reading of Herman's talk of grounds. I have assumed that acting on moral grounds requires appropriate sensitivity to moral constraints in the strong sense that, if things had been different, one would still have done what morality demands. In ch. 3, sect. 3.1, I argue that this is what Kant means when he speaks of acting on moral grounds or acting from duty. However, when Herman says that "the designation 'self-interest' signals the moral or justificatory status of the reason in question" (1993: 148), this might suggest that she has in mind a weaker requirement. On this

Despite the fact that Herman insists on the Kantian distinction between grounds and ends, and despite the fact that she appreciates its implications (the conception of *Morality as a Limit*), she repeatedly collapses this distinction.¹²⁰ One example is her talk of motivational overdetermination: the idea that it is possible to act on moral and on self-interested grounds at the same time (1993: 150). If acting on moral grounds was a matter of pursuing certain ends rather than others (we can call this conception the conception of “*Morality as Pursuit of Ends*”), this would be possible indeed. It could be, for example, that you visit your aunt in order to both cheer her up *and* get some extra cash. But if acting on moral grounds is a matter of being properly sensitive to moral constraints, then motivational overdetermination is not a possibility. For what would it mean for your pursuit to be and also not be sensitive to moral constraints?¹²¹ (Of course, if you visit your aunt to get some extra cash, you are still within the limits of morality in the sense that you are not doing anything impermissible, but this is not enough for your action to be based on moral grounds, as we defined it. The critical question is whether, if you did not expect her to give you any cash, you would still be going – and here the answer is either yes or no.) This confusion or ambiguity about whether ends are distinct from grounds or whether they are the same (the grounds-ends ambiguity), and the corresponding mix-up of two conceptions of morality is, in fact, a confusion that runs very deep in Herman’s account.

Recall that we concluded the previous section by highlighting the independence of Herman’s two conditions of defeat. We can capture the sense in which the two are taken to be independent in the following thesis.

reading, acting on moral grounds requires sensitivity to moral constraints only in the weak sense that one is sensitive to whether treating a certain consideration as a reason is licensed by morality. E.g., on this weaker reading, if your aunt’s illness justifies taking time off, then, in treating it as such, you’d be acting on moral grounds. This leaves open whether you would have taken time off even if your aunt had a harmless cold, and therefore whether you meet the stronger requirement as well. The important point, for my purposes, is that, even on the weak reading, acting on moral grounds entails that you did the right thing. I will get to this point shortly.

¹²⁰ In this respect, the account of moral motivation that Herman advocates in her recent writings is more consistent. See e.g. Herman 2007: 1-28.

¹²¹ Of course, you might be sensitive to some moral constraints but not others. But this is not what people mean when they speak of motivational overdetermination.

Independence Thesis: The fact that a presumptively impermissible action is performed on moral grounds (that condition (i) is satisfied) does not entail that it is permissible or obligatory (that condition (ii) is satisfied as well).¹²²

My main objection to Herman's account is this: *if* we assume that she endorses the *Independence Thesis* and take seriously her explicit remarks on ends and grounds, which point towards *Morality as a Limit*, then her account is inconsistent. Both antecedents will be scrutinized below, but for now let us proceed on the assumption that they are true and explore the inconsistency.

According to *Morality as a Limit*, whether we act on moral grounds or not is a function of whether we are appropriately sensitive to the moral constraints that regulate actions of the relevant kind across different circumstances. With this in mind, let us revisit the *Deceit for Aid* case. If you deceived A on moral grounds, then, in doing so, you were sensitive to the moral constraints that apply to acts of deception across different circumstances. This implies that you were sensitive to the kind of difference that the fact of B's life being at stake makes to the deontic status of deception. But if this is true, then it is no longer an open question whether your concern for B's life "outweighs" the concern for the value of truthfulness. If it did not, then you could not be said to have acted on moral grounds, even if you were driven by a concern for B's life, which is *prima facie* a morally laudable concern. In more general terms, if Herman's first condition of defeat is satisfied, then it is no longer an open question whether her second condition of defeat is satisfied as well. An action that is performed on moral grounds, be it presumptively impermissible or not, is *ipso facto* an action that is permissible or obligatory. So if Herman endorses the *Independence Thesis* and the conception of *Morality as a Limit*,¹²³ then her overall account is inconsistent.

¹²² Note that this is simply a reversal of the claim that motivational overdetermination is possible. The notion of motivational overdetermination is supposed to convey that one can act on self-interested grounds while being properly sensitive to moral constraints (which is impossible because, *qua* properly sensitive, one is *ipso facto* acting on moral grounds). The *Independence Thesis* says that one can act on moral grounds while failing to be properly sensitive to moral constraints (which, again, is impossible because, insofar as one's action is based on moral grounds, one is *ipso facto* properly sensitive to moral constraints).

¹²³ I say "if" because it is not entirely clear whether she really subscribes to these doctrines. On the one hand, given how we resolved the first ambiguity in her essay (the norm-vs.-decision-tool ambiguity), one could argue that her distinction between two conditions of defeat pertains only to the secondary level, the decision tool level. In that case, the distinction would be a distinction between two stages of decision making, so that, on a deeper level, the conditions wouldn't have to be independent after all. I will come back to this deflationary reading of the *Independence Thesis* in sect. 3.3.2 below. On the other hand, one could argue that the second ambiguity (the grounds-ends ambiguity) should be re-

Of course, if she endorsed the conception of *Morality as Pursuit of Ends* instead, then the *Independence Thesis* could be upheld. For, on this conception, acting on moral grounds is a matter of pursuing morally laudable ends, and surely it must be possible to pursue such an end *and* do something impermissible in the process. Imagine you abstained from deceit despite the fact that B's life was at stake. According to *Morality as Pursuit of Ends*, if your end or concern was a morally laudable one, e.g. if it was a concern for truthfulness, then you acted on moral grounds, even if what you did was impermissible because B's life was more important. In sect. 3.3.1, we will see why the conception of morality under consideration leaves room for this possibility.

But before we move on, let us ask why Herman's account takes the form that it does. The hypothesis that helped us to make sense of the structure of Korsgaard's *Two Standpoint View* comes in useful once again. If we abstract from the specifics of Korsgaard's case, the hypothesis was the following: the fact that the view contains incompatible elements can be understood as an unwanted side effect of an attempt to escape the *Trilemma* by dividing labour. In the following sections, we will find that the same explanation is available here, because, taken together, the two competing conceptions of morality that are mixed up in Herman's account can satisfy all three desiderata. More specifically, we will see that *Morality as a Limit* satisfies *Universality* and *Complexity*, but not *Asymmetry*, while *Morality as Pursuit of Ends* satisfies *Complexity* and *Asymmetry*, but not *Universality*. If this is indeed what each of the two conceptions can and cannot accomplish, then we can understand why Herman tries to combine elements of both, at least implicitly. *That* she does so is, once again, not immediately obvious, however, and that has to do with the grounds-ends ambiguity: with the fact that she goes back and forth between distinguishing ends from grounds and collapsing that distinction.

3.3. Two Responses on Herman's Behalf

Since, once again, the problem here takes the form of an inconsistency, we can, once again, try to abandon either one of the two incompatible elements, to see if one of the two resulting accounts satisfies all three desiderata after all. In sect. 3.3.1, I will

solved in favour of the conception of *Morality as Pursuit of Ends*, in which case Herman would not be committed to the conception of *Morality as a Limit*. I will explore this option in sect. 3.3.1.

consider a defence of Herman's account that is based on a rejection of *Morality as a Limit*, and in sect. 3.3.2 a defence that deflates the *Independence Thesis*.

3.3.1 The Competing Moral Concerns Response

Recall the grounds-ends ambiguity discussed in the previous section: the ambiguity in Herman's use of the terms "moral grounds" and "self-interested grounds". How we resolve this ambiguity bears on the conception of morality that we take her to be presupposing. Her explicit remarks on this issue suggest that she endorses *Morality as a Limit*, but some of her other commitments point towards a different conception, a conception of *Morality as Pursuit of Ends*. In this section, I want to explore the merits and demerits of taking these other commitments and the concomitant conception of morality more seriously.

In ch. 1, sect. 3, we considered Timmermann's claim that Kant's comments on grounds of obligation contain a theory of pro tanto duties. The conception of *Morality as Pursuit of Ends*, as characterized above, is exactly the kind of theory that he has in mind. According to this theory, there is a plurality of potentially competing moral ends or concerns with corresponding pro tanto duties to pursue those ends or honour those concerns. We already noted some of the advantages of such a view: that it makes room for moral residues and for the idea that moral concerns can outweigh each other. Now, with Herman's arguments on the table, we can add a further item to this list. A theory of pro tanto duties can accommodate the common intuition that, even when we misjudge what morality demands in a given situation, it is still possible that we are driven by a moral concern and not by self-interest. In the *Deceit for Aid* case, for example, where someone's life is at stake, it is impermissible to tell the truth, and if you told the truth out of a concern for truthfulness, you would be mistaken about which moral concern takes precedence in the case at hand. However, surely, one might say, you would still be driven by a *moral* concern, surely your grounds for being honest, i.e. your aspiration to be truthful, were not in any way *self-interested*. This line of thought suggests that Herman might be better off fully embracing *Morality as Pursuit of Ends* and the elements of her account that go along with that, such as the *Independence Thesis* and the possibility of motivational overdetermination.¹²⁴

¹²⁴ That Herman is attracted to these advantages of *Morality as Pursuit of Ends* comes out most clearly in the final section of her essay (1993: 151-7). There she speaks of the "weight or gravity" of considera-

As I indicated earlier, this would allow her to satisfy the *Complexity* desideratum and the *Asymmetry* desideratum. Let me explain how *Morality as Pursuit of Ends* satisfies the former, and postpone the discussion of the latter to the following section. Think of the presumption against deceit, which says that deceiving is by default impermissible. According to *Morality as Pursuit of Ends*, this presumption expresses only one of many moral concerns. In an exceptional case, where additional moral concerns are at stake, e.g. the concern for someone's life, the concern behind the presumption against deceit might be outweighed and hence defeated. In such a case, it would be permissible or even obligatory to deceive, and that would be because of features other than the ones that make the act an act of deception. So this is how *Morality as Pursuit of Ends* satisfies *Complexity*. But what about *Universality*?

Recall Herman's take on the deliverances of the Categorical Imperative: when a generic maxim is rejected, there is a presumption against actions of the designated kind. This presumption, though not itself universally valid, is embedded in a universally valid principle of duty, which says that it is impermissible to perform actions of this kind on self-interested grounds. So far, so good. The problem arises when we try to say the reverse: that it is always obligatory, or at least permissible, to perform presumptively impermissible actions on moral grounds or, more generally, that it is always, in all cases, obligatory, or at least permissible, to perform actions on moral grounds. If "on moral grounds" means "in pursuit of a moral end or concern", then these claims are false because moral ends and concerns can be outweighed by other moral ends and concerns. Think of a version of the *Deceit for Aid* case where an additional moral concern is at stake: the person you would have to deceive is the very person whose life is at risk because they want to kill themselves. Here one might argue that the concern for this person's autonomy strengthens the case for being honest and that, together with the concern for truthfulness, these two outweigh the concern for their life. Thus, the *only* universally valid principles of duty specifying what we are allowed or not allowed or required to do on moral grounds would be principles of pro tanto duty, i.e. principles located on the contributory level. The kinds of principles that are needed to satisfy *Universality*, namely universally valid principles of overall duty, are out of reach. Of course, it could be said that Herman is

tions, mentions the need for a "way to rank or compare or weigh different moral considerations", and takes herself to have reached "a conclusion ... in terms of *relative* value".

free to reject the *Universality* desideratum, especially because she has provided an alternative account of the role of Kant's Categorical Imperative. So the question is: is there any substantive reason, any non-exegetical reason that speaks against such a move?

When Kantians criticize theories of pro tanto duties, they often argue that judgments about weight are ultimately subjective, that such judgments lack criteria of correctness, and hence that they cannot be shared or fruitfully disputed (see e.g. Schüssler 2012: 3).¹²⁵ Herman is aware of this worry and of the fact that “any talk of weighing or comparing is supposed to be out of court in Kantian ethics” (1993: 153). Her attempt to show that her own account escapes this worry and avoids this talk is not entirely convincing, however. For although she introduces a criterion for ranking moral concerns (namely their relative significance with respect to the goal of respecting rational agency), and although she insists that this criterion honours the idea that moral concerns are not scalar (Herman 1993: 155), she ultimately fails to explain how this criterion would help: e.g., how it could be used to adjudicate disputes about the “weight” or significance of truthfulness, on the one hand, and other’s needs, on the other. In fact, she admits that translating such a question into value terms “does not make the answer obvious” (1993: 155). Thus, one is left with the sense that her judgments of “relative value” (1993: 155) are not intersubjectively intelligible after all, and that invoking such judgments threatens the objectivity of moral judgments. (We will return to the connection between *Universality* and objectivity in the following chapter.)

Before we move on, we should revisit the intuition that made the conception of *Morality as Pursuit of Ends* seem appealing in the first place: the intuition that overestimating the importance of a certain moral concern in relation to another (e.g. the importance of truthfulness in relation to the importance of someone’s life) does not make the resulting action self-interested. According to proponents of *Morality as a Limit*, acting on self-interested grounds is doing something merely because one feels like doing it. On the face of it, this is not what is going on in our example. But let us dig deeper and think about the kind of person, the kind of character who would pri-

¹²⁵ If this is true, if they lack criteria of correctness, then the problem is, in fact, more fundamental: not only can such judgments not be shared, they cannot even be thought of as consistent or inconsistent with one’s other judgments. This line of thought calls to mind Wittgenstein’s private language argument (1953: §§244–271). We will return to this point in ch. 3, sect. 2.3.

oritise being truthful over someone else's life. What kind of person would that be? Perhaps they are a cold-hearted person, someone who lacks empathy and so does not feel like helping other people, or maybe they are a pedantic person, a stickler for principles who feels discomfort at the thought of violating a rule that was hammered into their head when they were young. Whichever story we tell, it will be a story that makes reference to some kind of bias or vice, to some flaw in their affective dispositions. So once we fill in all the details, it is no longer implausible to say that our protagonist refuses to help merely because they don't feel like helping. For while their concern for truthfulness may *appear* like a moral concern when considered superficially (recall the above point about ends as mere labels), this appearance does not stand up to scrutiny.¹²⁶

3.3.2 The Deflationary Response

If abandoning *Morality as a Limit* threatens the objectivity of our moral judgments, then there is only one alternative: we have to re-interpret the remarks that suggest that Herman endorses the *Independence Thesis*. We already insinuated how this could be done. In sect. 3.2, we resolved the norm-vs.-decision-tool ambiguity in Herman's text in a way that allowed that Kant's Categorical Imperative and her own conditions of defeat might have two functions. Their primary function, we said, must be to serve as norms that determine the deontic status of actions. However, in settling on this reading, we left open the possibility that they could have a secondary function as well: that they could *also* serve as decision tools. With this in mind, we noted that some of Herman's claims might concern this secondary function only. What if this was true of the *Independence Thesis*? What if this thesis should be read in a deflationary way?

Recall why we thought that Herman's account was inconsistent: given *Morality as a Limit*, the satisfaction of the first condition of defeat entails the satisfaction of the second, so, at bottom, the two must be one – they cannot be independent. But perhaps we misunderstood her, perhaps Herman isn't actually saying that they are. Perhaps the distinction between two conditions is in fact a distinction between two deliberative steps, two steps that help us to determine whether the one and only condi-

¹²⁴ I think that this line of thought could be developed into an argument for why it is appropriate to reduce one of the two modes of defeat distinguished above (introduction, sect. 1.2) to the other, i.e. to treat cases of outweighing as cases of undermining.

tion of defeat is met.¹²⁷ On this deflationary reading of the *Independence Thesis*, the second step is a way of reassuring ourselves of the outcome of the first: by reflecting on the relative weight of the moral concern that seems most important, we seek to confirm our hunch that there are moral grounds for pursuing the corresponding course of action. We try to reassure ourselves that a *prima facie* defeater¹²⁸ is a genuine defeater. For example, knowing full well that a threat to someone's life tends to call for radical measures, you might guess that this fact defeats the default deontic import of deception, but in order to reassure yourself that this is how things stand, you might find it useful to ask yourself whether the concern for this person's life really is more essential to respecting rational agency than the concern for truthfulness, or whether it only seems to you that way, e.g. because you two have some special bond. At bottom, there is only one condition, but two methods for finding out if it is met.

As we have seen, the Deflationary Response does away with the inconsistency. But how does it fare in relation to the *Particularist Challenge*? Here we need to pick up where we left off in sect. 3.1. According to Herman, the rejection of a generic maxim yields a universal principle of duty of the form “ ϕ -ing on self-interested grounds is impermissible” (*Universality*). This principle contains a defeasible principle (a deliberative presumption) of the form “*Ceteris paribus*, ϕ -ing is impermissible” (*Complexity*). When considering this account of the deliverances of the Categorical Imperative, we wondered why cases of actions on self-interested grounds were being treated as if they were simpler. Why is it, for example, that, on Herman's account, when it comes to acts of deception on self-interested grounds, we can simply judge without having to think about potential defeaters, whereas, when it comes to acts of deception on moral grounds, we need to deliberate? How has Herman earned the right to make these claims that suggest that *Asymmetry* is satisfied? Now that we have disentangled the different threads that run through her argument, we are in a position to see that she has this right *only* if she insists on a non-deflated *Independence Thesis* and the con-

¹²⁷ There is plenty of textual evidence for a distinction between two stages or steps. Herman says, for example: “Prior to deliberation the agent must *both* identify her proposed action as of a particular moral kind and determine the nature of her interest in the action (or its end)” (Herman 1993: 151). But since she uses procedural terms all the way through, it is difficult to tell whether it is merely a distinction between stages or also a distinction between conditions or norms.

¹²⁸ In this context, “*prima facie*” *means* *prima facie*, and not *pro tanto*, as it does in Ross' writings. See fn. 44 above.

comitant conception of *Morality as Pursuit of Ends*. Notably, this is true despite the fact that the alternative view that emerges from the deflationary reading allows us to talk of defeaters.

According to *Morality as Pursuit of Ends*, each deliberative presumption refers to a moral end or concern. The presumption against deceit, for example, refers to the concern for truthfulness. So in saying that acts of deception on self-interested grounds are impermissible, we are basically saying that they are impermissible in a certain respect (qua untruthful) and that, in standard cases, where the only other ends or concerns at stake are self-interested ends or concerns, they are also overall impermissible. This is why the deliberative and explanatory burdens in standard cases are relatively light: in order to grasp or explain why an act of deception is overall impermissible, we only need to think about or mention that it is deceptive and, as such, untruthful. We do not need to think about or mention other moral ends or concerns. In saying that a given act of deception on moral grounds is permissible or obligatory, by contrast, we are already invoking further moral ends or concerns. We are saying that the deontic import of deceit-qua-untruthfulness is defeated. This is why the deliberative and explanatory burdens in exceptional cases are heavier: in order to grasp or explain why a given act of deception is overall permissible or obligatory, we need to think about or mention the moral ends or concerns that function as defeaters.¹²⁹

So if Herman endorsed *Morality as Pursuit of Ends*, she could explain why there is an asymmetry in deliberative and explanatory burdens, but if she opts for *Morality as a Limit* – as we are currently assuming – then no such explanation is forthcoming. Recall that, according to the deflationary reading, her two conditions of defeat are, at bottom, one and the same. The only question is whether, in performing a certain action, one would be acting on moral grounds (in which case the action is permissible or obligatory) or on self-interested grounds (in which case it is impermissible). But if this is right, then Herman’s take on the deliverances of the Categorical Imperative is in need of revision. On the revised account, the rejection of a generic maxim yields a universal principle of duty of the form “ ϕ -ing is impermissible on self-

¹²⁹ Actually, the burdens are heavier *whenever* an additional end or concern is at stake, even if and when it turns out that this end or concern does not defeat the end or concern behind the presumption, meaning that the case is not exceptional after all. In such cases, we still have to at least *deliberate about* the additional end or concern, if not mention it in an explanation of the action’s deontic status. This is exactly what the *Independence Thesis* would predict.

interested grounds and permissible or obligatory on moral grounds”. Taking into account what we said in sect. 3.2, this means: ϕ -ing is impermissible if, in ϕ -ing, one would fail to be properly sensitive to the *absence* of defeaters, and, vice versa, permissible or obligatory if, in ϕ -ing, one would be properly sensitive to their *presence*. But this, of course, is nothing but an overcomplicated way of saying that ϕ -ing is impermissible unless there are defeaters.¹³⁰ And if this is what principles of duty really say, then whether or not they contain actual lists of defeaters (cf. the Fine Print Reading horn),¹³¹ the explanatory and deliberative burdens are the same across the board.¹³²

4. Taking Stock: A Comparison and a Note on Defeasibility

In discussing Korsgaard’s and Herman’s attempts to escape the *Trilemma* and meet the *Particularist Challenge*, we discovered many structural similarities between their accounts: both appeal to the idea that, despite being universally valid, principles of duty are (or contain) defeasible principles and both end up with views that are not entirely consistent. More specifically, they both commit themselves to the claim that certain actions both *do and do not* violate the Categorical Imperative.¹³³ Moreover, in each case, the inconsistency is concealed. In Korsgaard’s account, the inconsistency is inconspicuous, in part, because the discussion is conducted on a metaethical level

¹³⁰ Why is it overcomplicated? As my discussion has brought out, acting on moral grounds entails that one is doing something permissible or obligatory, but it is, in fact, more demanding than that (see fn. 119). Hence, in order to explain why an action has the deontic status that it has, we do not need to refer to whether the agent has a good character or will, which goes beyond the question of whether they did what they ought to have done in this case.

¹³¹ If they do not, if they are hedged rather than unhedged principles, then the worry would be a different one, namely that the hedging clause “unless there are defeaters” seems to be the kind of hedging clause that renders the principle to which it is attached *trivially true*. For a discussion of this objection see e.g. McKeever and Ridge 2006: 121-3. Note that there is actually some evidence that Herman is advocating unhedged principles with a long list of unless-clauses or a fine print. For, like Korsgaard, she tries to distinguish between parts of principles that lie in the foreground and parts that lie in the background (1993: 144-5). To the extent that she does, she exposes herself to the criticism that we levelled against the Implicit Awareness Response in sect. 2.3.2.

¹³² It could be argued that there is an asymmetry nonetheless. After all, such principles do honour the idea that certain features have their deontic import by default and that this import can be defeated. This is true, but they honour it in name only. To say that defeaters are absent is to say that the case is a standard case; to say that defeaters are present is to say that the case is an exceptional case – but if, having said that, one goes on to treat their presence and absence symmetrically, one hasn’t really done justice to the difference between them.

¹³³ In Korsgaard’s account, this is true of actions from provisionally universal principles: they are not supposed to be morally problematic, but in the end they must be. In Herman’s account, this is true of impermissible actions that are based on moral grounds: insofar as they are impermissible, they must be violations of the Categorical Imperative, but insofar as they are based on moral grounds, they cannot be violations – at least on her official understanding of the notion of moral grounds, i.e. the *Moral-ity as a Limit* understanding.

and, in part, because *de facto false* moral views are treated as if they were merely *fallible* or *incomplete*. In Herman's account, the inconsistency is inconspicuous because her use of the distinction between moral and self-interested grounds is ambiguous in a way that evokes two different conceptions of morality. In both cases, we suggested that the endeavour to hold together two elements that are ultimately incompatible is nonetheless understandable, because, given the inability of any one element to satisfy all three desiderata on its own, this endeavour can be understood as an attempt to divide labour. That such a division of labour is indeed what Korsgaard's and Herman's accounts *would* require was confirmed by our unsuccessful attempts to defend them. We saw that abandoning *Universality* in its original form (2.3.1, 3.3.1) would saddle them with views that cannot make sense of the objectivity of our moral judgments (a worry to which we shall return in ch. 3), whereas abandoning the element that is in contradiction with *Universality* would leave them impaled on the Fine Print horn of the *Trilemma*.

In the following chapter, I will argue for a response to the *Particularist Challenge* that shares an important feature with Korsgaard's and Herman's accounts: like theirs, mine is an account of our principles of duty as both universal and defeasible. For this reason, I want to conclude the present chapter with a brief note on the two conceptions of defeasibility discussed above. Philosophers' reasons for presenting certain rules, claims or statuses as defeasible are many, and thinking about *why* someone chooses to appeal to defeasibility is usually a good guide to what they mean by it. One such reason is the wish to highlight the fallibility of our judgments, beliefs and theories, the wish to do justice to the common sense idea that what we take to be true at a given time may well turn out to be false later. Which conception of defeasibility is appropriate for this purpose? Think about a scientific hypothesis: a scientific hypothesis is defeasible in the sense that there might be counterexamples that disprove it. When such a counterexample is found, we have to admit that our hypothesis was wrong and hence reject or revise it. Korsgaard's remarks on the appropriate reaction to an exception (namely, going back and revising the original principle) suggest that this is the conception that she has in mind. On reflection, this is surprising, however. After all, *her aim* is not to highlight the fallibility of our moral judgments but rather to accommodate the idea the deontic purport of features varies with circumstances (of course, we can be mistaken about how they vary as well, but that is

not the point). Herman's choice is similarly curious. In sketching the problem and discussing rival accounts, she is very clear that her concern is with how to arrive at principles that mention all and only morally relevant features, given how that varies with circumstances (1993: 132-46). But the conception of defeasibility that her solution draws on is a conception that we would usually associate with a different aim. Hers is a conception of *pro tanto* defeasibility, according to which the default deontic import of a feature is defeasible not in the sense that it varies with or depends on circumstances, but in the sense that it can be *outweighed* by the default import of other features. This conception is, in my view, a more natural choice for someone who aims to accommodate the phenomena discussed in ch. 1, sect. 3, e.g. the idea that an obligation retains its force and leaves a residue even when it is defeated and that, therefore, there are reasons for regret.

The moral of the above story is this: if we are to give an account of defeasible principles of duty, we need to make sure that it is fit for purpose. In the following chapter, I hope to develop such an account.¹³⁴

¹³⁴ Of course, my account of defeasible principles is not the only one that was ever designed with a view to accommodating the idea that the normative purport of features varies with circumstances. Another example is the account presented by H.L.A. Hart in "The Ascription of Responsibilities and Rights". In this paper, Hart argues that "there are characteristics of legal concepts which make it often absurd to use in connection with them the language of necessary and sufficient conditions" (1948: 173). Instead, he argues, we should distinguish between initial conditions and defeaters. Under normal circumstances, the satisfaction of the initial conditions is sufficient for the applicability of the relevant legal concept; under exceptional circumstances, in which defeaters "bring[] the case under some recognized head of exception" (Hart 1948: 174), it is not. We can see how this account is fit for purpose.

Chapter 3

Case-Scope or Agent-Scope?

The Inferentialist Response

1. Korsgaard's Anti-Particularist Argument Revisited

In the previous chapter, we considered Herman's and Korsgaard's attempts to escape from the *Trilemma* that the *Particularist Challenge* was shown to generate. Both aim to show that Kant's principles of duty can admit of exceptions, and both hope to accomplish this by reinterpreting *Universality* such that universal validity and defeasibility come out as compatible. We found that these attempts are unsuccessful because they appeal to the wrong conceptions of defeasibility. The conceptions they appeal to, we said, are not fit for purpose: not fit for the purpose of accommodating the idea that our moral obligations vary with circumstances. In the present chapter, I will argue for a reinterpretation of *Universality* that, I think, is better suited for this task. In order to motivate this reinterpretation, let us first revisit Korsgaard's anti-particularist argument. Recall how the argument proceeds:

1. Willing means determining yourself to be a cause, not being moved by the causal operations of incentives within you.
2. Determining yourself to be a cause means following an incentive on the condition that doing so is representative of your practical identity.
3. Willing particularistically, if possible, would mean following an incentive in its full particularity, i.e.
 - a. as not "representative of any sort of type" of policy or standing commitment,
 - b. or as "in no way [...] further describable" (2009: 76).
4. Following an incentive in its full particularity means *not* making your following it conditional on whether doing so is representative of your practical identity.
5. Thus, it is impossible to will particularistically.

This argument aims to both refute particularism and to defend a Kantian brand of generalism – the brand that *Universality* is intended to capture. Let us consider how the argument is supposed to achieve the latter aim. In ch. 2, sect. 2.1, we noted, but

did not dwell on the fact, that Korsgaard offers two versions of the third premise. When fleshing out what it would mean to follow an incentive in its full particularity, she says, on the one hand, that it would mean following an incentive as not “representative of any sort of type” (of policy, say) and, on the other, that it would mean following it as “in no way ... further describable”. Focusing on the former (3a), we thought that she was making a point about the constitutive function of practical identities: since it is by adopting a practical identity that I constitute myself as a certain kind of person, as a person who does certain kinds of things and abstains from doing others, such an identity has to consist of standing commitments that are robust across the board. Accordingly, to follow an incentive on the condition that it is representative of such an identity must be to follow it as an instance of a form of conduct that we are committed to in this robust way. Let us call this the “standing-commitment thread” of her argument.¹³⁵ There is another thread, however. For if we focus on premise 3b, then Korsgaard’s main point seems to be a different one: if there was such a thing as particularistic willing, it would not be a *conceptual* state or activity¹³⁶ because, just in virtue of drawing on concepts, we already connect the particular case before us to other cases. But if we followed incentives *non-conceptually*, then we could neither succeed nor fail to act in a way that is representative of our practical identity, because practical identities are essentially conceptual entities.¹³⁷ After all, a practical identity is a “*description* under which you find your life to be worth living and your actions to be worth undertaking” (Korsgaard 1996b: 101, my emphasis). Let us call this the “willing-is-conceptual thread” of Korsgaard’s argument. Superficially, these two argumentative routes lead to the same conclusion, but we will see that the brands of Kantian generalism that they establish are very different.

The standing-commitment thread of the argument via premise 3a lends support to the view that willing is following incentives as instances of principles that hold *for*

¹³⁵ For a reading of Korsgaard’s argument in the *Sources* along these standing-commitment lines see Stern 2015: 58.

¹³⁶ We will briefly return to the question of whether willing is a state or an activity in connection with its temporal structure in sect. 3.2 below.

¹³⁷ This is a rough outline of an argument that we will elaborate on in sect. 2.1 below, albeit in a slightly different form. Instead of asking what character a choice would have to have so that it be able to agree or disagree with a practical identity, we will ask a more general question, namely, what character a representation of an object would have to have so that it be able to agree or disagree with other representations of that object.

all cases or in all circumstances, and, relatedly, that good willing is willing in accordance with principles that really are universally valid in this sense. Since this is a claim about willing in general, a claim that extends to willing in accordance with principles of duty (e.g. ignoring an incentive to say something because it would be lying), it implies that the possibility of morally good willing rests on there being principles of duty that really do hold for all cases or in all circumstances. As such, the standing-commitment thread of the argument is a consideration in favour of what I will call the Case-Scope Reading of *Universality* – the reading that, despite all their attempts to tone it down, still figures as a benchmark in Herman’s and Korsgaard’s accounts and, ultimately, saddles them with the *Trilemma*.

*Universality*_(CS): The features in virtue of which actions have their deontic statuses can be captured in principles of duty that hold for all cases or in all circumstances.

In addition, *Universality*_(CS) draws support from what Dancy calls the “*subsumptive* conception of rationality” (1993: 84). According to this conception, we achieve consistency in moral judgment by subsuming particular cases under general principles.¹³⁸ Thus, to morally judge a particular case in a way that is consistent with one’s other moral judgments, and hence to judge rationally, is to base one’s moral judgment on an argument of the following form:

P₁ All actions with features a, b, c, ... are impermissible.

P₂ This action has features a, b c, ...

C Thus, this action is impermissible.

The subsumptive conception presupposes that there is a comprehensive set of true Case-Scope universal principles of duty because, without them, arguments of this form would not be formally valid.¹³⁹

There is broad agreement that Kant subscribes to the subsumptive conception, and the textual evidence does indeed suggest as much. For not only does he define reason as the capacity to “cognize the particular in the universal” (CPR A300), he also identifies the will with practical reason on the grounds that “*reason* is required for

¹³⁸ For a more detailed explanation of why this conception is tempting, see McDowell 1998: 57-65.

¹³⁹ The set has to be comprehensive in the sense that, for any true judgment about the deontic status of a given action, we must be able to construct such an argument. If the set was not comprehensive, then some such judgments could not be adopted on a rational basis.

the derivation of actions from laws” (G 4:412). In developing an alternative reading of *Universality*, we will have to be sensitive to the fact that this idea of deriving actions from laws lies at the heart of Kant’s account of what it is to act for a reason. What we will not deal with, in the present chapter, is the specifically Korsgaardian concern that animates the standing-commitment argument. This concern will be addressed in ch. 5.

What about the willing-is-conceptual thread of the argument via premise 3b? In what sense of “universally valid”, if any, does the claim that willing is a conceptual state or activity presuppose that the way in which features determine the deontic status of actions can be captured in universally valid principles? In other words, which brand of Kantian generalism does this claim support? This is the question that we will seek to answer in sect. 2. And although we will approach this question from within Kant’s texts rather than Korsgaard’s, we will discover some parallels between Kant’s reflections on concepts and universality, on the one hand, and the willing-is-conceptual thread of Korsgaard’s anti-particularist argument, on the other (especially, if we see Korsgaard’s talk of describability there as anticipating her arguments for the publicity of reasons that is inspired by Wittgenstein’s arguments for the publicity of meaning (Korsgaard 2009: 196-206 and 1996b: 139-40, 144)).

2. The Agent-Scope Reading of *Universality*

In the present section, following the methodological approach outlined in sect. 2 of the introduction, I will try to uncover Kant’s most fundamental reasons for thinking that, in order to acquire objective knowledge of a given domain, we need to grasp the goings-on in that domain through concepts, serving as universal rules. For once we know which *role* universal rules play in Kant’s account of cognition, we will be in a better position to decide which *form* such rules must exhibit in order to play this role – i.e. in which sense they must be universal.

I will begin by tracing one of Kant’s arguments in the Transcendental Deduction (sect. 2.1) in the hope to clarify the role of concepts, qua universal rules, in cognition. Then I will show that a rule can play this role, and therefore count as strictly universal in the relevant sense, even if it does not hold in all circumstances, but only in a certain range of them, i.e., even if it is defeasible and admits of certain exceptions (sect. 2.2). Next I will exploit similarities between Kant’s view and inferentialism in

the theory of meaning in order to explain how it is possible to subsume particulars under defeasible rules (sect. 2.3). From there I will move on to the moral domain and draw on these insights about defeasible rules in order to make room for exceptions to Kant's principles of duty in a way that is sensitive to the aim of accommodating the circumstance-dependence of our moral obligations (sect. 2.4). I will conclude by proposing a reading of *Universality* that reflects these insights and show how this reading does away with the *Trilemma* (sect. 2.5).

2.1 Why Do We Need Concepts-Qua-Universal-Rules?

In the Transcendental Deduction, Kant sets out to establish that the categories¹⁴⁰ are objectively valid by showing that their objective validity is the condition of the possibility of experiencing objects, and therefore also the condition of the possibility of the objects of experience themselves (CPR A84-130/B116-169). According to one line of argument in the Deduction,¹⁴¹ the manifold of sense, as an aggregate of merely subjective modifications of the mind, could not amount to representations *of objects*, of entities existing independently of our awareness of them, if it was not for the fact that we combine the elements of the manifold in accordance with concepts, serving as universal rules. Here is a trenchant formulation of the basic idea from the Second Analogy.¹⁴²

How do we come to posit an object for these representations, or ascribe to their subjective reality, as modifications [of the mind], some sort of objective reality?... If we investigate what new characteristic is given to our representations by the *relation to an object*, and what is the dignity that they thereby receive, we find that it does nothing beyond making combinations of representations necessary in a certain way, and subjecting them to a rule. (CPR A197/B242)

¹⁴⁰ The categories are pure concepts of the understanding, such as the concept of a substance or the concept of a cause, which we cannot derive *from* experience but which we must draw on *in* experience, or so Kant argues. Kant's table of categories can be found in CPR A80/B106.

¹⁴¹ Since there are two versions of the Transcendental Deduction (A and B), each consisting of more than one line of argument, and since it is controversial how to best parse the text, the following sketch of the argument should not be seen as an attempt to do justice to the whole of Kant's argument in the Deduction.

¹⁴² Of course, this is not to say that the idea is not also articulated in the Transcendental Deduction itself. Here is an example: "We find ... that our thought of the relation of all cognition to its object carries something of necessity with it, since namely the latter is regarded as that which is opposed to our cognitions being determined at pleasure or arbitrarily ..., since insofar as they are to relate to an object our cognitions must also necessarily agree with each other in relation to it, i.e., they must have that unity that constitutes the concept of an object" (Kant CPR A105).

Kant says that the same rules “that give[] unity to the mere synthesis of different representations *in an intuition*,” namely concepts which serve as universal rules, also “give[] unity to the different representations *in a judgment*” (CPR A79/B104-5). And it is through this unity, that they bestow “the dignity” of a “relation to an object” on our representations. Our guiding question is this: *in which sense* must rules be universal in order to bestow this dignity on our moral judgments, i.e. our judgments that a given action has a certain deontic status? We shall answer this question indirectly, by asking *why* they need to be universal in order to bestow this dignity on intuitions.

So let us focus on the passage where Kant answers the latter question: a section from the A-Deduction entitled “On the Synthesis of Recognition in the Concept” (CPR A103-10). I want to offer a brief reconstruction of Kant’s argument in the first part of this section, along with supporting textual evidence. However, before we can delve into the main argument, we need to review some background assumptions carried over from foregoing sections. Earlier, in the beginning of the “Transition to the Transcendental Deduction of the Categories”, Kant observes that the kind of non-accidental correspondence or coming-together of representation and object that make for an objectively valid representation – or, as we might say: knowledge – is possible only in two cases: if either the object grounds the possibility of the representation or the representation grounds the possibility of the object (CPR A92/ B124-5). Now, in the present section, he reminds us why, with regard to how the manifold of sense (as the matter of intuition) is combined (or formed), it is the *latter* grounding relation that explains any non-accidental correspondence, not the former.

We have said above that appearances themselves are nothing but sensible representations, which must not be regarded in themselves, in the same way, as objects (outside the power of representation) ... outside of our cognition we have nothing that we could set over against this cognition ... we have to do only with the manifold. (Kant CPR A104-5)

By the time he reaches the Transcendental Deduction, Kant has established that empirical objects are appearances, and that sensations are the only thing that we ever receive. So while, in Kant’s view, it is *right* to say that, in the good case, the combination of the manifold of sense in intuition corresponds to the combination of the parts or aspects of the object represented, and that it does so non-accidentally, it would be *wrong* to say that the former combination is grounded in or is the way it is

because of the latter. In making this kind of statement, we would betray a longing to “step outside” of our experience. For this reason, his argument in the Transcendental Deduction presupposes the following assumption:

Background Assumption: With regard to the form or the manner of combination of the manifold, a non-accidental correspondence between object and representation is possible because the representation grounds the possibility of the object, not *vice versa*.

Of course, this doesn't mean that we can combine the elements of the manifold of sense in any way we like. In fact, it means the exact opposite, for to say that the very possibility of the object of experience hinges on something about how we combine the manifold of sense is to say that the objective validity of a given combination – its correctness, its being what it ought to be – is determined by a criterion that lies *within the bounds of experience*. According to Kant, this formal criterion of correctness is the *agreement* of the resulting representation with other representations of the object or, in other words, *unity* among our representations of an object.

We find ... that our thought of the relation of all cognition to its object carries something of necessity with it, since ... the latter is regarded as that which is opposed to our cognitions being determined at pleasure or arbitrarily ... insofar as they are to relate to an object our cognitions must also necessarily agree with each other in relation to it, i.e., have that unity that constitutes the concept of an object. (Kant CPR A104-5)

When Kant says that the very thought of an object “carries something of necessity with it”, he means that the thought of something that is independent of us is the thought of something that we can get right or wrong, of something that we *ought to* represent in certain ways and not others.¹⁴³ But how do we know whether we are getting it right, if we cannot step outside of our experience to compare our representations to the objects as they are in themselves? Kant's suggestion, to put it crudely, is that we know that something has gone awry when our various representations of

¹⁴³ This reading is certainly not uncontroversial. Kant could also be thinking of a more specific kind of necessity, namely *causal* necessity, and not just of the generic kind of necessity that we find wherever there is normativity. My reading has an important advantage, however. It brings out that being related to an object or being objective is something that *all* cognitive judgments, including *practical* judgments, have in common. The way in which this more generic notion of necessity figures in Kant's account of cognitive and aesthetic judgments has been explored in the work Hannah Ginsborg (see e.g. her 2006 and 2011).

an object do not fit together, for example, when an object appears to us as being a certain way even though it cannot possibly be that way, given what we know about objects of this kind. If an apple tree that I saw a moment ago suddenly disappeared, or changed its colour or its size, or moved whenever I move, then I would know that something is wrong: that I am misrepresenting it in some way, perhaps imagining it, or hallucinating it or being tricked.¹⁴⁴

Premise 1: An intuition of an object could not correspond to that object non-accidentally (i.e., be objectively valid or correct), if the manifold of sense in it was not combined in such a way that the intuition can agree or disagree with other representations of the object.

On this basis, Kant goes on to establish the need for concepts, for, in his view, there could be no such agreement if the manifold of sense was not combined in accordance with concepts, serving as universal rules.¹⁴⁵ Consider another experiential event that would make me suspicious:

Apple Tree: I approach an apple tree to touch its trunk, only to find that it gives in, allowing my hand to go right through to its core.

In which sense, if any, does the representation of this part of the object as penetrable fail to agree with my representation of the whole object as an apple tree? It fails to agree with the latter representation in the sense that it defies my expectations, given what I know about apple trees *in general*.

Recall that intuitions, as such, are representations of particulars: of *this* or *that* object (CPR B377). Kant realized that, if an intuition, as the representation of a particular, did not contain anything that, as it were, “reaches beyond” the particular item intuited, it could neither agree nor disagree with other representations of that item.¹⁴⁶ So in order for my representation of the relevant part of the object as penetrable to conflict with my representation of the whole object as an apple tree, I have to know

¹⁴⁴ James Van Cleve articulates this idea in the following way: “If a red, round apple lies on the table before me, the course of my experience will not be a matter of ‘anything goes’; rather, experiences of grasping and turning will be followed by further experiences of redness and roundness, experiences of lifting and biting by experiences of crispness and tartness, and so on” (1999: 92).

¹⁴⁵ Kant uses the term “representation” very broadly. His discussion of judgments, concepts, and intuitions and their relation to one another can be understood as an attempt to fill in the details: to say how our representations must be structured to afford objective knowledge.

¹⁴⁶ Now we can see that the argument under consideration is a more general version of the argument contained in the willing-is-conceptual thread of Korsgaard’s anti-particularist argument (see sect. 1).

something general about how things stand with apple trees under circumstances of this kind. I have to know a *rule*, e.g. that, unless someone or something has tampered with it, an apple tree has a completely impenetrable trunk. This rule is supplied by the concept. It is by seeing the object in front of me *as an apple tree*, by interpreting and combining the patches of green and brown, the rustling sound, etc., in accordance with that concept, that I form the expectation that I would feel something coarse and resistant if I were to touch a certain part of it. If things turn out otherwise, then my actual sensations will defy my expectations, and I will have reason to suspect that something has gone awry and to abstain from judgment.

We have seen why there could be no agreement or disagreement without rules. But why do the rules have to be strictly universal? After all, it is perfectly common to form expectations on the basis of mere generalizations or observational patterns, where the possibility of exceptions is left open. Nevertheless, Kant clearly insists on the strict universality of the rules in question or, as he also puts it, on the *necessity* of the connections that these rules assert. He says, for example:

The concept of body serves as the rule for our cognition of outer appearances by means of the unity of the manifold that is thought through it. However, it can be a rule of intuitions only if it represents the necessary reproduction of the manifold of given intuitions ... Thus in the case of the perception of something outside of us the concept of body makes necessary the representation of extension, and with it that of impenetrability, of shape, etc. (Kant CPR A106)

It is true, of course, that, even without any strictly universal rules, our manner of combining the manifold of sense would not have to be entirely random: we could pick up on patterns in the succession of sensations and form habits of association.¹⁴⁷ What is noteworthy, however, is that no distortion of such a pattern could ever amount to a real disagreement or contradiction, which is what we are trying to ensure is possible. If we did not regard the rules that constitute our concepts of empirical objects as strictly universal, as laws of nature that are valid for all actual and possible objects and events,¹⁴⁸ if, instead, we viewed them as something that holds true most

¹⁴⁷ This is how David Hume would explain the unity or coherence of our experiences: by appeal to customs or habits of the mind (2007: 5.1.5-6).

¹⁴⁸ When Kant says that a rule is “strictly universal”, he means that it is not just contingently universal but rather necessarily universal. This is the kind of universality that is characteristic of laws. We will come back to this distinction in ch. 5, sect. 1.

of the time, we could not say that certain ways of representing an object are *ruled out* by representing it as falling under a certain concept. So in order to play the role that Kant assigns to them, concepts have to do more than merely generalize over our past experiences. They have to *set standards* that determine what *has to be* true of objects falling under them in circumstances of a certain kind, no matter how we happen to experience these objects.¹⁴⁹ This brings us to the second premise of Kant's argument.

Premise 2: An intuition of an object could neither agree nor disagree with other representations of that object if the manifold of sense in it was not combined in accordance with a concept serving as a universal rule.

From these two premises, Kant can infer the conclusion that he is after:

Conclusion: An intuition of an object could not correspond to that object non-accidentally (i.e. be objectively valid or correct), if the manifold of sense in it was not combined in accordance with a concept serving as a universal rule.

Relatedly, an intuition could not have objective purport (i.e. be *of an object*), if, in combining the manifold of sense in it, we were not subject to a universal rule supplied by a concept.

For our present purposes, the main point of the argument is this. The significance of the criterion of agreement and, by extension, of concepts qua universal rules, lies in the fact that they ensure that “our cognitions [are not] determined at pleasure or arbitrarily” (CPR A104). We have to rely on universal rules because they alone make it possible, in principle, to distinguish *from within* between combinations of the manifold of sense that are thoroughly objective and those that reflect an undue influence of merely subjective factors, e.g. the perspective of the subject or their cognitive limitations. Accordingly, a combination is objectively valid insofar as it accords with the relevant rule, and subjectively distorted insofar as it does not. This is why Kant says that, in order to acquire knowledge of an objective world, we must represent objects in this world as governed by universal laws (CPR A114, A125, A127, B160-1, B163-5). Having explored Kant's argument for this conclusion, we can now focus on what he means when he says that a rule is universally valid.

¹⁴⁹ Note that, at this point, I do not mean to commit myself to any claim about the content or internal structure of such rules or any claim about the way in which they interact with one another. Speaking of strict universality, I only mean to indicate that these rules must differ from habits of association in a way that secures the possibility of agreement and disagreement.

2.2 Two Kinds of Counterfactual Robustness

In discussing the *Apple Tree* example, I said that my various representations of the apple tree could neither agree nor disagree with one another if they contained no general knowledge of what to expect of an apple tree *under these kinds of circumstances*. With this addendum, I was hinting at a point that it is now time to make explicit, namely that the significance accorded to universal validity in Kant's argument lends no support to the claim that the rules that we rely on in synthesising the manifold of sense in intuition must be robust across variations in circumstances insofar as these affect how things stand *with the object*. We can add some detail to the *Apple Tree* case to illustrate the point.

Imagine, once again, that, contrary to my expectations, the tree trunk gives in when I touch it. I could come up with various different explanations. Some of these explanations will disqualify (a subset of) my perceptions as grounds for perceptual judgment: perhaps there is actually no apple tree, so that my haptic perception of penetrability is veridical, whereas my visual perception of the tree is hallucinatory; or perhaps there is a tree, so that my visual perception of the tree is veridical, but my haptic perception of its trunk as penetrable is illusory; or perhaps it is dark and the object that I mistook for an apple tree is in fact a piece of clothing on a clothes line. However, there might be other explanations, explanations that do not disqualify *any* of my perceptions as grounds for perceptual judgment. After all, there are certain circumstances in which trunks of apple trees *do* have penetrable parts – for example, when woodpeckers have bored holes in them. In this case, all my representations of the object, taken individually, would be correct.

Given what we have said so far, this is a somewhat puzzling possibility. On the one hand, my perceptual expectations seem to have been generated on the basis of the rule “If x is an apple tree, x has a completely impenetrable trunk”, so the fact that my actual sensations fail to agree with the sensations that this rule makes necessary suggests that some of my representations of the object are incorrect or subjectively distorted. On the other hand, all my representations of the apple tree seem correct or objectively valid. If we want to make sense of the veridicality of my representations, while staying within the Kantian framework that we sketched above, we have three options. Either we argue 1) that the rule that I purportedly relied on, the one cited above, is not truly universal and therefore flawed, or 2) that my expectations were

actually generated by a different rule, or 3) that a rule can be universal, in the sense that matters to Kant, despite allowing for certain exceptions. For reasons that are similar to those that led us to reject Korsgaard's appeal to provisional universality (ch. 2, sect. 2.2) and those that we cited in support of the *Asymmetry* desideratum (ch. 2, sect. 1), I favour the third account.

Regarding the first account, it is worth noting that the mere fact that my actual sensations fly in the face of my initial expectations does not show that I did not know that this sort of thing can happen. Initially, I may not have been fully aware of the circumstances, not aware that the tree had become the nesting site of a woodpecker, but upon noticing the hole, I might say to myself: "Oh, wow, a woodpecker nest!". If that was my reaction, it would seem inappropriate to say that my perceptual experience was guided by a flawed concept of an apple tree, that it relied on a rule that fell short of the universality needed to establish unity among my representations, or that it was based on a misconception of the laws of nature that govern apple trees in their interaction with other things and beings. The reason why I had the wrong expectations is *not* that my perception of the tree was guided by a flawed concept, but that I was not aware of all the relevant facts, that I was not aware of the applicability of additional concepts.¹⁵⁰ In order to appreciate the difference between these two explanations, compare this version of the *Apple Tree* example to an example where the experiencing subject really *is* relying on a rule that is not properly universal: someone who has never seen or heard of black swans, for instance, and, therefore, assuming that all swans are white, misperceives a black swan as an odd-looking black cormorant. Here the perceptual experience is subjectively distorted. The lesson to be learned from this comparison is that it is one thing to think of the goings-on of a certain kind of object as sensitive to the presence of other objects and states of affairs, and another thing to be wrong about what those goings-on are in the first

¹⁵⁰ This example illustrates that Kant commentators are wrong when they say *either* that concepts-qua-universal-rules are to be understood as criteria for classifying experiences as veridical or non-veridical and that Kant was confused when he thought that, for objective knowledge, these rules had to play a role in synthesising the manifold (Bennett 1966: 61-2; Strawson 1966: 88-9); *or* when they maintain that these rules are imperatives that we ought to follow in synthesising the manifold, and that our experiences are veridical to the extent that we do (Beck 1998: 103; for this distinction see, i.a., Van Cleve 1999: 94). The truth lies in the middle: the veridicality of our experiences depends on whether we synthesise the manifold of sense such that our experiences *can be* unified in accordance with concepts-qua-universal-rules, no matter whether we manage to thus unify them on any given occasion, e.g. whether or not we realize that we have found a woodpecker nest.

place. The woodpecker version of the *Apple Tree* case falls into the former category, so the rule relied on is universal.¹⁵¹

Having ruled out the first account, we can now compare the second and third. Proponents of both accounts accept that the rule used to generate my perceptual expectations is universal, both trace the lack of agreement between expected and actual sensations to my initial ignorance of the nesting activity, and both agree that my appreciation of the causal impact of the woodpecker on the apple tree restores the agreement between my representations, thus rendering them fully intelligible. The point of disagreement between them is only whether the fact that the rule is universal entails that the knowledge required to grasp the goings-on as a causal interaction between woodpecker and apple tree is encoded *in the very rule* relied on in combining the manifold of sense in the intuition of an apple tree, or whether it doesn't, whether that knowledge could instead be knowledge of how this rule operates in different circumstances, e.g. knowledge of when it is valid and when it is not. According to the second account, it does. On this account, if my expectations were generated on the basis of a universal rule, the rule would have to read: "If x is an apple tree and x has not become the nesting site of a woodpecker, then x has a completely impenetrable trunk".¹⁵² According to the third account, it does not. On this account, a rule can be universal, in the sense relevant to Kant, even if its validity is circumstance-dependent. The question that we have now arrived at is the question from which we began, namely: can a rule be universal, in the relevant sense, if it does not hold in all cases or in all circumstances, if it is a defeasible rule that admits of exceptions? (This, recall, is what we need in order to reconcile our three desiderata. If we were to build all potential defeaters into the rule itself, as the second account has it, we would violate *Asymmetry*. So if both accounts are viable, the third is to be preferred).¹⁵³

¹⁵¹ This is the lesson that we accused Korsgaard of ignoring. The *reductio argument* against her *Two Standpoint View* (ch. 2, sect. 2.2) brought out that we should not treat the acknowledgment that there are defeaters as mistakes.

¹⁵² Of course, since the whole point of this move is to be able to insist that I relied on an appropriate concept or fully universal rule, the list of exceptional cases that would have to be ruled out would be much longer.

¹⁵³ My aim here is limited. It is merely to show that, given the role that universality plays in Kant's account, he *can* say that a rule is universal despite admitting of exceptions, not that he *should* say this. If I wanted to argue that he should favour this view over the second account, I would adduce the kinds of considerations that I presented in ch. 2, sect. 1 to motivate the *Asymmetry* desideratum. I would note, for example, that proponents of the theoretical counterpart of the Fine Print Reading would have us believe that it is part of mastering the rule that first enables us to perceive trees that we know about the nesting practices of a certain bird species, and that, if we did not possess this understanding,

Now that we know why universality matters to Kant, we can reformulate the question as follows: could we rely on the rules that constitute our concepts of objects to distinguish, from *within* experience, between objectively valid and subjectively distorted ways of synthesising the manifold of sense, if these rules did not hold in all circumstances? The answer to this question, I would argue, is yes. For the possibility of agreement and disagreement between my different representations of an object is secured by the standards set by all the rules that make up my understanding of the objective world *together*, not by any one rule on its own. So as long as an exception to a given rule can be justified from within the relevant system of rules as a whole, it is not the kind of exception that, if deemed legitimate, would erode the distinction between objectively valid and subjectively distorted ways of combining the manifold of sense and, as such, undermine the very possibility of agreement and disagreement.¹⁵⁴ (Of course, this claim raises questions about the logical form of these defeasible rules and about how they restrict each other; we will address these questions in a moment).

Given the cognitive role that Kant attributes to universal rules, it is thus unproblematic to say that the rule “If x is an apple tree, x has a completely impenetrable trunk” is universal, even though the woodpecker version of the *Apple Tree* example constitutes an exception to that rule; or that the concept of an apple tree, on its own, makes necessary the representation of a completely impenetrable trunk but that, in combination with the concept of a nesting woodpecker, it does not; or, again, that the operation of the laws that govern apple trees depends on circumstantial features insofar as these are deemed relevant by the system of laws that govern nature as a whole.¹⁵⁵ What the operation of these laws cannot and does not depend on are circumstantial features which *appear to be* relevant to the subject because of a distorting factor that bears on how they combine the manifold of sense, but which, according to the system in which these laws are embedded, are not. Think of someone who has taken a hallucinatory drug that makes them feel as if they can walk through walls. If it

this knowledge gap would prove that none of our perceptions of trees are ever veridical – which is a rather implausible thing to say.

¹⁵⁴ In this respect, it differs from the kinds of exceptions that would be deemed legitimate within a framework like Hume’s, in which the combination of the manifold of sense is thought to be based on mere habits of association. See fn. 147.

¹⁵⁵ This point is worth stressing: the above rule is *strictly universal* despite being defeasible and, unlike a Humean generalization (see fn. 147), it qualifies as a *proper law*, although it is a *ceteris paribus* law (see ch. 2, sect. 2.3.2). It is strictly universal or a proper law in the sense that, whichever conditions must obtain for the relevant link to obtain, whenever they do obtain, the link obtains as well.

was the drug intake that explained why they found themselves having the kinds of sensations that would otherwise indicate that the object that they are touching is penetrable, then their perception of penetrability would be non-veridical. And any attempt on their part to restore agreement between their representations by appeal to circumstantial features that are taken to bear on the object, e.g. the “fact” that it is witching hour, would result in a false judgment.¹⁵⁶ We can capture the sense in which conceptual rules have to be universal more precisely by distinguishing between two kinds of counterfactual robustness. A universal rule, such as “If x is an apple tree, x has a completely impenetrable trunk”, needn’t be robust across variations in *objectively relevant circumstances* (i.e. circumstances that bear on how things go with the object that is represented). But it must be robust across both actual and possible variations in merely *subjectively relevant circumstances* (i.e. circumstances that merely appear to bear on how things go with the object because of some distorting factor that bears on the subject that is doing the representing).¹⁵⁷ This means that universal rules can be *defeasible* in the sense that there can be objectively relevant features of the circumstances that function as defeaters.¹⁵⁸ In the next section, I will elaborate on this conception of defeasibility and examine whether rules that are, in this sense, defeasible can do the kind of work that motivates proponents of the subsumptive conception of rationality to posit Case-Scope universal rules. This is important if we want to make sense of Kant’s claim that we derive actions from laws.

¹⁵⁶ What they could do, however, is restore agreement between their representations by appeal to the distorting factor and its impact on their capacities: they could realize that they are intoxicated and on this basis discount the perception of penetrability as not indicative of the character of the tree.

¹⁵⁷ In bad cases the third-personal explanation of why a subject judged the way they did differs from the explanation or justification that the subject themselves would offer if challenged. The third-personal explanation would appeal to the relevant distorting factor, e.g. the drug intake, whereas the first-personal explanation or justification would appeal to the purportedly objective but in fact merely subjective circumstantial feature, e.g. the “fact” that it is witching hour. In good cases, by contrast, the two explanations coincide.

¹⁵⁸ It seems to me that Stephen Engstrom is expressing the same idea when he says: “A judgment has objective universal validity just if ... the act of predication ... is valid for all objects falling under its concept ... Such objective universality does not, however, imply that the cognition of each object falling under the judgment’s concept would actually agree with the cognition of every other in asserting the judgment’s predicate of its object. Such actual agreement cannot be assumed in the case of contingent judgments, such as those of experience ... In the case of such judgments, objective universal validity implies only that all the objects that can be brought under the judgment’s concept would have its predicate asserted of them in the cognition of them *provided that they were in such conditions as is the object of the judgment*. The judgment that the water in the pond is frozen, for example, in relying on the concept water, implicitly involves the universal judgment that any bit of water, *when in the conditions of the water in the pond*, must be frozen – a judgment in which such conditions, though unspecified in the judgment itself, are implicitly regarded as sufficient to determine the water to be frozen” (2009: 116, my emphasis). He does not explore the implications of this idea any further, however.

2.3 Defeasible Rules and Subsumption

If the rules that our concepts supply are defeasible in ways that are not encoded or anticipated in the rules themselves, then it seems that consistency in judgment – or, in Kant’s terms, agreement or unity – cannot be secured by subsuming particular cases under such rules. Consider how we might try to secure consistency in the *Apple Tree* case.

P₁ If x is an apple tree, x has a completely impenetrable trunk.

P₂ This is an apple tree.

C This has a completely impenetrable trunk.

If the rule in P₁ is defeasible, if it holds in some but not all cases, then the inference is formally invalid: the conclusion does not follow from the premises.¹⁵⁹ So from the point of view of those who accept the subsumptive conception of rationality, those who think that there is no consistency without formally valid inferences, our account of conceptual rules as defeasible must look problematic. But, as we will see, the subsumptive conception is not without alternatives. It is not the only way to make sense of Kant’s idea that we “cognize the particular in the universal through concepts” in order to secure “the highest unity of thinking” (CPR A299-300/B355-357).

Note that the subsumptive conception of rationality rests on an assumption. It assumes that, in saying “This is an apple tree”, I am not committing myself to any claim about the penetrability of its trunk. For if I did, if, in thinking “This is an apple tree”, I already thought “This has an impenetrable trunk”, then I would not need P₁ to derive C from P₂. So, according to the subsumptive conception, P₂ (even when combined with the claim that the circumstances are objectively suitable) does not *entail* C. Instead, whether C is true is *independent of* whether P₂ is true, each being a matter of whether they correctly represent the properties of the object referred to. Knowing the rule in P₁ might make it easier to figure out whether the trunk-part of the object

¹⁵⁹ At this point, we might be tempted to add a hedging clause and restate the rule such that it *does* hold in all cases after all, e.g. as follows: “*In objectively suitable circumstances*, if x is an apple tree, x has a completely impenetrable trunk”. In that case, one would need a further premise to reach a determinate conclusion about whether one’s judgment about a particular (“This has a completely impenetrable trunk”) is consistent with one’s other judgments. One would need the premise that the circumstances at hand are objectively suitable. However, the rule itself provides no criterion on the basis of which one could decide whether a set of circumstances is objectively suitable or not. So this hedging strategy does not seem to get us closer to our goal.

in front of me is penetrable if I already know that it is an apple tree and cannot be bothered to touch it, or it might help me to predict what I will feel if I touch it, but none of this is required for either of the judgments to be true or objectively valid.

This assumption is in direct opposition to Kant's view, as outlined in sect. 2.1 above. Recall that, according to Kant, we can only make objectively valid judgments, judgments that *refer to* and are *true of* objects, because we intuit and cognize these objects through concepts, such as the concept of an apple tree, which, *all by itself*, serves as a rule, allowing me to *infer* what else can, cannot or must be true of an object that falls under it, given the circumstances in which it is embedded.¹⁶⁰ This view, the view that inference comes before reference, is a view that Kant shares with so-called inferentialists in the theory of meaning.¹⁶¹ For this reason, it should not come as a surprise that it is in *their* writings that we find an account of how defeasible rules can figure in valid inferences that secure consistency in judgment.

According to inferentialists, we have to extend the inferential role account of meaning, which is usually taken to cover logical terms only, to non-logical terms, i.e., to terms whose meaning is usually understood in referential terms (see e.g. Peregrin 2014: 25). Consider the material conditional as an example. There is broad agreement that the meaning of the material conditional can be articulated by saying that "If x is

¹⁶⁰ Kant makes this point over and over again. He says, for example, that an "object is no more than that something, the concept of which expresses such a necessity of synthesis" (CPR A106) and that "an *object* is that in the concept of which the manifold of a given intuition is united" (CPR B137). Some commentators have read these remarks as indicating that Kant is a *phenomenalist*. James Van Cleve, for example, moves from the claim that "Kant defines an object as something *the concept of which* plays a certain role" – a claim I agree with and highlight above – to the claim that Kant endorses a "reductive account of objects". On this account, "to say that there is an apple before me is *equivalent* to saying that I am having certain sorts of experiences (intuitions), and that if I (or other observers similarly placed) were to perform certain actions, they would have further experiences of predictable sorts" (1999: 92-3). What Van Cleve is missing here is the normativity that concepts bring into the picture. To apply a concept, e.g. to say that the object before me is an apple, is not to *predict* that, if I perform certain actions, I *will* have certain experiences, but to say that, if I perform these actions, I *ought to have* these experiences, and that, if I don't, something must have gone wrong. As I said earlier, this is precisely how concepts make it possible for us to have experiences of objects.

¹⁶¹ The view that Kant is an inferentialist has recently been defended by David Landy (2015). Landy argues that this reading allows us to make sense of Kant's case against Hume. In his view, Kant raises two objections against Hume's theory of representation, namely, first, that Hume cannot account for the possibility of representing complex states of affairs as complex (Landy 2015: 62) and, second, that he cannot account for the unity of the proposition (2015: 91-2). Landy shows that, given these points of criticism, it is only by interpreting Kant as an inferentialist that we can see how his own theory of representation solves the problems that he finds in Hume. On this reading, Kant thinks of concepts as rules or functions that place intuitions in inferential relations, thus locating them in an inferentially structured system of cognitions (Landy 2015: 64-80, 92-101). Landy's reading is inspired by Wilfrid Sellars' reading of Kant and his talk of counterpart relations (see Sellars 1968: 26-30, 63-67). For other inferentialist readings of Kant, see Rosenberg 2005: 91-4 and O'Shea 2012: 128-32.

F, then x is G” is false, unless whenever “x is F” is true “x is G” is true as well. But many philosophers would argue that this is not how we should articulate the meaning of terms such as “apple tree”. Instead, they would suggest articulating the meaning of “apple tree” by citing the conditions that have to obtain in order for the claim “x is an apple tree” to be true. Here inferentialists disagree. In their view, we can articulate the meaning of “apple tree” by saying things like the following: in objectively suitable circumstances, “x is an apple tree” is false unless “x has a completely impenetrable trunk” is true.¹⁶² To say that the meaning of *both* logical *and* non-logical terms consists in their inferential role is to say that inference (2) below is just as self-contained as inference (1), that it is not an enthymeme, and that we should not assume that there is some additional premise that has been omitted (see Sellars 1953, Brandom 1994: 95-105, 2000: 52-4, Peregrin 2014: 27-9).¹⁶³

(1) If x is F, then x is G.

x is F.

Therefore, x is G.

(2) x is an apple tree.

Therefore, x has a completely impenetrable trunk.

If we think that (1) is valid as it stands, then this is only because we take the meaning of “if ... then ...” to be fixed. By the same token, we can treat the meaning of “apple tree” as fixed and regard (2) as valid and self-contained. Following Wilfrid Sellars (1953), inferentialists call the rule underlying (1) a *formal* rule of inference and the rule underlying (2) a *material* rule of inference. We need to highlight two implications of this broader conception of rules of inference.

The first concerns the issue that brought us here: the nexus between being an apple tree and having a completely impenetrable trunk is *defeasible*, so if we call the rule governing the inference in (2) a rule *of inference*,¹⁶⁴ then we have to abandon the idea that rules of inference are generally indefeasible. (However, as inferentialists are keen to emphasize, we might have good reasons to abandon this idea anyway, even

¹⁶² This, of course, is just a different way of saying what Kant says: that concepts serve as rules.

¹⁶³ In fact, according to inferentialists, if we thought that (2) relied on an implicit premise (as stated by Bertrand Russell in his 1914: 66), we would ultimately face the same infinite regress that Lewis Carroll (1895) cautioned against when reflecting on formal inferences such as (1).

¹⁶⁴ The alternative would be to call it a (suppressed) major premise.

when it comes to purportedly bulletproof rules of inference such as the rule in (1), i.e. modus ponens (McGee 1985)). Relatedly, we must abandon the idea that good inferences, i.e. inferences by means of which we can “cognize the particular in the universal” and thus secure consistency in judgment, have to be indefeasible, that they must survive the addition of any further premise whatsoever. I believe that we *can* abandon this idea. Recall what it takes for a judgment about a given particular to be consistent (in agreement, in unity) with our other judgments: it must fit in with what we know about objects of this kind. Note, further, that understanding *which* premises render a material inference (e.g. (2)) invalid is part of understanding the inferential role of the relevant term (e.g. “apple tree”). Thus, for an inferentialist, articulating the meaning of (at least) ordinary, non-logical terms, such as the term “apple tree”, is a matter of “construct[ing] ... inferential hierarchies with oscillating conclusions like [the following]” (Brandom 2000: 88).

- (2) x is an apple tree. Therefore, x has a completely impenetrable trunk.
- (3) x is an apple tree. And x has become the nesting site of a woodpecker. Therefore, x does *not* have a completely impenetrable trunk.
- (4) x is an apple tree. And x has become the nesting site of a woodpecker. And x has been patched up by a gardener. Therefore, x has a completely impenetrable trunk.

Etc.

This hierarchy expresses some of what we know about apple trees, namely what we know about the penetrability of their trunks, and therefore it is only in light of this hierarchy *as a whole* that a judgment about the penetrability of a particular apple tree trunk can be assessed for consistency.

With this in mind, we can restate a point from sect. 2.2 above in inferentialist terms. There we observed that the woodpecker exception is an exception that does not erode the distinction between objective and merely subjective combinations of the manifold of sense, because it is an exception that is intelligible from within the relevant system of rules as a whole. Now we can give the same explanation of why this exception is innocuous, albeit in slightly different terms: it is innocuous because the rule that “having become the nesting site of a woodpecker” is a defeater is part of

our shared understanding of the term “apple tree”, part of the meaning of this term. As such, it binds *anyone* who uses it. It is, in other words, not up to the individual user to treat this feature as a defeater. They *ought to* treat it as such, no matter what subjective condition they are in (no matter if they have taken drugs or if they are having hallucinations, etc.).

The second implication of the view that the inferential role account of meaning should be extended to non-logical terms is that there is no sharp distinction between analytic and synthetic truths, at least not in the contemporary sense. In the context of an interpretation of *Kant*, this might seem particularly problematic, given that he is the one who first introduced the distinction. The attempt to examine this implication in full detail would take us far beyond the scope of this thesis, but let me indicate how one could go about overcoming this concern. To begin with, we should note that, for Kant, the difference between analytic and synthetic judgments is not that the former are made true by *language*, while the latter are made true by *the world*. Instead, he defines analytic judgments as those where “the predicate B is (covertly) contained in [the subject] concept A”, and synthetic judgments as those where “B lies entirely outside ... A, though to be sure it stands in connection with it” (CPR A6). This way of understanding the analytic-synthetic distinction leaves open the possibility that the manifold of sense (the “world-element”) enters into cognition not only as material that we synthesise in accordance with ready-made concepts (as a “truth-maker”),¹⁶⁵ but also as something that shapes the formation of concepts (Sellars 1953: 336-7).¹⁶⁶ Accordingly, we can attribute a given nexus, e.g. the nexus between apple trees and completely impenetrable trunks, to the world, without having to deny that this nexus has made its way into our language and become part of our shared understanding of a term.¹⁶⁷ The judgment expressing this nexus will be synthetic, in *Kant’s* sense, if it

¹⁶⁵ Strictly speaking, we should not use the term “truth-maker” in connection with Kant’s theory. This term is misleading because, for Kant, the manifold of sense, simply as such, has no objective purport at all. Prior to being synthesised, it cannot make anything true or false.

¹⁶⁶ In fact, this is exactly what Kant says when he discusses the formation of empirical concepts. In his view, it is by comparing, reflecting on, and abstracting from our intuitions of particulars that we form concepts of empirical objects (JL 9:94).

¹⁶⁷ For a detailed discussion of how it is possible that our empirical concepts are both derived from experience *and* presupposed in perceptual synthesis, see Newton 2015.

can only be established on the basis of a *ground* (or reason) that is not already contained in it, and analytic if it can be established without any such external ground.¹⁶⁸

In fact, what initially seemed like a major drawback of the inferentialist reading, namely the blurring of the line between language-based and world-based truths, will turn out to be one of its most important advantages. For once we realize that even binding semantic norms are open to challenges, and that, in fact, we often challenge each other's views about what to believe and how to live *by* challenging each other's ways of speaking, we can see how it is possible to maintain that the intersubjective intelligibility of a domain depends on its being law-governed without denying that there is open-endedness and disagreement around the edges.¹⁶⁹

2.4 The I That Judges and the I That Acts

It is now time to come back to the moral domain and carry over what we have learned so far. Recall our starting point: at the end of sect. 1, we set out on a lengthy excursion into Kant's theoretical philosophy because we were hoping that, if we re-think Kant's talk of universality in light of the idea that, like perception, willing is a *conceptual* state or activity, we might come closer to a more suitable conception of defeasibility that dissolves the *Trilemma*. We found that a conceptual rule can be defeasible in the sense that it can play its role in cognition *without* being robust across objectively relevant variations in circumstances, as determined by the system of rules as a whole or, more precisely, by the inferential hierarchy that articulates the relevant aspect of the concept's meaning. In the following, we will draw on these insights to explain how Kant's principles of duty can be universally valid *despite* being defeasible and admitting of certain kinds of exceptions. In order to do this, we need to conceive of principles of duty as hierarchically nested rules of inference that articulate the meanings of what we might call "deontic concepts": concepts that mark the action kinds they designate as impermissible or obligatory by default, e.g. the concept of a lie or the concept of a promise.¹⁷⁰ (Note that, in doing so, we pick up where we left

¹⁶⁸ For a more text-based discussion of how an inferentialist interpretation of Kant can accommodate his synthetic-analytic distinction, see Landy 2015: 68-9.

¹⁶⁹ It is by admitting that there is this open-endedness (which I highlighted in ch. 2, sect. 2.3.2, when motivating the *Asymmetry* desideratum) that we earn the right to blur the line between language-based and world-based truths *without* committing ourselves to the absurd view that mastering an empirical concept requires complete scientific knowledge of the relevant phenomenon.

¹⁷⁰ I should clarify two points. First, it is worth noting that we can formulate principles of duty that contain no deontically loaded terms. Here is an example: "Making a believed-false statement to another

off at the end of sect. 1: following the lead of Korsgaard’s premise 3b, we explicate what it means to think of willing as a conceptual state or activity). Furthermore, we need to assume that these deontic-concepts-qua-principles-of-duty enable us to acquire objective knowledge. The first question is this: which objects do deontic-concepts-qua-principles-of-duty enable us to acquire knowledge of?

Principles of duty tend to be formulated in the following way: “Lying is impermissible”, “Breaking promises is impermissible”, “Helping people in an emergency is obligatory”, etc. This suggests that deontic concepts enable us to acquire knowledge of *actions*, that, by bringing a particular action under a deontic concept, we acquire knowledge of its deontic status. If we were to adopt this view, then we could use the inferentialist framework outlined above to say all the things that we wanted to say: that these principles are defeasible, that they have to be robust across variations in circumstances that are merely subjectively relevant but not across variations that are objectively relevant, that they admit of certain exceptions, namely exceptions that are licensed by our shared understanding of the relevant deontic concept, etc. This is not *Kant’s* view, however, and it is a view that, I think, is in various ways inferior to Kant’s.¹⁷¹ Let me begin by citing some textual evidence – the more substantive reason for rejecting the idea that we acquire knowledge of *actions* will come into view when the details of my interpretation are in place.

When drawing the distinction between theoretical and practical knowledge, Kant characterizes the former as knowledge of “what is” and the latter as efficacious knowledge of “what ought to be” (CPR Bix-x, A633/B661, CPrR 46). That practical knowledge is supposed to be causally efficacious, that it is supposed to *make its object actual* (CPR Bx), means that it must be knowledge not merely of what ought to be but, more specifically, knowledge of *what I ought to do*. After all, if I did not take my

er person with the intention that the other person believe that statement to be true is impermissible”. (This is the traditional definition of lying (see Mahon 2008a: §1)). However, far from disproving our hypothesis that principles of duty articulate the meanings of deontic concepts, this merely goes to show that the inferential role of a concept can be articulated without using it. In the case of a deontic concept, this can be achieved by simply linking the conditions of its application *directly* to the deontic consequences of its application. On this point see Peregrin 2014: 21-5. Second, for the time being, we are interested in principles of duty, so we can leave aside the question of whether there are concepts that mark actions as by default *permissible*. For to say that an action is permissible is simply to say that there is neither a duty to perform it nor a duty to abstain from it. We will come back to the domain of the permissible in sect. 3.2.

¹⁷¹ This indicates that my response to the *Particularist Challenge* consists of two potentially independent parts. I hope that my discussion will make clear why both parts are important and why they should be combined.

judgment that things ought to be thus and so to stand in any relation to *me*, then that judgment would not play any role in making it the case that things are thus and so (Engstrom 2009: 118-21). Other aspects of Kant's discussion confirm the idea that, for him, judging practically (and thus: judging morally) is a matter of predicating *of oneself* what one ought or ought not to do. Think of how he goes about establishing the possibility of hypothetical and categorical imperatives. In both cases, he aims to show that we have the right to say *of certain agents* that they ought to do something. In the former case, this is meant to be easy: that an agent who wills a certain end ought to take the means to this end is thought to follow analytically from their willing the end (G 4:417-20). In the latter case, the task is more difficult: that any rational agent, simply as such, ought to do what morality demands is a synthetic a priori judgment to which Kant devotes a whole section of the *Groundwork* (G 4:446-63). This, again, suggests that practical judgments are judgments of the form "I ought to do x" and, thus, that the *deep structure* of principles of duty qua general moral judgments is: "I ought not to lie", "I ought not to break promises", "I ought to help people in an emergency", etc. (In sect. 3.2, we will see that moral judgments differ from other practical judgments in that they are judgments about me *simply as a rational being*).

If principles of duty enable us to acquire knowledge of *ourselves*, if, as subjects who judge, we are also the objects about whom we judge, then it might seem that, in the moral domain, there are no objectively relevant features of the circumstances that could serve as defeaters at all. This is because, in the theoretical domain, we were able to classify features of the circumstances as objectively relevant, i.e. as genuine defeaters, by asking ourselves whether the judging subject figured in the explanation of why their judgment about the allegedly exceptional case was what it was. In the bad *Apple Tree* case, for example, the fact that the relevant explanation made reference to the subject and her drug intake was a sure sign that the alleged defeater was a merely subjectively relevant feature. However, given the identity of subject and object in the moral domain, *any* explanation of why a judgment is what it is will make reference to the judging subject. Does this mean that principles of duty are infeasible after all?

No, it does not. We can see that it does not if we pay attention to the distinction between the subject *as the one who is judging* and the subject *as the one who is acting (or object)*, for then we can appreciate that it's only in the former capacity that they should

not figure in an explanation of why they judged the way they did. With this in mind, we can reformulate the distinction between two kinds of counterfactual robustness as follows: a principle of duty need not be robust across variations in *objectively relevant circumstances* (i.e. circumstances that bear on what the subject as the one acting ought to do), but must be robust across actual and possible variations in merely *subjectively relevant circumstances* (i.e. circumstances that merely appear to bear on what the subject as the one acting ought to do because of some distorting factor that bears on the subject as the one judging).

2.5 Universal and Defeasible: The Agent-Scope Reading of *Universality*

Now that we know what Kant's principles of duty, as articulations of the meanings of deontic concepts, enable us to acquire knowledge *of* – namely, what we, as rational agents, ought to do – we are in a position to understand in which sense they have to be universally valid and how their being so is compatible with their defeasibility. To begin with, let me illustrate in which sense they are defeasible. Afterwards, I will elaborate on the sense in which they are universally valid (which, of course, is not the Case-Scope sense) and show how this alternative understanding of universal validity allows us to dissolve the *Trilemma*.

Let us compare three inferences or derivations of actions from laws. Suppose that the first occurs in a standard situation, the second in a situation that differs from the standard situation in objectively relevant respects, and the third in a situation that differs from the standard situation only in subjectively relevant respects. The cases that I have chosen are hopefully intuitive, but, for now, nothing hinges on whether my account of these cases is correct. (An account of how to determine the content of Kantian principles of duty will be given in sect. 3.3 below). I should note, further, that the below example serves to illustrate only one of the two kinds of principles of duty that Kant discusses: principles that prescribe or prohibit the performance of actions, not principles that prescribe or prohibit the adoption of ends or maxims.¹⁷²

¹⁷² Roughly speaking, principles of duty that concern ends or maxims could be defeated by a broader range of features than principles of duty that concern action kinds. Among other features, the former could be defeated by facts about what else the agent has done or plans to do. For example, if one had to explain why one is not obligated to show one's gratitude to a particular benefactor on a particular occasion, one might have to refer to the fact that one has already used the money that was available to support one's parents. We discussed this example in ch. 1, sect. 3.2, and the distinction between two perspectives on our moral lives which bears on the issues just noted in ch. 1, sect. 4.1.

- 1) In saying x, I would be lying. I ought not to say x.
- 2) In saying x, I would be lying. And in saying x, I would be saving a life. I am allowed or obligated to say x.¹⁷³
- 3) *In saying x, I would be lying. And in saying x, I would be benefitting a friend. I am allowed or obligated to say x.

Given that all our practical judgments, including moral judgments, are judgments about what we ourselves ought to do, any feature that we take to be a defeater is a feature that we predicate *of ourselves*. As the one who is lying, I would also be the one who is saving a life (as in 2)) or the one who is benefitting a friend (as in 3)). This bears out what we observed above: that, in the moral domain, *any* explanation of why a judgment is what it is will make reference to the subject. The decisive question, as we noted, is whether it makes reference to him as the one who is judging. Arguably, in case 2), it does not. Arguably, it is part of our shared understanding of the rules that govern our practices of information exchange, that, *ceteris paribus*, having the chance to save a life makes it permissible or obligatory to lie. (In other words, it is part of the meaning of the concept of lying, which, qua meaning, is essentially public (Korsgaard 2009: 196-206, 1996a: 139-40, 144)). As such, this feature is objectively relevant: something that *any* judging subject could agree is relevant to what one ought to do in this sort of situation. Accordingly, there is nothing about *me as the one judging* or *my perspective* in particular that would have to be mentioned when explaining why I judged the way I did. This is not how things stand in case 3). Here it seems that my treating the opportunity to benefit my friend as a defeater is not licensed by our shared understanding of when it is and isn't legitimate to lie. As such, this feature is only subjectively relevant: something that *only I* (and others with a similarly biased perspective) could take to be relevant to whether I am allowed to lie. Anyone else would be unable to share in this biased judgment because part of the explanation of why I judged the way I did is that my fondness for my friend influenced *me as the one judging*. My judgment is, quite literally, an instance of my "cognitions being determined at pleasure" (CPR A104-5).

¹⁷³ As above, I am ignoring the question of whether and when the presence of a defeater makes a by default impermissible action permissible and whether and when it makes it obligatory.

The comparison of these two cases reveals that, corresponding to the distinction between two kinds of counterfactual robustness, there is a distinction between two kinds of exceptions. On the one hand, there are morally legitimate exceptions that are either called for or at least licensed by the circumstances of action, and, on the other hand, there are morally illegitimate exceptions, exceptions that we might be tempted to make “for ourselves (or just this once) to the advantage of our inclination” (G 4:424). It is because of our propensity to make exceptions of this *latter* kind and to treat ourselves as special that Kant feels the need to repeat, again and again, that principles of duty are strictly universal. With this in mind, we can appreciate that there are not just textual but also substantive grounds for regarding *agents*, and not *actions*, as the objects which principles of duty enable us to acquire knowledge of. For while treating an action as exceptional when it is not exceptional does not seem like a distinctively *moral* failure, treating oneself as exceptional when one is not exceptional most certainly does.

Now it is time to examine in which sense Kant’s principles of duty have to be universally valid in order to play their role in cognition. We said that they have to be robust across variations in circumstances that only appear to bear on what we, as agents, ought to do because of some distortion in how we, as judging subjects, look at the case. But, *pave* Case-Scopers, they do *not* have to be robust across variations in circumstances that really do bear on what we, as agents, ought to do. Thus, we have earned the right to abandon the Case-Scope Reading of *Universality* in favour of what I shall call the “Agent-Scope Reading”, which relativizes the validity of principles of duty to the circumstances of action. Here are both, for comparison.

*Universality*_(CS): The features in virtue of which actions have their deontic statuses can be captured in principles of duty that hold for all cases or in all circumstances.

*Universality*_(AS): The features in virtue of which actions have their deontic statuses can be captured in principles of duty that can be agreed to hold by all rational agents (qua subjects) and for all such agents (qua objects) insofar as they are in the same circumstances.¹⁷⁴

¹⁷⁴Jochen Bojanowski draws a similar distinction when he says that “the categories of quantity [in Kant’s table of the categories of freedom] quantify not over cases but over agents” (2015: 215-6). My distinction is indebted to his.

How does this re-interpretation of *Universality* bear on the *Trilemma*? According to the Agent-Scope Reading, a Kantian principle of duty (e.g. “I ought not to lie”) can be universally valid and thus binding in some situations (standard situations), but not in other situations (exceptional situations), which call for different principles (e.g. “I am allowed to/ought to lie when I can save a life”). As such, the Agent-Scope Reading accomplishes the feat of dissolving the *Trilemma*: it allows that the kinds of actions that Kantian principles of duty are usually taken to strictly prohibit (e.g. lying, promise-breaking) can be permissible or obligatory when and because defeaters are present (*Complexity*), but without implying that, when they are impermissible, this is partly because these defeaters are absent (*Asymmetry*).

It accomplishes this feat, in part, because it avails itself of the inferentialist conception of defeasibility outlined above. For inferentialists, the defeasibility of material rules of inference reflects the fact that we organise our knowledge in generalizations that are implicitly understood to rest on a whole range of conditions. Because this dependence on conditions is always already understood and (ideally)¹⁷⁵ understood *by anyone* who has mastered the relevant concepts, the concomitant defeasibility does not cast a poor light on the judgments that such concepts-qua-universal-rules allow us to infer or derive. *Pace* proponents of the subsumptive conception of rationality, this account of our principles of duty as defeasible does not leave us puzzled as to how, in deriving actions from such principles, we can do something that is either consistent or inconsistent with our overall moral outlook or indeed with the moral outlook of others. In this, the inferentialist conception of defeasibility differs from both the view that defeasibility is a marker of fallibility (Korsgaard) and the view that it points to a plurality of competing concerns that we have to balance (Herman).

With this in mind, let us consider the broader implications of the Agent-Scope Reading, beginning with other aspects of Kant’s ethics.

3. Re-Interpreting Kant’s Ethics

The Agent-Scope Reading was introduced to deal with a specific range of problems: with Kant’s alleged failure to appreciate the normative significance of circumstances,

¹⁷⁵ I say “ideally” because, as we mentioned in sect. 2.3 above, the present account does not draw a sharp distinction between our views about the meaning of terms and our views about the world. So to the extent that we have disagreements about the latter, we will also have disagreements about the former, e.g. disagreements about whether the validity of a given principle really depends on a given condition or not.

to accommodate exceptions, and so on. However, even the sympathetic reader might worry that, with regard to other difficulties that exercise Kant commentators, this reading raises more questions than it answers. To allay this suspicion, I will now go through some of these questions, though not with the aim of offering solutions. My aim is less ambitious: it is merely to highlight how the Agent-Scope Reading bears on certain existing debates.

3.1 Acting from Duty

In the first section of the *Groundwork*, Kant introduces the distinction between acting *from duty* and acting merely *in accordance with duty* to explicate the concept of a good will (G 4:397-8). This distinction has provoked a lot of controversy, not least because Kant's Neo-Aristotelian critics have treated it as evidence of yet another variety of rigorism in Kant: motivational rigorism. The critics take issue with Kant's claim that only actions performed from duty have moral worth. In their view, Kant is not only wrong to deny that actions from certain other-regarding emotions such as love or sympathy can be morally worthy, he is also guilty of neglecting that, in many cases, expressing emotions is precisely what morality demands (Williams 1973: 225-9, and Stocker 1976: 453-5).

The most influential response on Kant's behalf is the so-called non-accidentality response.¹⁷⁶ According to this line of response, Kant's opponents misunderstand his objection to dutiful action from inclination. Kant's complaint is not that inclinations are essentially partial, nor that they are unreliable guides to dutiful conduct. His issue is rather that, even if someone's inclinations were such that they *did* lead to moral outcomes with perfect regularity, this would be a mere accident. It would depend on a "fortuitous alignment of motives and circumstances" (Herman 1993: 6). On this basis, Kantians can refute the charge of motivational rigorism by, on the one hand, allowing that emotions can be involved in various ways as long as their involvement does not render dutiful conduct accidental, and, on the other hand, debunking the critics' idea that anything more than this is required for a plausible account of moral worth. They can admit, for example, that acting from duty is compatible with being

¹⁷⁶ For statements of this response, see Herman 1993: 3-6 and Baron 1995: 130, 173. This is not the only Kantian line of response to the above objection, however. Among other strategies, Kantians have also appealed to a distinction between motive and manner of performance (Tannenbaum 2002: 324-7).

inclined to act in this way, with relying on one's affective dispositions to determine what exactly the situation demands, and even with acting from a sensible incentive, if the incorporation of that incentive into one's maxim is sensitive to moral demands.¹⁷⁷

My aim in this section is to show that this response will not satisfy the critics, unless it is modified so as to reflect the distinction between two kinds of counterfactual robustness that I introduced as part of the Agent-Scope Reading in sect. 2.2. We can begin by looking at Kant's discussion of dutiful action in the *Groundwork*. Having set aside all actions that are contrary to duty, he distinguishes between two kinds of actions that accord with duty but are not performed from duty: actions from mediate inclination and actions from immediate inclination. The former, he claims, are easily recognized for what they are, whereas, with respect to the latter, we are prone to confusion. For example, if a shopkeeper "keeps a fixed general price for everyone" when "there is a good deal of trade" (G 4:397) but overcharges his inexperienced customers when he considers himself to be unobserved, it is safe to assume that he is not acting from duty. Since he is only inclined to charge fair prices when the circumstances make this policy an efficient means to profit-maximisation, since he is only *mediately* inclined to do so, his conduct when unobserved will attest to his unworthy motives. By contrast, if someone is "so sympathetically attuned that, without any further motive of vanity or self-interest they find an inner satisfaction in spreading joy around them" (G 4:398), that is, if someone is *immediately* inclined to beneficence, then it is far more difficult to discern whether their actions are performed from duty or from inclination. It is harder to assess the worthiness of their motives because they have a more stable tendency to act in conformity with duty. So the hypothesis that they are acting from duty is harder to falsify.

According to proponents of the non-accidentality response, the two cases are alike in that, in both cases, the agents' dutiful conduct is a mere accident: it is a mere accident that the shopkeeper and the philanthropist do what duty demands. It seems to me that this account underestimates the difference between the two cases and, in

¹⁷⁷ For Kant's thesis that incentives have to be incorporated into maxims, see R 6:24. This thesis is known as Kant's "Incorporation Thesis" – a term coined by Henry Allison (1990: 39-40). Note, also, that, according to Herman, non-accidentality is necessary but not sufficient for an action's being done from duty and, thus, for its having moral worth. This is because, in her view, there are some actions – permissible actions – that are not done *from* duty, despite being in non-accidental conformity with moral demands. In these cases, the motive of duty operates as a limiting or secondary motive. Herman 1993: 13-7.

order to bring out how it does, it will be useful to briefly consider Herman's version of the response. Herman is alive to the fact that sympathy is a more reliable guide to dutiful action than greed (Herman 1993: 4), but, in her view, the difference is a difference in degree:

Is the motive of sympathy only fortunate when it hits on a right action? Doesn't it necessarily prompt a person to help others? Suppose I see someone struggling, late at night, with a heavy burden at the backdoor of the Museum of Fine Arts. Because of my sympathetic temper I feel the immediate inclination to help him out. (Herman 1993: 4)

Herman leaves it to the reader to fill in the gaps here. Her point is: at first glance, it might seem as if the philanthropist's dutiful action is *not* a mere accident because sympathy is a disposition to help others and, as such, a disposition to fulfil a wide duty. However, on closer examination, she finds that it is a mere accident after all, because, even for a philanthropist with a perfect "moral record", we can imagine circumstances in which their sympathy, their willingness to help, would lead them astray. For example, as someone who enjoys helping other people, *any* other people, the philanthropist would be as willing to cover for an art thief as they would be to rescue a child. What does this line of thought tell us about Herman's conception of (non-)accidentality? Presumably, she would not feel the need to show that there are circumstances in which sympathy leads us astray, if she did not think that a failure to come up with such examples was tantamount to a concession – to the concession that actions from sympathy are non-accidentally dutiful and hence morally worthy.

However, if this is right, the above line of thought commits her to a conception of non-accidentality that is grist to the mill of the critics. It is grist to the mill of the critics because, as Neo-Aristotelians, they agree that morally worthy actions cannot be accidental in *this* sense: such actions have to result from a proper responsiveness of motives to circumstances, not from a mere fortuitous alignment between the two. When these critics say that someone exhibits the virtue of sympathy or beneficence, they mean that this person is disposed to help others in circumstances where doing so is appropriate, and not in circumstances where it is not. As a virtuous person, the philanthropist is someone who has the *right* affective dispositions in this area; they are someone who would never help art thieves or any other criminals. Since sympathy in *this* sense cannot lead its bearer astray, Herman's non-accidentality response

would fail to convince the virtue ethicist that helping actions from sympathy have no moral worth. It is not convincing because it seems to overlook the possibility of someone having perfectly virtuous inclinations.

I believe that we can salvage the non-accidentality response if we pay attention to the above distinction between objectively relevant variations in circumstances and merely subjectively relevant variations. In fact, Kant himself implicitly draws on this distinction when he contrasts actions from mediate inclination with actions from immediate inclination. As many commentators have observed (e.g. Baron 1995: 146-7, Timmermann 2009: 46), each of his examples of actions from immediate inclination features a protagonist who goes through a change: we are asked to imagine, for example, that “the mind of [the] philanthropist [is] overclouded by his own grief, which extinguishe[s] all sympathy with others” (G 4:398). This thought experiment, I submit, serves the same function as the unobserved shopkeeper case. It is meant to draw attention to a set of circumstances in which we could judge the agent’s motives on the basis of their outward conduct. If, having lost all sympathy with others, the agent stopped helping, then we would have reason (though perhaps not conclusive reason)¹⁷⁸ to believe that he used to help others from inclination, not from duty. This scenario is importantly different from the unobserved shopkeeper scenario, however, because, in this case, the grieving agent would not be led astray by his *sympathy* (whereas the unobserved shopkeeper is led astray by his *greed*). Before he was struck by grief and lost his sympathy, he may well have had perfectly virtuous inclinations that disposed him to do his duty in each and every situation – as the Neo-Aristotelian insists. But the fact that he conformed with duty would still have been a mere accident. For although, for some time, his disposition *was* “virtuous” in the sense that it was properly responsive to how the circumstances bear on what he, as an agent, ought to do (e.g. to the import of whether the recipient of his helping action is a thief), it only took a distorting factor like grief to throw him, as a judging subject, off balance.¹⁷⁹ Once this factor prompted a change in his inclinations (or, as a Neo-

¹⁷⁸ Herman points out that he could have become a bad person overnight. This is a possibility that Kant seems to rule out when he says that revolutions of the heart can only go one way: from bad to good (R 6:47).

¹⁷⁹ I have put the term “virtuous” in inverted commas because Kant would not call such a person virtuous – and perhaps a virtue ethicist wouldn’t either. It is worth keeping in mind that Kant is not arguing against virtue ethicists here, but rather against sentimentalists. If and to the extent that virtue ethicists maintain that we are responsible for our affective dispositions and for failures to retain them in the face of adverse circumstances, their view is quite close to Kant’s. Note, also, that there is an

Aristotelian might say: in his moral sensibility), his moral judgments and actions changed as well.

This conception of accidentality is not only better suited to refute the charge of motivational rigorism, it also makes better sense of what Kant says. For if Herman was right that actions from mediate and immediate inclination differ only in degree, and, importantly, that their being in accord with duty is accidental in the same sense, then Kant's discussion of changes in inclinations would seem rather odd. In order to make his point, he would not have had to imagine a philanthropist who goes through a change; instead, he could have imagined a philanthropist who is confronted with an art thief or some other criminal.

3.2 Hypothetical and Categorical Imperatives

Of all the claims and doctrines that make up Kant's ethics, his use of the distinction between categorical and hypothetical imperatives as a way of marking the difference between how agents are bound by moral norms and how they are bound by instrumental norms is perhaps the most widely accepted (G 4:414-6).¹⁸⁰ Nonetheless, there is far-reaching disagreement about each of the constituents of this distinction. In this section, I want to focus on one of the main controversies concerning hypothetical imperatives: the controversy about the scope of the "ought" that figures in them.¹⁸¹ This question is very important, partly, because it is, at bottom, a puzzle about how Kantians can make sense of the relation between moral norms, on the one hand, and norms of instrumental rationality, on the other.

The controversy goes back to a seminal essay of Thomas Hill's, entitled "The Hypothetical Imperative". In this essay, Hill argues that, despite never having said so explicitly, Kant must have assumed that there is a "fundamental rational principle" (1973: 430), *the* Hypothetical Imperative, that enjoins all rational agents as such to either take the necessary means to their ends or abandon these ends. Although en-

interesting parallel between someone who is "virtuous" in this deficient sense and someone whose representations of the world are accurate due to a pre-established harmony – a possibility that Kant entertains in the much discussed §27 of the first *Critique*. In both cases, the judgments or representations are perfectly accurate throughout, but only accidentally so.

¹⁸⁰ At least, this distinction seems to have found its way into contemporary debates in metaethics that are otherwise quite independent of debates in Kantian ethics. See e.g. Wallace 2001 and Setiya 2007.

¹⁸¹ This question has been discussed in tandem with a further question, namely the question whether, besides the Categorical Imperative, there is also a principle called *the* Hypothetical Imperative, in the singular. See e.g. Schroeder 2005. I shall leave this second question aside.

dorsed by many (e.g. Korsgaard 1997: 234, Darwall 1983: 16, 47), this *wide-scope* reading of the dictum “If you will the end, then you ought to take the necessary means” has also met with some resistance.¹⁸² Hill’s opponents argue for a *narrow-scope* (or consequent-scope) reading, according to which hypothetical imperatives enjoin subsets of rational agents, namely those who will certain ends, to take the means to these ends. In a nutshell, the disagreement between the two camps concerns both *whom* hypothetical imperatives enjoin to do something and *what* they enjoin them to do. The difference between them can be represented as follows.

$HI_{(ws)}$ I ought (If I will end e, then I take the necessary means m to e).

$HI_{(ns)}$ If I will end e, then I ought (I take the necessary means m to e).

One reason why Hill’s wide-scope interpretation has met with resistance is that it seems to lack textual support. As Hill’s opponents have noted, the narrow-scope interpretation is much easier to square with what Kant says (Schroeder 2005: 359-61, Siyar 2013: 445-53). Their main point is the following: if we take Kant at his word and assume that hypothetical imperatives differ from categorical imperatives in that the former apply to agents under a condition (this is thought to be the condition of willing an end),¹⁸³ while the latter apply unconditionally, to all finite rational agents as such, then, *by Kant’s lights*, $HI_{(ws)}$ is a categorical imperative.¹⁸⁴ This suggests that the wide-scope reading blurs the very distinction that it purports to shed light on. Relatedly, it makes “complete nonsense” (Schroeder 2005: 361) of Kant’s claim that there is a special difficulty in demonstrating the possibility of moral imperatives because, qua categorical, they are synthetic, whereas hypothetical imperatives are taken to be analytic. To say that hypothetical imperatives are analytic is, roughly speaking, to say that, in willing an end, we already commit ourselves to willing the means to this end. Establishing the possibility of such imperatives is trivial because, in thinking of an agent *as* someone who wills an end e, we already think of them as someone who is committed to taking the means to e, and so there is nothing puzzling about the fact

¹⁸² The term “wide-scope” is supposed to convey the idea that the scope of the “ought” is wide in the sense that it ranges over both the antecedent and the consequent of the above dictum. See the contrast of $HI_{(ws)}$ and $HI_{(ns)}$ below.

¹⁸³ I think that this is not, strictly speaking, correct. I would argue that the bindingness of hypothetical imperatives is conditional on the correctness of the agent’s judgment that the end is good. Of course, in part, this judgment is based on the agent’s perception of the end as pleasurable. See Newton 2013.

¹⁸⁴ In fact, this point is already acknowledged in Hill’s classic discussion (1973: 440).

that they are required to take the relevant means. To say that categorical imperatives are synthetic is accordingly to say that there is nothing in our thought of a finite rational agent as such that would suggest that they are required to only act on maxims that can be willed as universal laws. This is why the possibility of such imperatives is a puzzle and why establishing it is so difficult. But if Hill was right, if the Hypothetical Imperative was a wide-scope requirement that applied unconditionally to all finite rational agents as such, then its possibility would pose the very same difficulty, because, surely, there is nothing in our thought of a finite rational agent as such that would suggest that they are required to either abandon their ends or take the means to them. In light of these problems, we must ask why so many Kantians opt for a wide-scope reading anyway.

The motivation for interpreting hypothetical imperatives as wide-scope requirements is that the narrow-scope reading seems to have unacceptable implications. For if we suppose that agents can will evil ends, e.g. the end of murdering someone, then, using an instance of $HI_{(ms)}$, we can deduce (or “detach”) the claim that they ought to take the necessary means to this end, e.g. load their gun, point it etc. – and that is plainly false. This is where the wide-scope reading is thought to come to the rescue. When combined with the Categorical Imperative, $HI_{(ms)}$ says that, in such a case, the agent ought *not* to take the means to their evil end. What they ought to do, instead, is abandon the end. In response, proponents of the narrow-scope interpretation reject the supposition from which their opponents begin. Given that Kant identifies the will with practical reason (G 4:412), and that the Categorical Imperative is meant to be a constraint *on willing* (and not a constraint on the decision to follow through with what one already wills), it is, in their view, simply impossible to will an evil end, in the Kantian sense of “will”. As finite rational agents, we can certainly *choose* such ends, but we cannot will them (Engstrom 2009: 44-9 and Schroeder 2005: 367-70).¹⁸⁵ Now, all of this is right (after all, Kant does have a normative conception of willing), but it leaves us with the feeling that a potentially substantive disagreement has been settled by sheer stipulation. Let me explain why this worry might arise.

Consider Simon Rippon’s (2014) objection to the narrow-scopers’ response that one cannot will an evil end. In his view, this response cannot completely defuse the detachment objection to the narrow-scope reading because, even if we grant that

¹⁸⁵ For Kant’s account of the free power of choice, see MM 6:213.

willing impermissible ends is impossible, there are still other false ought claims that this reading allows us to detach. This additional detachment problem arises because many of our ends are merely *permissible* rather than *obligatory*. If I will an end that is merely permissible, e.g. to make myself a cup of tea, then it seems to follow from $HI_{(ns)}$ that I ought to take the necessary means, e.g. take out a cup, boil water, etc., and that doing so is the *only* rational thing for me to do. If this were true, it would be a problem indeed. After all, Rippon is right when he says that “it would be highly implausible to claim that you would always be irrational to revise your ends” (2014: 785). But does the narrow-scope reading really have this implausible implication? On the face of it, there is, once again, a simple way of showing that the objection rests on a misunderstanding. In this case, the misunderstanding concerns the temporal structure of willing. As Hill observes, willing an end is not deciding or resolving to pursue it. Willing is not a momentary occurrence, it is an ongoing commitment (Hill 1973: 433-4). With this in mind, proponents of the narrow-scope interpretation could reply that it is only *as long as* the fickle tea drinker is committed to making a cup of tea that they satisfy the condition under which they are subject to the relevant instance of $HI_{(ns)}$, which enjoins them to take out a cup, boil water, etc. Once they abandon this commitment, which they are free to do at any time, they are no longer required to take the necessary means. However, the fact that Hill adduces this consideration when arguing for the wide-scope reading should make us pause.

It seems to me that Rippon’s objection touches on a deeper issue, and that, in light of this, the temporal structure response only makes things worse, not better. The deeper issue is that the narrow-scope reading seems to distort our understanding of instrumental reasoning by modelling it on moral reasoning. Let us consider how it represents our reasoning from ends to means. When I will an obligatory end, e.g. to help someone in need, and do so for the right reasons, there is something that I am getting right. My judgment that I ought to pursue this end is based on *objective* grounds. As such, this judgment is a perfectly good basis for other ought judgments, such as the judgment that I ought to take out some cash. When I will a merely permissible end, by contrast, e.g. to have a cup of tea, then part of what I am responding to are *purely subjective* sensible desires – and this, we want to say, is how it should be. But the involvement of sensible desires seems to change the character of my state. Unlike willing to help someone, willing a cup of tea seems like a state that I just hap-

pen to be in. So if it were to serve as the basis for another judgment, e.g. the judgment that I ought to take out a cup, it could only be a deficient basis: a *cause*, not a valid *ground*. In light of this, the appeal to the temporal structure of willing only makes things worse because it actually highlights this apparent deficiency of my judgment that I ought to take out a cup, that is, its lack of valid grounds. For, on reflection, to say that I can make and then abandon this judgment like a passing fancy is simply to concede that it does not have the temporal structure of a judgment proper, which, as such, is meant to be sustained.¹⁸⁶

Now we can see that, although Rippon's objection is not strictly speaking valid, he is nonetheless onto something: the narrow-scope interpretation exerts a certain pressure to conceive of our commitments to permissible ends as obligations. This is also what Robert Brandom has in mind when he complains that Kantians "assimilate[] all reasons for action to [moral reasons]" and that, in their view, desires have to always be "accompanied by the acknowledgment of some corresponding obligation or commitment" (2000: 92). These considerations make the wide-scope reading and its account of the involvement of sensible desires seem very appealing. For what this reading seems to get right is the idea that, within the realm of the permissible, we can choose to pursue or abandon our ends as we please without thereby violating any imperatives of reason. Does this mean that the narrow-scope interpretation should be abandoned?

In my view, it does not. To see why, we need to consider the above concern in light of the distinction between the Case-Scope and the Agent-Scope Reading of the *Universality* desideratum. First, it is worth noting that the idea that our commitments to permissible ends are in some sense binding follows very naturally from the kinds of considerations that Korsgaard seems to cite in support of Case-Scope universal principles, namely the thought that our ability to determine ourselves depends on our having standing commitments or being persons of principle (see her premise 3a, as cited in sect. 1).¹⁸⁷ The Agent-Scope Reading helps us to see how proponents of the narrow-scope reading can evade this implausible commitment and how they can accommodate the intuition that the fickle tea drinker does not violate any imperatives

¹⁸⁶ Engstrom 2009: 103-4. The state that we are considering would have the form of the "I feel", not the form of the "I think". See Newton 2013.

¹⁸⁷ Of course, this is not to say that proponents of the Case-Scope Reading could not possibly evade this commitment. Instead of considering this question any further, I will limit myself to showing how narrow-scope advocates can evade it.

of reason. The key to understanding why they don't is the aforementioned distinction between the subject as the one who is judging, on the one hand, and the subject as the one who is acting, on the other.

In sect. 2.4, we considered this distinction in relation to how circumstances can enter into an explanation of why someone judged the way they did, and in this context – notably: in the context of a discussion of *moral* rules – feelings seemed to come in only as distorting factors, e.g. as part of the explanation why the fact that a given lie would benefit my friend made me inclined to tell it. However, in fact, we can draw the very same distinction with respect to how feelings themselves can enter into our explanations of why someone judged the way they did. If they enter as features of the subject as the one acting without affecting the subject as the one judging, then their bearing on the judgment is, again, harmless: it does not reflect a distortion or lack of valid grounds; it does not attest to the violation of an imperative of reason. Instead, it reflects a preference, considered as a property of the object being judged, and, as such, it has every right to enter into a judgment *about* this object. It is one of the three conditions under which it is appropriate to judge, of oneself, that one's pursuit of a certain end is overall good, that is, to adopt the end – the other conditions being that it is permissible and realizable. And it is *as someone of whom this judgment is true* that one is bound by the narrow-scope hypothetical imperative to take the necessary means to this end.¹⁸⁸

With this in mind, we can now clarify how *Universality_(AS)* relates to Kant's Formula of Universal Law, which will be the topic of the next section. According to the latter formula, we ought to act only in accordance with maxims of which we can at the same time will that they become universal laws, i.e. in accordance with maxims which are universalisable. Kant uses the term "maxim" to refer to our principles of

¹⁸⁸ In addition to dispelling the concern that the narrow-scope reading allows us to detach false ought-judgments, the above account of hypothetical imperatives has the great advantage of making sense of Kant's belief that hypothetical imperatives are analytic, whereas categorical imperatives are synthetic. Recall what we said about Kant's analytic-synthetic distinction in sect. 2.3: a judgment is analytic if the nexus between subject and predicate can be established without appeal to an external ground or third term, and synthetic if it can't. Now consider the judgment "I ought to take out a cup". If the "I" in this judgment refers to me *as someone who ought to make themselves a cup of tea* (which is simply a different way of saying that it is *overall good for me* to make myself a cup of tea), we can see that this judgment is true without appealing to any additional ought-judgments – although, of course, we do need to appeal to an is-judgment, namely the judgment that taking out a cup is a necessary means to making a cup of tea. Next consider the judgment "I ought not to lie". The "I" in this judgment refers to me *as a finite rational agent*. Clearly, in this case, we cannot establish that the judgment is true without appealing to a further ought-judgment that says something about what finite rational agents as such ought to do.

action. Principles of action capture our reasons for action. Some of these reasons are *moral* reasons. They are the reasons we have independently of our feelings, and, as such, they can be captured in universal principles of duty: in principles that bind all rational agents in the same circumstances (as *Universality*_(AS) says).¹⁸⁹ Other reasons of ours are *non-moral* reasons. Which non-moral reasons we have depends, in part, on our feelings. Since some maxims specify non-moral reasons, it would not make sense to demand that maxims, as such, should be *universal*, in the sense that others in the same circumstances would necessarily have to have the same reason that we have.¹⁹⁰ What they do have to be is *universalisable*: they must specify reasons that anyone else in the same circumstances would be allowed to treat as such.

3.3 Applying the Formula of Universal Law

It is now time to turn to the most famous component of Kant's ethics: the so-called Formula of Universal Law, and its close relative, the Formula of the Law of Nature (G 4:421).

Formula of Universal Law: Act only in accordance with that maxim through which you can at the same time will that it become a universal law.

Formula of the Law of Nature: Act as if the maxim of your action were to become by your will a universal law of nature.

To illustrate how these formulas are to be applied, Kant provides four examples, which have prompted a long-standing debate consisting of cycles of counterexamples and re-interpretations.¹⁹¹ Among the questions discussed in the context of this debate are the following: can we derive substantive moral precepts from these formulas and, if so, how? Or are these formulas actually empty, as Hegel contends?¹⁹² What is the nature of the two kinds of contradictions, contradictions in conception and in will-

¹⁸⁹ This is why, in our discussion of principles of duty in sect. 2.4, feelings figured only as distorting factors: because they have no bearing on what moral reasons we have.

¹⁹⁰ Only *some* maxims have to be universal, namely those that specify moral reasons or, to put it in Kant's own terms, those that specify obligatory ends.

¹⁹¹ For an overview, see Allison 2011: 176-203.

¹⁹² Hegel's famous objection to Kant's Formula of Universal Law has come to be known as the Empty Formalism Objection (see e.g. 2011: §135). According to Hegel, the universalisability test cannot be applied without presupposing moral content. For example, in order to show that there is a contradiction in willing a maxim of theft together with its universalised counterpart, we need to presuppose that there ought to be property. But if we presuppose this, then we do not need the universalisability test to tell us that others' property should be respected. In this section, I hope to indicate why the Empty Formalism Objection may not be warranted.

ing, to which the so-called universalisability test is supposed to give rise? Is there a unified account of these two kinds of contradictions that preserves the difference between them? And do the results of Kant's test accord with our moral intuitions? These questions are, of course, closely intertwined – to provide a satisfactory answer to one is to provide satisfactory answers to all.

Even among Kantians it is generally acknowledged that a fully satisfactory set of answers to these questions is yet to be given, and this is certainly not the place to close this gap. My aim in this section is far more limited: I want to defuse the worry that the Agent-Scope Reading makes things worse. This is an understandable worry, given that most of the interpretations on offer – including the strongest contender, Korsgaard's so-called practical contradiction interpretation – do not merely happen to involve a commitment to the Case-Scope Reading, but actually depend on this commitment to generate contradictions.¹⁹³ Take Korsgaard's practical contradiction interpretation as an example. According to Korsgaard, the contradiction involved in willing an immoral maxim together with its universalisation is practical in the sense that, in so willing, the agent wills a world in which her action is no longer a means to her end – either the end specified in the maxim itself (contradiction in conception) or an end that is essential to her will as a finite rational agent (contradiction in willing) (Korsgaard 1996a: 92-4). Consider Kant's example of a maxim of false promising.

False Promising: “When I believe myself to be in need of money, I shall borrow money and promise to repay it, even though I know that this will never happen” (G 4:422).

“Promises,” Korsgaard notes, “are efficacious in securing loans only because they are believed, and they are believed only if they are normally true” (1996a: 92). If they were normally false, then the practice of promising would break down and promises, true or false, would no longer be means to securing loans. So, in Korsgaard's view, both kinds of contradictions arise, in part, because willing an immoral maxim as a universal law means willing that an existing practice be violated or abused with such frequency that it would inevitably break down.¹⁹⁴

¹⁹³ In this regard, Engstrom's reading of the Categorical Imperative is an exception (2009: 188-9, 195).

¹⁹⁴ This is why she struggles to show that acts of natural violence are ruled out (Korsgaard 1996a: 97-101).

Note that, in order to generate Korsgaard's practical contradictions, the Formula of Universal Law would have to be interpreted along Case-Scope lines. For if, by "universal law", we meant that a principle should hold for all agents *in the circumstances under consideration* (Agent-Scope Reading), then universalising a maxim wouldn't be a matter of imagining that the relevant action is performed *whenever* someone has the relevant end, nor that it is performed with any high frequency. No two situations are exactly alike, so even if my maxim was deeply immoral, willing that any rational agent in the same situation be allowed to act on it would not mean willing the collapse of any practice or convention on which I have to rely – either in this very action or otherwise. These observations suggest that proponents of the Agent-Scope Reading are left without *any* account of how contradictions are meant to arise, let alone an account that could satisfy the Hegelian critic. For how are we supposed to determine whether one can will a given maxim together with its universalization in the Agent-Scope sense without presupposing any moral content (e.g. that the practice of promising is valuable)?

In what follows, I will offer a reading of the universalisability test that takes its cue from the reading offered by Stephen Engstrom (2009: 185-229), but that also goes beyond the latter, especially in its appeal to the inferentialist framework outlined above.

First, it is important to appreciate that, within the framework introduced in sect. 2, the Empty Formalism Objection takes on a different form. Let us recall what we said about concepts and their role in securing the possibility of agreement and disagreement between our representations of objects of experience. The role of concepts, we said, is to set standards that determine what has to be true of objects falling under them under various kinds of circumstances, and to thereby secure the distinction between how things are with the object and how a particular subject may experience them as being. If a given object falls under the concept of an apple tree, for example, and the circumstances are such that no woodpecker or other animal has tampered with it, then it will have a completely impenetrable trunk. To judge that it does not – for instance, because, when touching it, one has the illusion that it gives in – is to treat merely subjective grounds as objective and, thus, to misjudge. It is to judge on the basis of a rule ("If this is a tree, it has a partly penetrable trunk") that *contradicts* one of the universal rules that make up the concept of a tree ("If x is a tree,

x has a completely impenetrable trunk”). Can we exploit these observations to shed light on the kinds of contradictions that the universalisability test is meant to reveal?

There is an obstacle: the content of deontic concepts cannot be derived from experience. Deontic concepts are artefacts of our ways of life which are arguably as ever-changing as these ways of life themselves.¹⁹⁵ So while we can learn *by observation* that, if they are not tampered with, apple tree trunks are completely impenetrable, the question of whether someone in need is allowed to make a false promise cannot be answered in this fashion (in fact, Kant cautions against basing practical judgments on feelings in the way that theoretical judgments are based on sensations, CPrR 5:58). It is at this point that we re-encounter the Empty Formalism Objection, albeit in an importantly different form: if we *presuppose* that it is part of our shared understanding of the concept of a promise that, in these kinds of circumstances, promises ought to be made in good faith, then we can see how my acting on a maxim of false promising in this situation gives rise to a contradiction – but who says that we should presuppose this? The concern behind this critical question is no longer a concern about having to presuppose moral content at all. We already know that concepts could not secure the objective purport of our moral judgments if they did not set standards that determine what agents acting under these concepts ought to do in various kinds of circumstances. The question is merely *which* standards they set. This point is crucial because it does away with the misguided notion that willing the frequent violation of a practice or convention is essential to contradicting oneself in the sense that Kant envisages. If the rule that, in a given set of circumstances, a promise ought to be made in good faith belongs to the concept of a promise, then, being in these circumstances, I cannot promise in bad faith without contradicting myself, and it is then immaterial whether the frequent violation of that rule would lead to an atmosphere of mutual distrust (for a similar point, see Engstrom 2009: 188-9, 195). With this in mind, we can now turn to the question of content.

Most of the time, what agents, qua objects of practical judgment, ought to do in a given set of circumstances depends on their ends: whether an agent ought to drink depends on whether they will the end of quenching their thirst; similarly, whether an object has a trunk depends on whether it is a tree. However, in some instances, what an agent ought to do is independent of their ends. It is worth noting that analogous

¹⁹⁵ This may not have been Kant’s view, but it is, in any case, very plausible.

cases can be found in the theoretical domain as well: whether an object is a squared circle, for example, does not depend on what kind of object it is, and whether an event has a cause does not depend on what kind of event it is. I want to argue that the former case is, in some respects, analogous to a contradiction in conception, and the latter to a contradiction in willing. Let us begin with the former and examine why the false promisor's employment of the concept of a promise entails a commitment to promising in good faith.

On the face of it, the false promisor's concept seems like a perfectly coherent concept that does not involve any contradictions in the way that the pseudo-concept of a squared circle does. This impression is misleading, however. To begin with, it is worth noting that there are some circumstances in which to say "I promise to do x" is not generally presumed to express a genuine commitment to doing x: when it is said on stage, for example, or when one is being coerced. So perhaps the promisor in Kant's example believes that they are in such a situation. Perhaps it is part of the concept of promising that they are relying on that, in circumstances of dire need, to say "I promise" is, basically, to beg for charity. If this were so, then their knowledge that they will not be able to repay the promisee would not contradict any of the commitments incurred by making the promise. (Compare this to a game in which "square" means round; in the context of such a game, the concept of a squared circle would not be contradictory). Moreover, it is clearly possible to *think of*, and indeed to *will*, this rule as a universal law, as one of the laws that together determine what agents ought to do in various kinds of circumstances and thus secure the possibility of objective practical knowledge.¹⁹⁶ That is all well and good, except this is *not* the concept that the promisor is relying on. In fact, the promisor wants to have their cake and eat it. They might "mean" to beg, but they also want to be understood as making a genuine promise. Their maxim is fit for purpose only because it exploits

¹⁹⁶ Kant might disagree. In his view, promises and acceptances are two acts that co-constitute a contract. And despite the fact that the parties to a contract make their respective declarations at different times, he argues that the two acts are really one: a joining of two wills, which creates an obligation to deliver what is being promised, e.g. a service or commodity (MM 6:272-3). In light of this, it might be objected that I have not managed to defend the view sketched above, namely that the principle "One ought not to make false promises" can be universally valid even if it does not hold in all circumstances. All I have shown, it might be said, is that, in some circumstances, a certain string of words ("I promise") does not amount to a promise at all. Similarly, it could be argued that a white lie is not a real lie, because a thin-skinned friend asking for my opinion on their new haircut simply cannot expect to be told the truth and hence should not rely on my response. This brings us back to the discussion of the Casuistical Questions Response in ch. 1, sect. 4. There I argued that derivative uses of concepts deserve to be taken seriously; that they are still uses *of the relevant concepts*. The same point applies here.

two inconsistent concepts of promising. This is why it cannot even be *conceived* as a universal law or part of a genuine concept. (Compare this to a game where one can use “square” to mean square or round, as one pleases).¹⁹⁷

Let us turn to contradictions in willing. To contradict oneself in willing is not to act under such a pseudo-concept. It is more akin to judging that a given event has no cause: something that can be *thought* without a contradiction, but cannot be *experienced* by finite beings like us. Consider what our agential finitude amounts to, according to Kant. For finite rational agents with a lower faculty of desire, happiness is a necessary end. Happiness is a state in which the agreeable is attainable, a state in which we are self-sufficient, where “everything goes according to wish and will” (CPrR 5:124). But happiness is not the same as blessedness: while both are states of self-sufficiency, the former, unlike the latter, depends on *fortune*. Accordingly, our pursuit of happiness is a pursuit that is informed by the recognition that our powers are limited and that our ability to attain self-sufficiency depends on factors beyond our control. One of the factors beyond our control is the extent to which the attainability of our goals will depend on the actions and the cooperativeness of other people. It is, for example, beyond our control whether we are born with a disability that makes us dependent on other people, or whether we inherit enough money to fund our own education, or whether the talents that we happen to have are valued at a given point in time. With this in mind, we can return to the question of whether and how morality can impose substantive constraints on the deontic concepts that communities of finite rational

¹⁹⁷ This account of contradictions in conception may seem to render Kant’s ethics *relativistic*. After all, it seems to imply that the permissibility of acting on the above maxim of false promising depends, in part, on how one’s linguistic community happens to understand the concept of promising at a given point in time, or on the characteristics of an existing practice in a given cultural and historical context. So perhaps Hegel is right after all: perhaps we do have to presuppose some moral content, though not as individuals applying the universalisability test (subjectivism), but rather as communities (relativism). Although this concern would merit a far more detailed discussion, I shall limit myself to two points. First, it is true that my reading leaves considerable room for local variations in moral codes. It assumes (perhaps against Kant) that, *to a certain extent*, what is right and wrong does depend on the cultural and historical context (on this point, see Herman 2007: 143-5). But – and this is the second point – there are constraints, and these *can be* derived from the Formula of Universal Law and the other formulas of the Categorical Imperative, though not without doing argumentative work. Oftentimes, Kant’s critics look at the arguments in the *Groundwork* and jump to the conclusion that Kant’s ethics is empty. But Kant does not just presuppose judgments such as the judgment that there ought to be property or contracts or a practice of charity. He argues for them in the *Metaphysics of Morals*, using variations of the principle established in the *Groundwork*. The account of contradictions in willing that I present in the following paragraph gives some indication of what an argument for a practice of charity might look like, for example. I will return to the threats of subjectivism and relativism in ch. 4.

agents form and use. Consider Kant's example of a maxim of never helping others.¹⁹⁸ It is certainly possible for a community of finite rational agents to use the concept of helping in such a way that helping others is completely voluntary and that receiving help is demeaning. But since only *some* members of this community would possess the means to help, it would be entirely up to them and no one else how the concept of helping is used and how it evolves. Accordingly, this community would not have a *truly* shared understanding of this concept, even if all or most of its members acquiesced to using it in the way that its privileged members did.¹⁹⁹

4. Taking Stock: The *Particularist Challenge* and Beyond

To conclude the first part of this thesis, I want to briefly reflect on what my Kantian brand of generalism can contribute to the contemporary debate between generalists and particularists. In the introduction (sect. 1.2), I noted that the *Particularist Challenge* to Kantian ethics is an instance of the broader challenge that particularists mount against their generalist opponents: the challenge to devise moral principles that are both deserving of their name *and* suited to accommodate holism. More specifically, I pointed out that principles proper are expected to fulfil two demands, which holism seems to render incompatible: the demand to cover a determinate range of cases and the demand to play the role of guides and/or explanatory standards. By showing how the Agent-Scope Reading of *Universality* and the concomitant account of principles of duty dissolve the trilemma, I have already indicated how my brand of generalism meets some of these demands. Now I want to broaden the perspective further and ask the more general question of how it fares in comparison to other generalist approaches.

My argument in this chapter aims to *meet* the challenge that particularists mount against their opponents, not to reject it, as other generalists have (e.g. McNaughton and Rawling 2000: 267, Raz 2006: 107-13). Accordingly, I have sought to *dissolve* the trilemma instead of embracing one of its horns, as some other Kantians have (e.g. Timmermann 2001: 65-67, Cholbi 2013: 447). Broadly speaking, generalists' attempts to meet their opponents' challenge fall into two categories. Some appeal to unhedged

¹⁹⁸ For an illuminating discussion of how Kantians can appeal to Kant's notion of happiness and its role in practical reasoning to address empty formalism concerns, see Bojanowski 2017: 4, 6, 8, and 10.

¹⁹⁹ This account of contradictions in willing is, of course, evocative of Rawls' original position (1999: 15-19), who I think would agree with my main thesis that moral and political principles should be understood as Agent-Scope universal.

principles (see e.g. Hooker 2008: 22-3), others to various kinds of hedged principles (e.g. Lance and Little 2004, Väyrynen 2009). I touched on some of the difficulties that haunt theories of unhedged principles in ch. 2., sect. 2.3.2 and 3.3.2. Roughly speaking, such principles fulfil the demand of covering a determinate range of cases, but they do not seem suited for the role of guides or explanatory standards, partly, because they obscure the relevant features by embedding them in a host of irrelevant features. In this chapter, I have presented an account of *hedged principles*. The main difficulty that proponents of such accounts face is to devise a type of hedging clause that does not trivialise the principles to which it is attached. Think of the principle “*Ceteris paribus*, lying is impermissible”. If this principle boiled down to the claim that lying is impermissible except when it is not, then this principle, although trivially true, would be utterly useless because it would leave undetermined when lying is impermissible and when it is permissible or obligatory, and why. As such, this principle would fail to fulfil both of the demands that proper moral principles are supposed to fulfil. Does my hedging clause fare any better?

One of the conclusions reached in the above discussion is that Kant’s principles of duty come with the hedging clause “in objectively suitable circumstances”. As I noted in sect. 3.3, one might worry that this hedging clause is *empty*,²⁰⁰ that it is one of the hedging clauses that trivialise the principles to which they are attached. This is a worry that I tried to address by further unpacking this clause. First, I pointed out that the objective suitability of circumstances is determined by our shared understanding of the deontic concept that the relevant principle contains. Consider the principle “In objectively suitable circumstances, lying is impermissible”, and how it applies in the *Murderer at the Door* case. Whether the circumstances in this case are objectively suitable, and thus whether this lie is impermissible, depends on whether it is part of our shared understanding of the concept of a lie that the addressee’s intention to use the information provided to murder an innocent victim functions as a defeater: whether it makes lying permissible or obligatory. This is a first step, but, as I acknowledged in sect. 3.3, it is not enough to fully address the emptiness concern. More substantive constraints are required. I therefore added the following second point. In order to determine whether treating a given feature as a defeater is warranted by our shared

²⁰⁰ We discussed this in the context of Hegel’s Empty Formalism Objection in sect. 3.3 above; see, in particular, fn. 192, 197, and 198.

understanding of the deontic concept in question, I need to ask myself, first of all, whether, if all parties involved shared this understanding of the relevant concept, my envisaged action or interaction could be successful or whether I am presupposing a pseudo-concept (contradiction in conception). If my action does not presuppose that my own understanding of the relevant concept diverges from that of others, then I need to ask a more intricate question: could a community of finite agents like us, with our needs and limitations, choose this concept, so understood, to organise their lives together, or is the fact that this concept and this understanding of it is indeed ours a result of the fact that this concept was *not* chosen or developed together, that it evolved under the influence of power structures, say? (contradiction in willing). For if it is a result of the fact that this concept was not chosen or developed together, then the purportedly shared understanding of it is not *truly* shared.

Once the hedging clause “in objectively suitable circumstances” is unpacked in this way, it becomes apparent that it is not empty: that, although the list of defeaters that this hedging clause admits is open-ended and subject to ongoing negotiations,²⁰¹ it does provide some guidance on how to conduct these negotiations, how to identify defeaters, and how to determine the range of cases in which a given principle of duty is valid.²⁰² As a result, the principles of duty to which this clause is attached are suited to play their role in explanations and to serve as guides. What my argument has also shown, however, is that explaining and guiding are not the only and not the primary

²⁰¹ I think it is the wish to acknowledge this open-endedness, the de facto impossibility of compiling a complete list of defeaters, that leads McDowell to argue that the knowledge of the virtuous person is uncodifiable (1998: 57-65). He does not mean to say that, in morally deliberating, or explaining and justifying our actions to each other, we aren't guided by the thought that we share an understanding of the deontic import of features that could, in principle, be codified as well as articulated. In short, he isn't saying that principles don't have any role to play at all or that we can make sense of a given moral situation without referring to others, as Dancy takes him to be saying, when he presents McDowell as the main source of inspiration for his particularism (e.g. in his 1993: xii, 79-80).

²⁰² My account of hedged principles can be understood as an instance of the broader framework or theory of hedged principles put forward by Pekka Väyrynen (2009). Väyrynen shows that, if we hedge moral principles by reference to so-called normative bases, then they will satisfy the two demands on principles outlined above. They will cover a determinate range of cases and they will be able to serve as explanatory standards. Väyrynen defines a normative basis as the factor because of which a given feature has a normative import and which explains that normative import (2009: 96). Let us assume, for example, that the normative basis of the deontic import of lying is that it destroys trust. If this is so, we can hedge the principle “Lying is impermissible” by adding the clause “unless the lie does not destroy trust”. Väyrynen points out that normative bases need not be features with the same deontic import on a higher level (e.g., higher-order reasons or wrong-makers) (2009: 101-2), and he notes that such bases might be able to contribute to explanations even if they are purely formal (2009: 117). If this is so, then my hedging clause fits the bill.

functions of Kantian principles of duty (see introduction, sect. 1.1). Their primary role is much more basic: qua constituents of concepts, they first enable us to think about and act in a shared objective world.

PART TWO

In the following chapters, I will argue that the Agent-Scope Reading of *Universality* can strengthen the position of Kantians in two contemporary debates: the debate concerning the so-called Problem of Relevant Descriptions (ch. 4) and the debate between Humean and Kantian Constructivists (ch. 5). In each case, I will begin the discussion without presupposing any of the conclusions reached in ch. 1-3, and show how Kantians run into problems without them. Then, towards the end, I will present the Agent-Scope Reading as a solution.

Chapter 4

A Kantian Standard of Moral Relevance?

Dissolving The Problem of Relevant Descriptions

1. The Problem of Relevant Descriptions

There is a persistent suspicion among Kant's critics that in applying the Formula of Universal Law we have to rely on moral determinations that are logically prior to the determinations that the formula itself can afford. The critics are worried that once it is settled which principle of action is at issue, various morally significant choices will already have been made: which goal to pursue and how to pursue it, which features of the situation to attend to and which to ignore, how to conceive of and how to characterize what one is doing. This perceived "belatedness" lies at the root of many different objections and counterexamples to Kant's ethics: the charge that it yields counterintuitive results, that it is rigoristic, anachronistic and out of touch with the complexities of the moral domain. Anscombe captures this common root when she argues that Kant's

rule about universalizable maxims is useless without stipulations as to what shall count as a relevant description of an action with a view to constructing a maxim about it.²⁰³

²⁰³ Anscombe 1958: 2. Anscombe levels the very same objection against utilitarianism, but utilitarians have mostly dismissed her challenge. On this point see e.g. Lyons 1965: 35.

Anscombe frames the worry in terms of discriminatory power.²⁰⁴ Without additional “stipulations”, the requirement expressed by the Formula of Universal Law seems to underdetermine the deontic status of actions because it is a *description-relative* requirement. It discriminates between obligatory, permissible and impermissible actions on the basis of the respective principles on which they are performed, and thus partly on the basis of how these actions are described. As a result, an action that is deemed permissible or obligatory under some descriptions may be deemed impermissible under others. Based on this observation, Anscombe concludes that Kant’s requirement is useless unless supplemented with stipulations that determine under which description an action *should be* assessed. Following Anscombe’s lead, O’Neill has labelled the problem of identifying such stipulations the “Problem of Relevant Descriptions” (O’Neill 2013: 61).²⁰⁵

Let us consider an example.²⁰⁶

Asylum₁: An immigration officer faces an asylum seeker who is at risk of being executed for engaging in banned homosexual activities in her home country. Since the host country does not recognize persecution of sexual minorities as grounds for asylum, the officer has to make a difficult decision. He decides to reject the application. We can describe what the officer has done in at least two ways. “He complied with the law” (that is, with the legal order of his nation state), we may say, or “He denied asylum to a victim of sexual persecution”.

On all widely accepted interpretations of Kant’s Formula of Universal Law (e.g. Rawls 2000: 162-177, Korsgaard 1996a: 77-105, O’Neill 1989: 81-104), it would seem that a principle of law compliance meets the requirement that it expresses, whereas a principle of denying asylum to victims of sexual persecution does not.²⁰⁷ After all, anyone could find themselves attracted to people of the same gender, so no one can will a law that leaves homosexuals to the discretion of homophobic despots.²⁰⁸

²⁰⁴ Below we will see that there are various other ways of framing it, e.g. in terms of a threat of subjectivism (sect. 2.3), relativism (sect. 4.1) or empty formalism (see ch. 3, sect. 3.3, fn. 192, 197, and 198).

²⁰⁵ For now, I won’t try to offer a more precise formulation of the Problem of Relevant Descriptions because this is my aim in sect. 2: to understand what the problem is, exactly.

²⁰⁶ I borrow this example, with some modifications, from Angelina Maccarone’s film *Unveiled*.

²⁰⁷ Note that, in this chapter, our primary concern is with maxims or principles of *action* and their universalisability, not with principles of *duty* and their universal validity, as it was before.

²⁰⁸ Let me elaborate on this brief argument by considering the case in light of Korsgaard’s reading of the Formula of Universal Law (1996a: 77-105): if the immigration officer’s maxim to deny asylum to victims of sexual persecution became a standard practice among officials aiming to comply with the

Anscombe's problem is this: under *one* description the officer seems to have discharged his obligation, under *another* he seems to have violated it. Which of them should we focus on? Without an answer to this question, Kant's universalisability requirement²⁰⁹ classifies the officer's action as both obligatory and impermissible.²¹⁰

My aim in this chapter is twofold. It is, first, to clarify what the Problem of Relevant Descriptions is and, second, to show that it is rooted in what we have called the Case-Scope Reading of Kant's conception of universal validity.²¹¹ This reading, I will argue, also precludes its solution. Let me outline the course of my argument. Although the problem has received considerable attention among Kantians, several misunderstandings have impeded its satisfactory treatment. I will therefore begin by examining and dismissing three conceptions of the problem found in the literature (sect. 2.1-2.4). I will dismiss these because I do not think that the problem *so* conceived casts genuine doubt on the discriminatory power of Kant's universalisability requirement. I will then go on to present a conception of the problem that does render it a real threat to the latter requirement, as it is usually understood (sect. 2.5). This conception is such that a solution to the problem would have to appeal to the notion

law, then everyone would have to expect to be denied shelter if they were persecuted for their sexual orientation. Just as anyone could find themselves in need of help, so anyone could become a victim of sexual persecution, e.g. if they discovered that they are attracted to people of their own gender or if some other preference of theirs became a subject of disdain. Since security is vital to our ability to pursue ends, it should be seen as a purpose "which belong[s] essentially to the will" (Korsgaard 1996a: 96). So if the maxim at issue were to become a universal law, then a purpose that is essential to our will would be in danger of being thwarted. Hence, we cannot *will* the maxim as a universal law.

²⁰⁹ By "universalisability requirement", I mean the requirement expressed by Kant's Formula of Universal Law: the requirement to "act only in accordance with that maxim through which you can at the same time will that it become a universal law" (G 4:421). For a clarification of the distinction between universality and universalisability, see ch. 3, sect. 3.2.

²¹⁰ One may object that this contradiction is merely apparent. If one believes that the relevant article of the asylum law is unjust and that officials ought to violate it, then one may want to say that a principle of law compliance cannot be willed as a universal law; whereas if one assumes that it is never permissible to break the law, then one may want to say that a principle of granting asylum to the persecuted cannot be willed as universal law, as there will be occasions where doing so would mean breaking the law. Either way, the universalisability requirement would not yield any contradictory verdicts. But this objection is question-begging, for this is exactly what we would expect the universalisability requirement to tell us: whether the circumstances at hand justify breaking the law. And yet it seems that it can only do so once we have settled how the case is to be described. On this point see also Herman 1993: 140.

²¹¹ The Case-Scope Reading was introduced in ch. 3, sect. 1 and discussed in ch. 3, sect. 2.5, among other places. There it was presented as a reading of the *Universality* desideratum. We can detach ourselves from the specific dialectic of that discussion by speaking of the Case-Scope Reading of Kant's *conception of universal validity* or, more precisely, Kant's conception of universal validity *in the moral domain*. For the sake of brevity, I will omit this last clause in what follows.

of a moral sensibility,²¹² which Kantians tend to disregard or actively resist. I will defend this conception against such suspicions by comparing the challenge it presents to that encountered by sensibility theorists (sect. 3).

This comparison will also further my second aim, since it will show that a satisfactory solution to the Problem of Relevant Descriptions would have to meet a special demand: it would have to supplement Kant's universalisability requirement with an account of moral sensibility that does *not* make that requirement itself superfluous. With this demand in mind, I will review two solutions to the problem that have been proposed in the literature (sect. 4). My discussion will bring to light that the demand cannot be met because there is no principle of moral relevance²¹³ that is not also a moral principle, which, as such, would replace or transform the universalisability requirement into an altogether different requirement. These observations will reveal the root of the problem and point towards a strategy for its *dissolution* (sect. 5). As I noted above, the problem is rooted in the Case-Scope Reading of Kant's conception of universal validity: the view that, for Kant, a principle is universally valid only if it is valid in all cases or circumstances. If this is how universal validity is understood, then Kant's universalisability requirement faces Anscombe's problem because it does not demand any sensitivity to the circumstances of action and indeed demands the opposite. As such, it stands in need of a supplementary principle of moral relevance. We can preempt this criticism, however, if we embrace what I have called the Agent-Scope Reading of Kant's conception of universal validity, for, according to this reading, the universalisability requirement *does* demand sensitivity to the circumstances. Note that, in this chapter, when I use the term "universalisability requirement" and related terms, I'm presupposing the Case-Scope Reading, unless otherwise indicated.

2. What Is the Problem of Relevant Descriptions? Three Misconceptions

2.1 First Misconception: A Puzzle Concerning Kant's Notion of Maxims

Sometimes the Problem of Relevant Descriptions is either unwittingly conflated with or explicitly reduced to a puzzle concerning Kant's notion of a maxim.²¹⁴ It is not

²¹² I use the term "moral sensibility" as it is used by Neo-Aristotelians in contemporary metaethics; it is *not* meant to refer to elements of Kant's ethics (e.g. the subjective conditions of receptiveness to the concept of duty, MM 6:399-403).

²¹³ I will use "principle of moral relevance" and "standard of moral relevance" interchangeably.

²¹⁴ While I will argue that Anscombe's objection cannot be reduced to a puzzle concerning Kant's

very surprising that there is a tendency to frame the problem in this way, given that Anscombe's objection seems to trade on a misunderstanding of this notion. Anscombe makes it sound as though any principle that an act *could* instantiate counts as a maxim. The Kantian will reply that a maxim is not just *any* principle. Kant defines maxims as subjective principles of volition (G 4:400, CPrR 5:19). This definition suggests that the relevance of a given description is determined by whether or not it reflects the agent's intentions. This is the reply that O'Neill seems to have in mind in the following passage:

Kant's universality test includes an explicit solution to the Problem of Relevant Descriptions. The principle ... whose universality it is relevant to test is the one on which the agent acts or proposes to act on a given occasion. This principle ... is the agent's *maxim*. (O'Neill 2013: 61)

As it stands, however, this reply is not sufficient. Kantians must flesh out what a maxim is, and that has proved difficult (see e.g. Timmermann 2000, Kitcher 2003). In particular, there has been an ongoing debate about whether maxims are general policies, perhaps even life rules, that range over a large number of actions and occasions (Bittner 1974: 488-9),²¹⁵ or whether they are specific principles that are tightly bound up with the circumstances of action. Anscombe's reflections on intentional descriptions remind us that there are many ways of describing an action that capture the agent's intentions – some more general, others more specific (Anscombe 1957: 37-47). Some Kantians assume that the infamous difficulty of determining the appropriate level of generality for maxims lies at the very heart of Anscombe's challenge, that the threat of conflicting verdicts that she perceives reflects an uncertainty regarding whether maxims express agents' final ends or their higher-order or lower-order means. (Suppose, for instance, that the immigration officer in the *Asylum*, example intended to deny asylum to a victim of sexual persecution *in order to* comply with the law). They therefore shift their attention from the Problem of Relevant Descriptions to the generality puzzle, taking the former to ultimately boil down to the latter (see Kitcher 2003: 217, O'Neill 1989: 84, 2013: 108-9, and Wood 1999: 102-5).

notion of a maxim, I think that some other objections and counterexamples to the Formula of Universal Law can be, and that critics were wrong in taking these to reveal issues with the formula itself. On this point see Kleingeld 2017: 113-4.

²¹⁵ This talk of general policies and life rules brings to mind Korsgaard's idea that having a practical identity is a matter of having standing commitments (see ch. 2, sect. 2.1, and ch. 3, sect. 1).

In what follows, I focus on O’Neill’s solutions to the generality puzzle because, although she does not fully reduce the Problem of Relevant Descriptions to this puzzle, tracing the development of her thoughts on the latter helps to get a sense of the debate as a whole. Initially, O’Neill argues that Kant’s universalisability requirement is a requirement to “assess *all* aspects and phases of what we do” (O’Neill 2013: 109), thus implying that actions are based on as many maxims as there are intentional descriptions of them. Later she comes to believe that this view is problematic (O’Neill 1989: 129) and therefore revises her position. Her revised view is that maxims are “underlying [or fundamental] intention[s] by which [agents] guide and control [their] more specific intentions”.²¹⁶ It seems to me that the latter view is more problematic, so I shall consider it first.

What distinguishes maxims-qua-underlying-intentions from specific intentions, in O’Neill’s view, is that only specific intentions make reference “to incidental aspects of the particular act and situation” (1989: 129), that is, to aspects whose bearing on the carrying out of the underlying intention depends on the circumstances. Since the suitability of means is always circumstance-dependent, this explanation suggests that maxims reflect commitments to ends pursued for their own sake. One may worry that this account of maxims is too narrow because it implies that the universalisability requirement is not “enough to determine the rightness or wrongness of particular acts”.²¹⁷ But even if we accept that the sole purpose of Kant’s requirement really is to assess the moral quality of our underlying intentions insofar as these indicate the moral quality of our wills, it is doubtful that it could serve even this limited purpose if the *only* descriptions maxims contained were of final ends.

Consider the following versions of the *Asylum*₁ case.

*Asylum*₂

Officer A, B and C pursue the end of being law-abiding officials for its own sake.

[A] Officer A thinks that, in standard circumstances, his end of being a law-abiding official is best served by complying with the asylum regulations, and he does not see the circumstances at hand as exceptional. So, for him, the fact that the refu-

²¹⁶ O’Neill 1989: 84. In arguing for this solution, O’Neill appeals to the fact that Kant uses the term “maxim”, which suggests that he is thinking of the first major premise of a poly-syllogism (see also Beck 1960: 81). My argument aims to show that the way in which O’Neill makes sense of this choice of terminology cannot be the right way.

²¹⁷ O’Neill 1989: 130. Allison argues that this view cannot make good sense of Kant’s examples (1990: 93-4).

gee appeals to invalid grounds for asylum makes rejecting the application a suitable means to his end.

[B] Officer B shares A's view of the default, but regards the circumstances at hand as exceptional. In his view, if a refugee appealing to invalid grounds is suffering, then, other things being equal, this shows a respect in which the asylum regulations fall short of the constitutional right of asylum. So, for him, the fact that the refugee is suffering defeats the bindingness of the asylum regulations and makes granting asylum a suitable means to his end.

[C] Officer C shares A's and B's views of the default, and also B's view of the defeating import of the fact that a refugee is suffering. And he does recognize that the refugee before him is suffering. However, in his view, if the suffering of a refugee appealing to invalid grounds is self-inflicted, then, other things being equal, it does not show a respect in which the asylum regulations fall short of the constitutional right of asylum. And as a homophobe, this is how he sees the persecution of sexual minorities. So, for him, the (alleged) fact that the refugee's suffering is self-inflicted restores the bindingness of the asylum regulations and makes rejecting the application a suitable means to his end.

For now I want to compare officers A and C. The difference between them is neither a matter of whether they do the right thing (after all, it is the same act) nor a matter of deliberating more or less carefully. It is a difference in how they understand *what it means* to be a law-abiding official, a difference that only comes into view once we look at counterfactual scenarios in which their judgments about the conduciveness of certain means to what, initially, appears like the same end diverge.²¹⁸ The problem with O'Neill's more recent solution to the generality puzzle is that it is simply impossible to determine what someone's underlying intentions are, and what these intentions say about the quality of their will, until we know how they would implement them under various kinds of circumstances.²¹⁹ As a result, we seem forced

²¹⁸ We made the same point in ch. 2, sect. 3.2, when discussing the ambiguity in how Herman speaks of grounds and ends. There we used the metaphor of ends as "mere labels" to indicate that, so conceived, ends do not suffice to reveal the moral character of our actions.

²¹⁹ In *Acting on Principle*, O'Neill seems to agree with this point. This is manifest in her response to counterexamples to the universalisability requirement that rest on overly detailed "maxims" (e.g. Allen Wood's Hildreth-Milton-Flitcraft example in his 1999: 102). Her response to such examples is that "a person cannot simply *claim* a highly specific maxim ... He must in fact intend his act to be contingent on those restrictions and not merely pursued by these means" (2013: 156). What she seems to have in mind is this. Knowing that the agent is willing to make a false promise (action) in order to secure

to turn back to her initial solution, according to which, for every action, there are as many maxims as there are intentional descriptions of it.

It seems to me that O'Neill abandons this initial solution to the generality puzzle because she worries that it not only fails to deal with the threat of conflicting verdicts but makes dealing with it a matter of conceptual impossibility. Recall that we wanted to know *which one* of the many intentions we should regard as the agent's maxim because we realized that picking one would yield one result, while picking another would yield a different result. In light of this, being told that *all* intentions are maxims does indeed seem to make things a lot worse. However, as Marcus Willaschek points out, we can block this problematic implication of O'Neill's initial solution by combining it with the following modified version of the universalisability requirement: an action is permissible if and only if all the maxims that it is based on satisfy the universalisability requirement.²²⁰ Although this combination of views is not free of ambiguities and problems,²²¹ it suffices to show what we wanted to show: that there are ways of dealing with the multiplicity of intentional descriptions and, thus, that conceiving of the Problem of Relevant Descriptions as a consequence of the generality puzzle renders it a non-issue for Kant's Formula of Universal Law. The foregoing discussion also draws attention to a point to which I will return in sect. 5: that a satisfactory reading of Kant's Formula of Universal Law and of his conception of maxims, taken together, must ensure that the agent's deliberative route *as a whole* is subjected to scrutiny through the universalisability requirement. Otherwise, the deontic assessment afforded by this requirement will be superficial, an assessment of ends as mere labels, as it is on O'Neill's revised account.

some extra money (underlying intention) is not enough to determine which maxim they are acting on. We need to know more about the nature of their pursuit, e.g. whether, if the promisee was not called Hildreth Milton Flitcraft, they would still see the false promise as a good means to their end, or whether their willingness to deceive Hildreth may reflect a lack of respect for this person, in particular.²²⁰ Willaschek 1992: 299. According to this view, our maxims come in hierarchies. I agree with this view and see it as a natural extension of the view that principles of duty come in hierarchies (see ch. 3, sect. 2.3). However, I do not think that the universalisability requirement demands that, in each case, we assess the whole hierarchy. Instead, I argue that one principle from this hierarchy is assessed *in light of the circumstances*.

²²¹ Let me mention one problem with the view that we must assess whole hierarchies of maxims. According to Willaschek, we have to take into account all our steps in the derivation of means from ends (vertical perspective), but he does not say whether the principles that specify the means should make reference to defeaters as well (horizontal perspective). Here, as before, the combination of the *Asymmetry* desideratum and the *Complexity* desideratum, discussed in ch. 2, sect. 1, 2.3.2, and 3.3.2, constrains our options. When discussing the *Murderer at the Door* case, for instance, Willaschek includes all the features that are intuitively relevant in the case at hand and leaves aside others that seem irrelevant (1992: 300), but he does not give an account of how this is justified and where to draw the line.

2.2 Second Misconception: The Opacity of the Human Heart

There is a further problem with what O'Neill, in the passage cited in sect. 2.1 above, calls Kant's "solution" to the Problem of Relevant Descriptions. While we know that the universalisability requirement applies to the agent's maxim,²²² we often do not know what someone's maxim on a given occasion really is. It is not only that other people can be wrong about our intentions, we ourselves are often deceived about them as well. Kant famously maintains that "the depths of the human heart are unfathomable" (MM 6:447) and that we have a tendency to "flatter ourselves by falsely attributing to ourselves a nobler motive" (G 4:407). This so-called "opacity thesis" (Ware 2009: 672) has inspired another reading of Anscombe's objection, according to which the Problem of Relevant Descriptions is the problem of determining which maxim a given action was truly based on – a task that proves difficult, even from the first-person perspective.²²³ Recall the above *Asylum*₂ case and immigration officer C, whose intention to reject the application was contingent on the applicant's being a homosexual. Had C convinced himself that his intention was based on a non-discriminatory consideration (on the impression that the applicant was not really suffering, say), then asking whether the corresponding maxim meets the universalisability requirement would not have been a reliable guide to the deontic status of his action.

If this is how we understand the Problem of Relevant Descriptions, then solving it cannot be a matter of further explicating Kant's Formula of Universal Law or the concepts employed therein. Lack of self-knowledge is not a theoretical problem. It is a practical problem that we face in our everyday lives. At the level of theory, we could at best supplement the Formula of Universal Law with a sincerity clause demanding that the tested principle express what the agent, after having carefully scrutinized his or her intentions, sincerely takes to be their content and deep structure (see e.g. Rawls 2000: 167-8, O'Neill 2013: 248). But, of course, such introspective scrutiny need not be crowned with success, and so it seems that the Problem of Relevant Descriptions, so conceived, is a real threat to a core element of Kant's ethics.

²²² Given what I said in the previous section, it might be more appropriate to speak of maxims in the plural, but I shall use the singular, partly because I do not accept Willaschek's account and partly for the sake of simplicity.

²²³ O'Neill indicates that she sees this as part of the problem when she recommends trying to avoid self-deception as a method for selecting relevant descriptions. See O'Neill 2013: 225-27, 248-49.

It is indeed a threat if one assumes that the Formula of Universal Law is primarily meant to provide a decision tool that agents can use to navigate their way around the moral domain. It is not a threat, however, if its primary role is to articulate a norm that determines the deontic status of actions.²²⁴ For the universalisability requirement to be a reliable decision tool, the agent would have to know which maxim will underlie the future action that they are considering performing, and that is something they may not be able to know for sure.²²⁵ But this epistemic limitation is not an issue if the main point of the universalisability requirement is to determine the deontic status of actions on the basis of the maxims on which they were, *in fact*, performed. It is of no importance to the discriminatory power of the universalisability requirement whether anyone has epistemic access to these facts. Kant's conception of his own project as an exercise in the metaphysics of morals favours this latter interpretation. To conclude, if the Problem of Relevant Descriptions is taken to arise from the opacity thesis, it casts doubt on the reliability of Kant's universalisability requirement as a decision tool, but it does not undermine its discriminatory power as a norm that determines the deontic status of actions, which is what it is primarily intended to be.²²⁶

2.3 Upshot: Is There a Problem Left to Solve? The Subjectivizing Move

The two approaches to the Problem of Relevant Descriptions that I have discussed so far rely on the thought that O'Neill articulates in the passage cited in sect. 2.1 above, namely, that Kant himself offers a solution to the problem by pointing us to the agent's maxim and to the description it contains. The challenges that remain, according to these approaches, stem from the obstacles encountered in trying to de-

²²⁴ This question, the question of whether the Categorical Imperative is primarily a decision tool or a norm that determines the deontic status of actions, already came up. We discussed it in ch. 2, sect. 3.2.

²²⁵ In any case, it is far from uncontroversial that the opacity thesis denies that agents can know their maxims. Some commentators argue that it is absurd to deny that we can know our specific intentions, and thus they conclude that Kant must have meant the fundamental maxim (or *Gesinnung*). See e.g. Allison 1990: 93.

²²⁶ The tendency to conflate these two ways of viewing the Formula of Universal Law stems, in part, from a confusion about constructivist approaches in ethics and metaethics. Constructivists argue that a practical judgment is correct insofar as it results from a certain "construction procedure". This talk of a procedure should not be taken to refer to a decision procedure or tool that guarantees success, however. For constructivists, this is merely a metaphor, a way of illustrating fundamental principles of practical reason. In using this metaphor, they do not mean to say that any judgment an agent might arrive at when sincerely employing that procedure is correct. What they mean is rather that the fundamental principles of practical reason, which their procedure illustrates, define what it is to properly exercise one's capacity of practical reason – an ideal that any agent can fall short of, no matter how hard they try. We will come back to this in ch. 5.

termine what the agent's maxim *is*, both generally speaking (first misconception) and in specific instances (second misconception). I have argued that these obstacles do not in fact threaten the discriminatory power of Kant's universalisability requirement. Does that mean that there is no problem left to solve? I shall now argue that it does not.

First, however, an exegetical point. If we pay close attention to Anscombe's phrasing, we can see that the two conceptions of the problem discussed above fail to make good sense of what she says. According to her, the Problem of Relevant Descriptions arises "with a view to constructing a maxim". This suggests, first, that she is well aware of what O'Neill presents as a solution to her problem: the fact that Kant was not interested in descriptions other than the one contained in the agent's maxim. Second, and more importantly, her use of the phrase "constructing a maxim" indicates from which perspective she takes the problem to arise: the perspective of a *morally deliberating agent* trying to formulate a principle of action.²²⁷ The perspective that she has in mind is neither the perspective of a commentator who is fiddling with Kant's notion of a maxim, nor the perspective of a self-scrutinizing agent who is trying to get to the bottom of their "unfathomable heart". It would be odd to describe what the latter two are doing as "constructing maxims". This indicates that Anscombe is not actually worried that Kant's theory might lack the resources to determine which description it is that an agent's maxim contains, but rather that the description that the maxim does in fact contain could be irrelevant and that Kant's theory lacks the resources to rule out *this* possibility. Some Kant scholars deny this possibility.²²⁸ I will argue that they do so at too high a cost, but first let me illustrate what it is that they are denying. Consider the following example.

Grandmother: You promised to help your grandmother on Sunday afternoon, but when the time arrives, you just stay on your favourite bench in the park, playing the guitar. "Why did you stay in the park?" your friend asks perplexedly. "I was practicing in order to develop my guitar skills," you say. In saying this, you are

²²⁷ It must be admitted that the evidence cited here is not conclusive. The fact that Anscombe speaks of "constructing a maxim *about it*", for example, might be taken to suggest that she is thinking of an observer describing an action that has already been performed from a third-person perspective. The point that I go on to make is not affected by this: the two approaches discussed thus far make even less sense of her choice of words.

²²⁸ Wood's criticism of the Formula of Universal Law, for instance, rests on denying this possibility (1999: 102-7).

being honest, you are not keeping anything from your friend, and you are not deceived about your intentions.²²⁹

There is certainly a problem with your maxim, but it is not that it cannot be willed as a universal law. It can (or so it seems). The problem is that, given the circumstances, the considerations from which you are acting are irrelevant, or at least they are not the only considerations that are relevant to what you are allowed or obligated to do in this case. The opponent I am envisaging would say that this intuition is misleading. If, in forming your maxim, you complied with an attentiveness clause akin to the above sincerity clause,²³⁰ then you simply “can do nothing more to determine that [your] maxim does not match [your] situation” (O’Neill 2013: 249). According to my opponent, your action is right in the only sense that is available to Kantians: *subjectively right*.²³¹ If you had acted under a different description, e.g. from a maxim of promise-breaking, then your action would have been wrong. But you did not, and so it is right.²³²

Such a claim will strike many people as deeply un-Kantian, but this is the cost of denying the possibility that even an attentively formed maxim might contain an irrelevant description: one must deny that Kant’s ethics can deliver on its promise of objectivity and must settle for subjective rightness instead. If this is indeed the cost of putting Anscombe’s objection to one side, it is surely a cost that Kantians should refuse to pay. Instead, it seems, they should admit not only that the intentions with which agents act may be morally unworthy (this possibility is already acknowledged in the very idea of a Categorical Imperative), but also that such shortcomings may be rooted in a mismatch between maxim and circumstances, and not necessarily in a failure to meet the universalisability requirement. In acknowledging the possibility of

²²⁹ This is an example that Stratton-Lake uses to illustrate his objection to what he labels the “justificatory conception of the moral law”. On this conception, the “lawlike nature of [one’s] maxim is sufficient to make [the] action morally right” (Stratton-Lake 2000: 58). I will say more about the justificatory conception and its flaws in sect. 3 below, but the fact that my behaviour in this example seems so oddly unmotivated (Have I forgotten? Do I properly understand what promising is? Am I weak-willed? Etc.) is part of what Stratton-Lake takes issue with. For current purposes, I want to leave it open *why* I ignore the promise and postpone the discussion of this question to sect. 2.5 and 3.

²³⁰ According to O’Neill, the deliberating agent must not only try to avoid self-deception, she must also do her best to avoid ignorance and bias (2013: 249). That’s what I mean by “attentiveness clause”.

²³¹ Even though O’Neill endorses this subjectivizing move (2013: 251-3), her view is more nuanced than the view of the opponent I envisage in this section, because she does not deny that descriptions can be irrelevant.

²³² For simplicity’s sake, I am here adopting O’Neill’s way of speaking, so I speak of “right” and “wrong” actions, rather than “permissible”, “impermissible”, and “obligatory” actions, as I have done so far. I will do the same in sect. 4.2, when I discuss Mark Timmons’ solution.

such mismatch-based shortcomings, we make room for a different approach to Anscombe's challenge. According to this new approach, she is not asking for stipulations that would determine which description a maxim *does* contain (or what the agent's maxim *is*), but rather which description it *ought to* contain, given the circumstances.²³³ If such stipulations can be found, then there is hope that Kant can make good on his promise of objectivity and that his ethics can overcome Anscombe's challenge. However, even if it is accepted that this is the point of her challenge, there is still room for a further misconception.

2.4 Third Misconception: Inadvertent Ignorance

The *reductio argument* against the view that "there is no problem left to solve" that I presented in the previous section invites a response. My opponent may grant that subjectivizing Kant's ethics is an inadequate way of dealing with the problem, but still maintain that no additional stipulations are needed. All that is needed to solve the problem, they may argue, is a conditional restatement of the Formula of Universal Law, according to which the latter determines the deontic status of an action if and only if the action is based on a maxim that matches the circumstances. This proposal assumes that the Problem of Relevant Descriptions arises from the possibility of *epistemic* failure in the contemporary sense of the term. In what follows, I argue that this conception of the problem overlooks an important distinction between two ways in which one may fail to choose a relevant description, thus making it impossible to account for the core instances of such failure.

We find an example of such a conflation in O'Neill's early work. In sect. 2.3 above, I argued against those who deny the possibility of acting under irrelevant descriptions altogether. O'Neill admits this possibility, but, as a proponent of the subjectivizing move, she treats it as deontically insignificant.²³⁴ Let us consider how O'Neill explains the occurrence of actions performed under descriptions that are

²³³ Recall that this is Herman's objection to the view that the Categorical Imperative is an algorithm for moral deliberation: that, if it was, there would be no way of ensuring that the resulting judgments take account of all and only the morally relevant features of the case. For this reason, she aims to develop an account of moral deliberation on which the classification of features as morally relevant is regulated. See ch. 2, sect. 3.1.

²³⁴ On this point see sect. 2.3 and O'Neill 2013: 230. O'Neill does not distinguish between approaches that take Anscombe to be worried about what an agent's maxim *is* and those that take her to be worried about what it *ought to be*; instead she treats the two worries as different aspects of the problem. For the former aspect see, in particular, 2013: 106-10 and 1989: 83-9; for the latter 2013: 223-32 and 246-77.

irrelevant from the “bird’s-eye point of view” (O’Neill 2013: 257). She lists three types of error that can afflict agents when they are forming their maxims of action: self-deception, ignorance, and bias (O’Neill 2013: 225-30). In sect. 2.2, I showed that the first type is not a genuine instance of the problem at all,²³⁵ so we can focus on the latter two. O’Neill treats them both as corollaries of the pervasive epistemic limitations that affect human agents in decision situations. In doing so, however, she neglects a key difference between them, which bears on the prospects of resolving disagreements about whether a description is relevant or not.

Recall the *Grandmother* example. If you stayed on the bench playing the guitar, thereby ignoring a promise that you made to your grandmother, it would be rather odd for you to reply to your friend’s why-question in the way you did (“I was practicing in order to develop my guitar skills”) unless the promise had completely slipped your mind. Let us assume that this was the case. Now imagine your friend were to point out to you that you had made that promise. Naturally you would spring to your feet and get going. In instances of inadvertent ignorance, instances where an agent fails to remember or notice something or where they draw a false conclusion, disagreements about the relevance of their description are usually easy to resolve. There is, at any rate, a truth of the matter, visible from the bird’s-eye point of view, on which the parties to the disagreement can converge.²³⁶ Now recall the *Asylum*₂ cases. Of officer A we cannot say that he inadvertently ignored the fact that he was facing a distressed victim of sexual persecution (they spoke face to face, he could not have deemed the ground she appealed to invalid if he wasn’t aware of it, etc.). If the description under which he acted seems irrelevant to us, then this is because we regard the above fact as morally relevant and he does not. We may want to accuse him of being biased, but how could we justify this accusation? How would we go about resolving such a disagreement? An advocate of Kant’s Formula of Universal Law seems forced to say that *this* kind of disagreement, unlike the first one, is potentially intractable, because there is no standard of moral relevance prior to the universalisa-

²³⁵ Self-deception, in the sense emphasized by Kant, is usually taken to lead to a mismatch between the maxim one is actually acting on (perhaps the *fundamental* maxim one is acting on, see fn. 225) and the maxim that one takes oneself to be acting on, not a mismatch between maxim and circumstances. Accordingly, self-deception and the errors that are pertinent here operate on different levels. This is not to deny that they can go hand in hand. In fact, the former often helps to sustain the latter.

²³⁶ The assumption that in cases of ideally informed factual disagreement normal methods of rational discourse will persuade the mistaken party is not beyond doubt. On this point see Cassam 2015.

bility requirement (no “stipulations”) that the parties to the disagreement could appeal to, *even if* they were to take up the bird’s-eye viewpoint.²³⁷ There seems to be no objectively relevant description that represents the situation as it really is, morally speaking, and thus no possibility of getting it wrong.

Instead of properly appreciating the difference between these two varieties of error, O’Neill contradicts herself by claiming both that there is a meaningful distinction between relevant and irrelevant descriptions and that “Kant’s theory of right does not provide any method for determining the relevant ... act description under which to assess an act when we take the bird’s-eye point of view” (O’Neill 2013: 257). What she should have said is that there is a distinction between epistemic failure and success, between getting the facts right and getting them wrong, but that, for a Kantian, there is no (or does not seem to be)²³⁸ any meaningful distinction between correct and incorrect claims concerning the moral relevance of facts. But given that some mismatches between maxims and circumstances do result from epistemic failure, one may wonder why the conception that we considered here is a misconception rather than a partially correct conception.

2.5 The Correct Conception: A Flawed Moral Sensibility

In order to answer this question, let us recall what we have established so far. It emerged from our discussion that Anscombe rejects Kant’s ethics because, in her view, his universalisability requirement misclassifies actions based on maxims that contain irrelevant descriptions. We now have to ask how much force this objection would carry if it targeted cases where the mismatch between maxim and circumstances arose from an inadvertent epistemic failure. The answer is “not much”. After all, it is not the primary task of a *moral* theory or its supreme principle to rule out epistemic errors. Such errors do occur, and we admitted that they can lead to wrongdoing (e.g. to unkept promises), but they are at best a subsidiary matter in the context of moral theorizing. Their marginal importance is manifest in the fact that we tend not to *blame* people for committing such errors. Recall the *Grandmother* example: if you had simply forgotten, we would not blame you or at least not fully, because we do not see forgetfulness as a moral failure. This suggests that actions from epistemi-

²³⁷ O’Neill concedes that such disagreements are intractable (2013: 254-5).

²³⁸ In sect. 5, I will argue that Kantians *can* draw the desired distinction, but that they can do so only if they abandon the Case-Scope Reading of Kant’s conception of universal validity.

cally flawed maxims are merely *peripheral* instances of the class of actions that exemplify Anscombe's problem. Its *core* instances are actions from biased maxims. This is why an epistemic success condition (sect. 2.4) cannot dissolve the threat that the problem presents, why it cannot replace the stipulations that Anscombe is after, namely, an objective standard of moral relevance.

To grasp what we are looking for, we must elaborate on the notion of bias that we adopted from O'Neill. In the context of the present discussion, the notion should be read quite broadly, as referring to all the different ways in which one may misapprehend the moral relevance of a circumstantial feature that one is otherwise aware of. Biases are not singular instances of such misapprehension, however. They are objectionable dispositions, such as misogyny, homophobia, or racism, that produce corresponding patterns of behaviour.²³⁹ Going back to our *Asylum*₂ cases, we can imagine different scenarios: that officer A acted under an irrelevant description because he is particularly submissive to authority or because he is generally insensitive to human suffering, say. In contemporary metaethics, discussions of biases and objectionable dispositions are oftentimes couched in terms of a flawed moral sensibility. We could therefore reformulate the task associated with Anscombe's challenge by saying that Kant needs an account of moral sensibility.

Before moving on to some more specific worries about this task, let me forestall a general objection. The search for an objective standard of moral relevance may appear as un-Kantian as the subjectivizing move (see sect. 2.3) because it presupposes that there is moral content prior to the universalisability requirement. I agree that positing such content would go against some of Kant's most basic commitments, and this is precisely why, in the end, I argue that the task associated with the Problem of Relevant Descriptions cannot be discharged. Nonetheless, I think that an objective standard of moral relevance is what *would be* needed to solve the problem, given the strongest interpretation of it.

²³⁹ For further remarks on biases and objectionable dispositions see ch. 2, sect. 3.3.1 and ch. 3, sect. 2.4, where I associate the notion of bias with the notion of making exceptions for oneself.

3. Does Kant Need a Sensibility Theory? Clarifying the Task

The notion of a moral sensibility is familiar from a number of Neo-Aristotelian views known as sensibility theories.²⁴⁰ My claim that this notion is key to developing a consistent reading of *Kant's* ethics will thus be met with suspicion. These suspicions need to be addressed, not only in order to show that my preferred conception of the Problem of Relevant Descriptions is indeed a threat to Kant's universalisability requirement, but also because, in doing so, we will identify a demand that an adequate solution to the problem would have to meet. In sect. 1, I observed that the universalisability requirement is susceptible to Anscombe's criticism because it is *description-relative*. In order to dismiss the suspicions at hand, we need to sharpen this notion in two respects.

The first suspicion is rather superficial. It may seem that, by using the term "description", Anscombe has misled us from the start. A sensibility theorist, who thinks of moral reasoning in perceptual terms, might say that such reasoning aims at appropriate descriptions, for example, at a description of the applicant in the *Asylum* cases as someone who is desperately seeking shelter rather than as someone who is committing asylum fraud. But to ascribe such a claim to *Kant* seems inapt, because to do so would be to neglect his distinction between practical and theoretical reasoning. Open-ended reasoning towards a description of a given object, situation, or act, for Kant, is an exercise of reflective, that is, theoretical judgment. However, as O'Neill points out in a more recent piece,²⁴¹

in *practical* judging we are not judging a particular act. The task in practical judgment is to shape action *that is not yet done*. There is no particular act [or situation] to be judged. The aim of practical judgment is to shape the world (in small part), not to identify some way in which the world is shaped. ... The different direction of fit shields practical judgment from the problem of relevant descriptions.²⁴²

O'Neill's emphasis on the distinction between practical and theoretical judgment is true to Kant's text, of course, but her attempt to use it to dismiss the Problem of

²⁴⁰ See Wiggins 1998: 87-138, 185-214, McDowell 1998: 131-150, 77-94, and (less obviously) Murdoch 2001: 1-44. In recent discussions, the label "fitting-attitude" theories has become more fashionable.

²⁴¹ We already mentioned a closely related worry of hers when discussing the Casuistical Questions Response to the *Rigorism Objection* in ch. 1, sect. 4.1. My response there was similar to the response I give here.

²⁴² O'Neill 2004: 312-3. She reaffirms these doubts about the Problem of Relevant Descriptions in a new introduction to *Acting on Principle*. O'Neill 2013: 10-2.

Relevant Descriptions rests on an ambiguity in how we use the term “description”. In its *narrow* sense, the term is indeed associated with the mind-to-world direction of fit, making it inapplicable to practical judgments or “prescriptions”. But we also use the term in a *broad* sense to refer to any form of conceptualization. With this disambiguation in mind, we can concede the terminological point, namely, that Kant’s universalisability requirement is description-relative only in the broad sense, while holding on to the substantive point: that there is a relativity in it, which, for lack of a better term, we could call “prescription-relativity”. Is there a Problem of Relevant Prescriptions, then? Prima facie, we have no reason to assume otherwise, at least not on the basis of O’Neill’s observations about directions of fit. For we still need to know how to get from the Formula of Universal Law to a specific course of action (a specific way of “shaping the world”) that constitutes a morally appropriate response to a given situation, other than by relying on our moral sensibility, for which Kant apparently failed to set a standard.²⁴³ But here a second, more weighty suspicion about my appeal to the notion of a moral sensibility rears its head.

Unlike sensibility theorists, Kant does not think that an action’s deontic status is determined by whether it constitutes an appropriate response to the circumstances (that, after all, is the task that he assigns to the Categorical Imperative). So it would be quite surprising if he thought of moral reasoning in terms of exercising one’s moral sensibility. Let us grant, for now, that this is correct. But a weaker claim might be true nonetheless: an action’s deontic status might *depend on* whether the action constitutes an appropriate response to the circumstances or not, so that the Categorical Imperative, at least in its first formulation, would need to be supplemented with an account of moral sensibility. To evaluate this dependence claim, we must examine what it would mean to deny it. There are two options.

A radical denial of the dependence claim would amount to the view that Kant’s universalisability requirement is practical reason’s *only* guide, both necessary and sufficient to determine what we ought to do. It is obvious that this view has implausible implications: if it were true, we would always be bound by a single overarching imperative with disjunctive content that would prescribe acting on *any one* of infinitely

²⁴³ In fact, O’Neill goes on to argue that there is no Problem of Relevant Prescriptions either, because the appearance of indeterminacy in practical reasoning results from an artificial focus on one principle of action at a time. However, her argument for this conclusion conflates hypothetical and categorical imperatives. O’Neill 2004: 313-4.

many maxims that meet the universalisability requirement, no matter what the circumstances.²⁴⁴ Many of the actions that would then be open to us in any given situation would be outrageous or instrumentally pointless or both. This observation merely confirms what we knew all along: that the universalisability requirement cannot do all the work by itself or that categorical imperatives are not the only imperatives there are. It also points towards a weaker, more plausible denial of the dependence claim, to a view that we can label the Naive Stance on the Problem of Relevant Descriptions.

Naive Stance: Kant's universalisability requirement applies to whatever principles of action instrumental reason proposes. Of course, when we reason instrumentally, we need to be responsive to the circumstances, but only insofar as circumstances bear on which means we have to take to accomplish our ends. The responsiveness involved is theoretical, not moral; it does not require a moral sensibility.

To assess the *Naive Stance* we need to go back to the third misconception. Recall that, in sect. 2.4, we drew attention to mismatches between maxims and circumstances that defy explanation in terms of epistemic malfunctioning. The inadequacy of the *Naive Stance* shows up in the fact that it cannot accommodate this possibility because it casts practical reason's answerability to the circumstances in purely theoretical terms.²⁴⁵ Or, if it purports to accommodate it, then it effectively sidesteps the problematic aspect of Kant's theory: the universalisability requirement. Let me explain why this is so.

Consider what proponents of the *Naive Stance* would have to say about the *Asylum₂* cases. Officer A, we said, rejects the application simply because the grounds appealed to are invalid; officer B grants asylum to the victim of sexual persecution because he takes her suffering to make a difference to what he is allowed or obligated to do. Both are fully aware of the facts, rely on sound theoretical judgments, and act on maxims that satisfy the universalisability requirement. If *Naive Stance* theorists want to avoid saying that both decisions are right (on this subjectivizing move, see

²⁴⁴ This is Stratton-Lake's *reductio argument* against the justificatory conception of the moral law (see fn. 229).

²⁴⁵ In this respect, it is similar to Timmermann's account of the relation between moral and instrumental reasoning, which we discussed in ch. 1, sect. 3.2.

sect. 2.3), they have to say that only one of the officers *has adopted the right final ends*.²⁴⁶ This takes us back to the discussion of O'Neill's maxims-as-underlying-intentions account in sect. 2.1. This account of O'Neill's does not come out of nowhere. There are well-known passages in Kant's writings where he himself appeals to a system of obligatory ends, which is organised around the end of self-perfection and the end of promoting other people's happiness. In light of the results of our discussion in sect. 2.1 above, we can rule out one way of reading this appeal: if we want to account for the difference between officers A and B, we have to do more than insist that one of them failed to make other people's happiness his end, for the difference may lie in how they understand what that entails. So when *Naive Stance* theorists appeal to the thought that furthering other people's happiness is an obligatory end, they must be appealing to a conception of that end on which it is more than a label, on which it has a deep structure: an end that contains within it an understanding of how to balance various concerns in various sorts of circumstances.²⁴⁷

But if appeals to obligatory ends take this form, they cannot be part of an argument that is supposed to rescue the Formula of Universal Law from Anscombe's objection, because this argument would render the very object it purports to rescue superfluous. Instead of showing that, when taken together, the universalisability requirement and the deliverances of instrumental reason secure an adequate sensitivity to the circumstances, such an argument would just replace the universalisability requirement with a different requirement:²⁴⁸ the requirement to adopt certain ends, namely those that contain an appropriate understanding of what we are allowed or obligated to do under various circumstances.²⁴⁹ Does a response to Anscombe really require doing the former? Let us go back to the end of the previous section. When I said that Kantians need an account of moral sensibility, I was drawing on the forego-

²⁴⁶ For Kant's notion of an end that is also a duty see, in particular, MM 6:382-88 and ch. 1, sect. 4.1 above. It is a matter of controversy how to reconcile these passages with some of the more formalistic sounding passages that are prominent in the *Groundwork* and the *Critique of Practical Reason*.

²⁴⁷ This is how Kant's talk of obligatory ends has to be understood. These are ends prescribed by duties of virtue, and their pursuit is *practically necessary* (ch. 1, sect. 4.1), so, in doing wrong, one is not pursuing them.

²⁴⁸ If the tension mentioned in fn. 246 is merely apparent, then it is not, strictly speaking, an alternative requirement, but rather an alternative perspective on the very same requirement. But if they *do* express the same requirement, then one of them should not, as Anscombe puts it, be "useless" without the other.

²⁴⁹ This is the problem with Herman's recent attempt (2013) to show that Kant's ethics can accommodate the role of circumstances. Her appeal to both obligatory ends *and* the Categorical Imperative amounts to a problematic duplication. See ch. 2, sect. 3.2, fn. 115.

ing discussion: they need one, *if* they want to respond to Anscombe's objection, which concerns the Formula of Universal Law. It is thus beside the point that some *other* requirement that we can find in Kant's oeuvre can get by without an appeal to moral sensibility. Keeping this in mind, we can finally conclude that the dependence claim survives both attacks.

Ironically, it takes the breakdown of the *Naive Stance* to remind us of the kernel of truth in the second suspicion, the suspicion that we are mistaking Kant for a sensibility theorist. If we are to solve the Problem of Relevant Descriptions, we cannot make the even stronger claim that, for Kant, an action's deontic status is *determined by* whether it constitutes an appropriate response to the circumstances because, in doing so, we would throw out the baby with the bathwater. There may be nothing wrong with this position as such. In fact, it is the position of sensibility theorists proper, for whom "seeing" the circumstances in a certain light, conceiving of them as requiring some course of action, and acting accordingly is *all* that morality demands.²⁵⁰ "If I attend properly," Iris Murdoch says, "I will have no choices and this is the ultimate condition to be aimed at" (2001: 38). The idea that there is a further requirement, for instance, a universalisability requirement, is, in her view, a dangerous idea, both in philosophy and in life (Murdoch 2001: 15-6, 37-8). For sensibility theorists, acting under a relevant description is necessary *as well as* sufficient for moral conduct, which seems to make their task quite different from Kant's.²⁵¹ To bring this difference into focus, it will be useful to draw a further distinction between two senses in which a theory can be description-relative.²⁵² We shall say that sensibility theories and Kant's theory have in common that they are description-relative in the *general* sense of conceiving the deontic status of actions as dependent on the descriptions under which these actions are performed.²⁵³ But, within Kant's theory, this general description-

²⁵⁰ McDowell, for example, conceives of moral conduct as a matter of seeing the circumstances in the light in which a properly brought up person would see them (1998: 100-1). For Murdoch, it is a matter of attending to the infinite complexities of human reality and coming to see things and persons justly and lovingly (2001: 33). For both the quasi-visual experience at issue is conceptual.

²⁵¹ In sect. 5, we will see that this is actually a mistaken impression.

²⁵² I borrow the following terminology from Timmons, who distinguishes between a general and a special problem of moral relevance (1997: 402).

²⁵³ It has to be admitted that this is a somewhat unusual take on sensibility theories, so let me say a bit more. A sensibility theorist would take the deontic status of an action to be determined by whether the description under which it is performed is a description under which a virtuous person might have acted in this situation, and, if so, by whether she would have considered it necessary or merely possible to act in this way. I think that this correctly captures the view, if we assume, in addition, that the virtuous person is not prone to epistemic malfunctioning.

relativity takes a *special* form: the deontic status of actions is taken to depend on the satisfaction of a distinct requirement, whose satisfaction, in turn, depends on the descriptions under which these actions are performed. What Kant needs in order to deal with the Problem of Relevant Descriptions, then, is not a sensibility theory – or so it seems. What he needs is an account of moral sensibility that does not make the universalisability requirement *superfluous*. Let us call this the “non-superfluous demand”. In the remainder of this chapter, I argue that there is no account that meets this demand. On this basis, I conclude that we must regard the Formula of Universal Law *itself* as an expression of Kant’s account of moral sensibility.

4. A Kantian Standard of Moral Relevance? Two Unsuccessful Solutions

4.1 First Solution: Culturally Imparted Rules of Moral Salience

Let us consider Herman’s attempt to accomplish the above task. Although she does not address the Problem of Relevant Descriptions explicitly, her account of rules of moral salience can be understood as a solution to it. Here is her starting point.

In order to use the CI as a principle of judgment or assessment, the agent must first produce his maxim. ... If actions are to be assessed (directly) by the CI through their maxims, there is a need in the Kantian system for some kind of independent moral knowledge. An agent who came to the CI procedure with no knowledge of the moral characteristics of actions would be very unlikely to describe his action in a morally appropriate way. (Herman 1993:75)

Herman argues that Kant responds to this need by presenting “moral judgment as an activity with a customary context of occurrence”.²⁵⁴ She invites us to think of the structure of our moral sensibility in terms of so-called rules of moral salience that communities impart to their children through moral education. These rules, once internalized, make us perceive a world with moral features and help us “pick out those elements of [our] circumstances or ... proposed actions that require moral attention”.²⁵⁵

²⁵⁴ Herman 1993: 76. She cites Kant’s account of moral education (MM 6:477-484) as evidence.

²⁵⁵ Herman 1993: 77. This account of how features of the circumstances are picked out as relevant is slightly different from the account that Herman puts forward in “Moral Deliberation and the Derivation of Duties”, which we discussed in ch. 2, sect. 3. She comments on the relation between the two in her 1993: 151.

In developing this solution, Herman is attentive to the non-superfluous demand. She stresses that rules of moral salience should not be confused with rules of pro tanto duty, which assign independent deontic import or moral weight to the features they pick out. Rules of pro tanto duty fully determine the deontic status of actions in accordance with the balance of these features' import or weight, and make no appeal to whether the maxims on which these actions are based meet the universalisability requirement. So a theory of pro tanto duties would render the latter requirement superfluous. But, of course, Herman's rules of moral salience do influence and shape our moral judgments. And this causes trouble for Herman's approach.

The rules that we pass on through moral education represent the ethical self-understanding of our community. It goes without saying that different communities (cultural, subcultural, religious, generational) vary in this self-understanding, creating differences in how members of these communities "see" and describe their circumstances, what they take to be appropriate in a given case, and which maxims they form. As Herman admits, these differences "raise[] the spectre of relativism" (1993: 89), a close relative of the subjectivism that we encountered in sect. 2.3. For if the members of different communities have a dispute about what one is allowed or obligated to do in a given situation, they may not be able to settle this dispute by appeal to the Formula of Universal Law. Once again, determinacy comes at the cost of objectivity.

However, unlike O'Neill, Herman refuses to settle for this un-Kantian conclusion. Rules of moral salience, she adds, are objectively better or worse, depending on how well they accord with Kant's conception of the person as a moral agent or end-in-themselves (Herman 1993: 86). It is thus "a *practical* task for a community of moral agents" (Herman 1993: 87) to continually revise their defeasible rules in light of this conception.²⁵⁶ But how does this collective responsibility bear on the assessment of individual actions, which is the subject matter of Anscombe's criticism? Herman's remarks on this point are inconclusive. On the one hand, she seems to embrace relativism on the individual level, saying that "moral fault ... occurs only when an agent's maxim contains a principle that cannot be willed as a universal law" (Herman 1993:

²⁵⁶ Here defeasibility is understood as fallibility. This is the conception of defeasibility that we criticized when discussing Korsgaard's appeal to provisional universality. However, in this context, it is appropriate, because we do want to say that having to revise our rules of moral salience would mean that we got it wrong before. See ch. 2, sect. 4.

89), and not when it is based on rules of moral salience that fall short of Kant's conception of the person. On the other hand, she suggests that there are some cases in which individuals *can* be criticized for not having discovered the shortcomings in their rules of moral salience (Herman 1993: 90-2). Nazis are such a case, she believes. Let us ask whether going one way or the other gets us what we need to solve the Problem of Relevant Descriptions.²⁵⁷

We have already established that the relativistic route does not lead to a satisfactory solution. Think of the *Asylum*₂ cases and imagine that the prosecutor who ordered the persecution of the refugee in her home country was asked to pass judgment on officer B's decision to grant asylum. Presumably, he would say that the latter violated his duty because, having adopted the homophobic outlook of his community, the prosecutor would not regard a gay person's suffering as morally relevant (in this respect, he would be similar to officer C). In sect. 2.4, we saw that a disagreement of this kind could not be settled by appeal to the Formula of Universal Law. Prima facie, Herman's alternative route does offer a way out. The appeal to Kant's conception of the person allows us to judge officer B's rules of moral salience to be superior to those of his opponent because they pay heed to the fact that the refugee is a moral agent. This proposal for a solution is quite similar to Mark Timmons' proposal, which is the topic of the next section. We will find that it fails to satisfy the non-superfluous demand, albeit in a less obvious manner than a theory of pro tanto duties.

4.2 Second Solution: The Formula of Humanity

Timmons maintains that there is a "pretty obvious solution" (1997: 399) to the Problem of Relevant Descriptions in Kant's work. That it has been overlooked, he conjectures, is due to the fact that critics and commentators alike have disregarded the different roles assigned to the formulas of the Categorical Imperative in the architec-

²⁵⁷ Perhaps Herman is thinking that we do not have to choose, that the question whether an agent is to be blamed for relying on flawed rules of moral salience is to be decided on a case by case basis. I think that this is actually the most accurate account of our practice of praising and blaming. But the fact that someone ought not to be blamed does not mean that they did not do anything impermissible. So, as far as the Problem of Relevant Descriptions is concerned, this middle ground is tantamount to saying that the deontic status of actions does, in fact, depend on whether the rules of moral salience relied on are in accord with Kant's conception of the person.

tonic of Kant's system.²⁵⁸ The difference between these roles, he says, becomes apparent when we reflect on the two main aims of moral theorizing: the practical aim of providing agents with a decision tool (or "procedure", as he calls it) and the theoretical aim of finding a criterion of right that systematizes the fundamental right- and wrong-making features of actions.²⁵⁹ The requirement that is expressed in the Formula of Universal Law is generally expected to achieve both these aims, thus running into the Problem of Relevant Descriptions. According to Timmons, however, we can avoid this problem if we treat the latter as a mere decision tool and turn to the Formula of Humanity for Kant's criterion of right.²⁶⁰ It is the Formula of Humanity, he maintains, that "represents Kant's theory about [the] ultimate features in virtue of which an action is right or wrong," insofar as it tells us that they are "facts about how a concrete action, were it performed, would affect 'humanity' (whether in others or one's self)" (Timmons 1997: 401, see also 395-7). Timmons argues that, as such, Kant's Formula of Humanity supplies an objective standard of moral relevance or a norm for the ideal exercise of our moral sensibility. It is this: in forming their maxims, agents ought to be sensitive to how their proposed action will affect their own and other people's autonomy and the conditions thereof in the circumstances at issue and choose a description that makes this impact explicit.²⁶¹

If we abstract from inconsequential interpretative differences and attend to the substance of the proposals, we can see that this is exactly the solution that we arrive at if we take seriously Herman's claim that individual actions based on flawed rules

²⁵⁸ Timmons 1997: 390. A further reason is that too much attention has been paid to the first formula (1997: 415).

²⁵⁹ For a detailed account of the two roles, see Timmons 1997: 391-5. We distinguished these roles in the introduction, sect. 1.1. There we spoke of guides and standards. In sect. 2.2 above and, earlier, in ch. 2, sect. 3.2, we also considered how this distinction applies to Kant's Categorical Imperative. In these contexts, we spoke of the Categorical Imperative as either a decision tool or a norm that determines the deontic status of actions. The latter notion is related to Timmons' notion of a criterion of right and the notion of an explanatory standard that we used in the introduction, but it is not equivalent to them. In my view, the Categorical Imperative is primarily a norm that determines the deontic status of actions, but it does not specify the features in virtue of which these actions are permissible, impermissible or obligatory or explain why they are. For this, we need to look at specific principles of duty.

²⁶⁰ Kant's Formula of Humanity, recall, says to "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).

²⁶¹ Timmons 1997: 406-8. According to Timmons, Kant specifies the content of the idea of respecting autonomy in his system of duties, as outlined in the Doctrine of Virtue.

of moral salience *can* be classified as impermissible.²⁶² Herman does not speak of roles, but she distinguishes between two “aspects” of the moral law: the “CI procedure [which] interprets [the moral law] in a manner suited to the limits of the human understanding and the conditions of human judgment” (Herman 1993: 84), and the more fundamental “conception of oneself as a moral agent, [which is] the aspect of the moral law [that] provides the foundation for pre-procedural moral rules”.²⁶³ The former is akin to Timmons’ notion of a decision tool, the latter to his notion of a criterion of right. Both believe that at the heart of Kant’s account of moral sensibility lies his conception of autonomy, and that the moral sensibility of an agent is adequate to the extent that it reflects a proper understanding of autonomy, its extension, and the conditions of its exercise. In assessing this idea, I will focus on Timmons’ account.

Note, first of all, that this proposal seems to violate the demand that we uncovered in sect. 3. According to Timmons, the Formula of Humanity is a criterion of right, that is, a requirement that by itself *determines* the deontic status of actions. As such, it is not just a standard of relevance. It prescribes treating autonomous agents in a certain way, not merely attending to their being such agents for the purpose of maxim formation, and it deems facts about how our actions will affect autonomy right- or wrong-making, which is not the same as deeming them morally relevant. As far as Anscombe’s objection is concerned, the Formula of Humanity is thus a stipulation of the wrong kind. Instead of supplementing the Formula of Universal Law, it replaces it. Timmons is aware of this pitfall. In a section entitled “Isn’t the Universal Law formulation really superfluous on the differential roles interpretation?”, he tries to show that this is a false impression (Timmons 1997: 411-3). The Formula of Universal Law, he points out, is still needed as a decision tool because it helps agents to contemplate their actions and the situations that they are in from an impartial point of view. As such, it safeguards their judgment from the distorting influence of inclination and helps them adjudicate conflicting grounds of obligation.²⁶⁴

²⁶² Herman argues that the conception of persons as ends-in-themselves is not derived from the Formula of Humanity. Instead, she maintains that it lies behind all the formulations of the Categorical Imperative. Herman 1993: 87, n. 12. This exegetical difference between her and Timmons’ accounts is inconsequential, however.

²⁶³ Herman 1993: 86. For the claim that this conception is more fundamental, see her 1993: 87, n. 12.

²⁶⁴ For a discussion of grounds of obligation, see ch. 1, sect. 3.

In order to evaluate this response, we have to elaborate on an idea already touched on in sect. 2.2: the idea that Kant's Formula of Universal Law supplies a *mere* decision tool. In the context of Timmons' differential roles interpretation, terms such as "decision tool" or "decision procedure" are not metaphorical: they are not mere illustrations of a norm that determines the deontic status of actions (if they were, then the distinction between decision tool and criterion of right would collapse).²⁶⁵ Instead, these terms refer to a tool that helps finite rational agents to control the distorting effects that inclinations have on their judgment. Such an interpretation of Kant's Formula of Universal Law not only underestimates the importance accorded to all the formulations of the Categorical Imperative in what is, after all, meant to be a *metaphysical* inquiry, it also misconstrues the relation between the formulas. While Kant asserts their equivalence (G 4:436), this interpretation treats the first formula as auxiliary to the second, as a mere control device that cannot be applied on its own. In light of these remarks, Timmons' reply seems almost question-begging: he can only avoid making the Formula of Universal Law superfluous by arguing that it was more or less superfluous all along.

The proposal to solve the Problem of Relevant Descriptions by appeal to Kant's Formula of Humanity faces a second objection that is graver still. The very same criticism that Anscombe levels against the Formula of Universal Law seems to apply to the Formula of Humanity as well. We were hoping that this latter formula would bridge the gap between the circumstances of action and the agent's conception of them, but on closer examination it, too, expresses a requirement that is description-relative in the special sense. While there is no mention of maxims, there is mention of treating something as a means or an end, so that the question of whether the requirement is met hinges, once again, on how agents conceive their actions and circumstances, but not on whether their conceptions fit the circumstances that they are actually in.

Compare the above *Grandmother* example to Kant's famous example of someone who acts on a maxim of false promising, which we discussed in ch. 3, sect. 3.3.

²⁶⁵ It is tempting to think that they are because Kantian Constructivists couch their views in procedural terms, too. See fn. 113 above.

False Promising: “When I believe myself to be in need of money, I shall borrow money and promise to repay it, even though I know that this will never happen” (G 4:422).

When acting on a maxim of making a false promise in order to save oneself from financial embarrassment, one relies on their being a *promisee* who will accept one’s promise and give one money. There is thus a straightforward sense in which one is using them as a means, and as a mere means at that. But what would it mean to say that, in acting on a maxim of practicing in order to develop your guitar skills, you are treating your grandmother as a means or mere means? In sect. 2.4, we noted that you may not have thought about your grandmother at all, let alone planned on using her as a means to your end of guitar skill development. Had you acted on a maxim of breaking the promise that you made to her, you would have treated her as a mere means; but, as things stand, it seems that the Formula of Humanity would deem your staying in the park permissible or even obligatory. Here, again, it seems that Kant’s formula would need to be supplemented with stipulations determining how one *ought to* think of one’s means and ends in order for it to yield objective verdicts.²⁶⁶ Kantians will object. They will say that you are using your grandmother as a mere means because she cannot agree to being treated in this way, and this is what it means to treat someone as a mere means.²⁶⁷

My reading of the Formula of Humanity will seem uncharitable because it assumes that the grandmother cannot object to being left in the lurch unless she has some *general* objection to practicing the guitar for the sake of guitar skill development (in the way that one might have a general objection to making false promises for the sake of financial relief, say). The reading assumes that treating humanity as an end does not require adopting ends and means that fit the circumstances, that it does not require being sensitive to the morally relevant features of the case. This assumption makes it susceptible to an objection that is similar to Anscombe’s. We have seen that there is a more charitable reading, however. On this reading, one is being treated as a mere means if one cannot agree to being treated *in this way*. The demonstrative adject-

²⁶⁶ We saw that such stipulations would make any *further* requirement superfluous, be it the universality requirement or the requirement to always treat humanity as an end. This is related to the point we made about obligatory ends in sect. 3 above.

²⁶⁷ This interpretation of the Formula of Humanity, suggested by Kant’s discussion of his four examples in G 4:429–30, is a widely accepted one. See e.g. Wood 1999: 148–9.

tive indicates that the italicized phrase refers to a *particular* action, an action that is embedded in the circumstances of its occurrence. When the injunction to always treat humanity as an end is taken to apply to actions in context, it gives us grounds for criticizing agents whose actions result from their flawed moral sensibility.²⁶⁸ With that said, we must ask whether the Problem of Relevant Descriptions needs a solution at all, or whether an analogous reading of the Formula of Universal Law would allow us to *dissolve* it.²⁶⁹ Let me conclude with a few remarks on the prospects of solving and dissolving it, respectively.

5. Taking Stock: A Dissolution of the Problem of Relevant Descriptions

We have shown that the two solutions considered in sect. 4.1 and 4.2 fail to meet the non-superfluous demand. Instead of supplementing Kant's universalisability requirement with a neutral principle of moral relevance, they replace it with a different requirement or *moral* principle. That they suffer from this flaw is not a coincidence. In fact, given how we have construed the problem, every attempt to solve it is bound to suffer from this flaw because there is no robust theory of moral relevance that is not, at the same time, a moral theory. If pleasure is morally relevant, it is relevant *because* it ought to be promoted; if all the considerations mentioned in divine commands are morally relevant, they are relevant *because* divine commands ought to be complied with. R.M. Hare gets to the heart of this idea when he says that it is

a great mistake to think that there can be a morally or evaluatively neutral process of picking out relevant features of a situation, which can then be followed by the job of appraising or evaluating the situation morally ... When we decide what features of the description are morally relevant, we are already in the moral business.²⁷⁰

Some solutions to the Problem of Relevant Descriptions may conceal the fact that a principle of moral relevance is always already a moral principle. Herman, for instance, goes to great lengths to convince the reader that her rules of moral salience

²⁶⁸ This reading of the Formula of Humanity goes hand in hand with the Agent-Scope Reading of Kant's conception of universal validity, as it figures in his Formula of Universal Law. In both cases, the deontic status of actions is assessed *relative to the circumstances* in which these actions are embedded.

²⁶⁹ To *solve* the problem, one would have to provide the stipulations that Anscombe demands, that is, an account of moral sensibility or, in other words, an objective standard of moral relevance. To *dissolve* the problem, one would have to provide an interpretation of the Formula of Universal Law that shows that no such stipulations are needed.

²⁷⁰ Hare 1978: 193. On this point see also Lyons 1965: 101-2 and Timmons 1997: 403.

are there *only* to direct our attention to “moral danger” (1993: 78) and not to replace the universalisability requirement. Nevertheless, if we supplement Kant’s Formula of Universal Law with such a set of merely attention-guiding rules that we decide to privilege over other such rules (which, without a robust moral theory in the background, would be arbitrary), we arrive at a compound principle that is different from and thus a substitute for the universalisability requirement. Either way, the non-superfluous demand cannot be met. So where does this leave Kantians?

In sect. 4.2 above, we saw that there is, in principle, as much reason to think that the Formula of Humanity faces the Problem of Relevant Descriptions as there is in the case of the Formula of Universal Law. But no one seems to be worried that it does. In fact, with the Formula of Humanity, the Kantian retort was at our fingertips: of course, being treated as an end in itself means being able to agree to being treated *in this way*, being able to agree to *this* particular action in *these* particular circumstances. In order to understand why there is this difference, why the Problem of Relevant Descriptions seems so urgent and pressing in one case, whereas, in the other, even raising it seems dumb, we need to remind ourselves of the distinction that we highlighted in ch. 3, sect. 2.5. According to the standard Case-Scope Reading of Kant’s conception of universal validity, the Formula of Universal Law expresses the requirement to only act on principles that one can will to hold in all cases or in all circumstances. It is a demand to act on principles whose moral standing is not only independent of whether they fit the specific circumstances that one is in, but in fact precisely a matter of whether they generalize to all other cases. With this reading in mind, one is bound to wonder how the universalisability requirement is meant to ensure that the maxims it countenances refer to all and only morally relevant features of the case at hand.

Here the Agent-Scope Reading comes to the rescue. According to this reading, the universalisability requirement is the requirement to only act on principles that one (*any-one*) can will to be morally viable for any rational agent who is *in the same circumstance*.²⁷¹ The great advantage of this reading lies in the fact that it subjects the agent’s assumption that their maxim fits the circumstances to critical scrutiny

²⁷¹ Recall that this is the difference between universality and universalisability (see ch. 3, sect. 3.2). Not all principles of action are principles of duty. Many of them reflect our non-moral ends and projects. Such principles must be *morally viable* for anyone in the same circumstances in the sense that anyone must be allowed to act on them, but, of course, they do not *bind* people who don’t share the relevant ends.

through the universalisability requirement itself. In short, to satisfy the Agent-Scope version of the universalisability requirement, one has to act on a principle that fits the circumstances. As such, this requirement does not call for any Anscombian relevance-stipulations. In fact, to think of the universalisability requirement in this way is to think that it *is* Kant's account of moral sensibility:²⁷² that it is description-relative in the general sense, not in the special sense, and that it urges us to ask, about each deliberative step that we take in forming our maxim, whether it can be justified to all rational agents. Unlike the Case-Scope version of the universalisability requirement, satisfying *this* version is not merely a matter of adopting or dismissing principles that were formed in accordance with another standard or through the employment of some random sensibility beforehand.

Recall the *Asylum*₂ cases and immigration officer A who rejects the application simply because the grounds appealed to are invalid. There might be many situations in which his maxim would satisfy the Agent-Scope version of the universalisability requirement. But whether it does so in the circumstances at hand depends on whether all rational agents can agree that the fact that the applicant is a victim of sexual persecution has no bearing on what one is allowed or obligated to do in the officer's position. One may well think that it does.²⁷³

²⁷² In fact, this interpretation is supported by textual evidence. In a section entitled "Of the Typic of Pure Practical Judgment", where Kant discusses the problem of subsuming actions possible in the sensible world under the moral law, he claims that "the rule of judgment under laws of pure practical reason is this: ask yourself whether, if the action you propose were to take place by a law of the nature of which you were yourself a part, you could indeed regard it as possible through your will" (CPrR 5:69). This suggests that the Categorical Imperative *itself* bridges the gap between abstract principles and concrete actions.

²⁷³ Here, again, one might worry about the Empty Formalism Objection. See ch. 3, sect. 3.3, esp. fn. 192, 197 and 198 for comments on this point.

Chapter 5

Kantian Constructivism – Formal or Substantive?

A Response to Street

1. The Point of Contention between Humean and Kantian Constructivists

The ambitions associated with the term “Kantian Constructivism” have certainly grown since it first appeared on the stage of philosophical debate in 1980. Initially devised as a theory of political justification meant to account for deep moral disagreement,²⁷⁴ Rawls’ idea has been developed into a full-fledged metaethical view that aims to secure both the objectivity of moral demands and their authority over us.²⁷⁵ Although critics have acknowledged the appeal of the constructivist project, they have cast doubt on its viability. In particular, many question whether Kantian Constructivists can vindicate their claim that moral judgments are strongly objective: that they hold for all rational agents as such.

My main aim in this chapter is to defend Kantian Constructivism against Sharon Street’s version of this criticism, but what I have to say seems to me to apply more widely. As a Humean Constructivist, Street subscribes to the main tenet of the constructivist doctrine: the idea that, even though judgments about practical reasons are subject to standards of correctness, these standards should be thought of as internal to the practical point of view.²⁷⁶ Whether our judgments about reasons are correct is not, as moral realists have it, a matter of how they represent the relevant attitude-independent moral facts, but rather a matter of their withstanding a procedure of

²⁷⁴ This reading of Rawls’ lectures on “Kantian Constructivism in Moral Theory” (1980: 515-72) has been advocated by Miriam Ronzoni and Laura Valentini (2008: 416-7).

²⁷⁵ For a comparison of this view with other metaethical views, see Korsgaard 2008: 321-4 and Street 2010: 363-84.

²⁷⁶ Street 2008: 207-8. Three preliminary remarks on the focus of this chapter. First, since my primary concern is Street’s criticism, I will not address the ontological question of whether reasons exist attitude-independently, because that is a point on which Humean and Kantian Constructivists agree. They are united in their opposition to moral realism (for arguments that Kant is a moral realist, see e.g. Wood 2008: 111-4 and Stern 2012: 26-40, 2015: 31-33; for an argument that he is not a moral realist, see e.g. Bojanowski 2012). What I will focus on instead is a question about the correctness of our judgments about practical reasons. Here, too, the two camps seem to agree on a response: the correctness of our judgments about reasons is determined by standards internal to the practical point of view. However, as we shall see, this response is only superficially identical. Second, I am concerned exclusively with practical reasons, so when I say “judgments about reasons” or “normative judgments”, this is short for “judgments about practical reasons”. Third, I will follow Street in speaking about the correctness or validity of judgments about reasons rather than their truth. Nothing in what I say below hinges on this point.

scrutiny or construction from the standpoint of (a subset of) our other judgments about reasons.²⁷⁷ In sect. 3.1 below, I will outline Street's taxonomy of constructivist positions, but for now it will suffice to pin down her disagreement with the Kantian Constructivist.

To bring the point of contention into focus, it is useful to distinguish between a weak and a strong conception of objectivity. As a constructivist, Street prides herself on being able to account for the objectivity of judgments about reasons. What she means by that, however, is merely that her procedure of scrutiny enables everyone to reach agreement on what a particular agent has reason to do given his or her commitments.²⁷⁸ For Street, a statement of the form "X is a reason to Y" is elliptical in that it cannot be assessed before we specify whose reasons are at issue. Kantian Constructivists, by contrast, hope to establish that judgments about reasons can be objective in a stronger sense. They claim that, with their procedure of scrutiny at hand, certain judgments about reasons (namely judgments about moral reasons) can be agreed to hold not only *by* all rational agents as such but also *for* them. In their view, if a statement of the form "X is a reason to Y" picks out a moral reason, we can assess its validity without knowing to whom it is meant to apply. I shall label this claim, which is the main point of contention between Humean and Kantian Constructivists, the *Strong Objectivity* claim.²⁷⁹

Strong Objectivity: There are judgments about reasons, namely judgments about moral reasons, that can be agreed to hold by and for all rational agents as such.

²⁷⁷ In what follows, I will use the term "procedure of scrutiny" rather than "procedure of construction". When it comes to metaethical constructivism, I would say that the former term is more appropriate because the term "construction" suggests that we use some specific set of materials to construct judgments "from scratch". According to metaethical constructivists like Street and Korsgaard, however, this is not what we are doing when we make judgments about reasons. What we are doing, on their view, is scrutinizing a candidate judgment that presents itself (e.g. through an impulse) in light of a fluid system of normative commitments the content of which cannot be presupposed as part of the procedure itself.

²⁷⁸ Street 2008: 224-5. Drawing on a distinction introduced by Scanlon, we could say that, on Street's view, reasons are "judgment-independent" in the sense that it is possible for us to go wrong in our judgments about them, but they are not "choice-independent" in the sense that the standards for assessing judgments about reasons depend on what we have chosen or adopted (Scanlon 2012: 233). For Kantians, moral reasons are both judgment-independent and choice-independent.

²⁷⁹ Note that this is not how Street characterizes the disagreement (2012: 41). My reason for diverging from her characterization will become apparent in the course of my argument. Briefly, my account of the controversy is broader than Street's because I want to leave room for an alternative understanding of Kantian Constructivism, which I will contrast with hers.

Note the twofold significance of the phrase “as such” in the above formulation. On the one hand, it renders the claim more modest than it would otherwise be, since it registers the well-known fact that Kantians are not concerned with *actual* agreement – that they leave room for our cognitive limitations. To be correct, a judgment about reasons only has to be agreed to hold by agents insofar as they take up the standpoint of rationality. On the other hand, the phrase renders the claim more ambitious, and the sense in which it does is pertinent to understanding how Kantian Constructivism differs from Humean Constructivism. To say that a judgment holds for all rational agents *as such* is to say that its validity for each and every one of them doesn’t depend on any contingent overlap among their individual commitments.²⁸⁰ Consider the judgment “The fact that my country is at war is a reason to sacrifice my life”. If, as a matter of coincidence or due to some evolutionary force, all people (past, future, and present) happened to be fanatical nationalists, all rational agents would indeed be able to agree that the above judgment holds for them: that their country’s involvement in a war gives them a reason to sacrifice their lives. But this judgment would not hold for all rational agents *as such*. It would hold for them only *as nationalists*. For Humean Constructivists, our moral judgments are at most contingently universal (Street: 2008: 225); for Kantian Constructivists, their universal validity is necessary.

Street’s argument against Kantian Constructivism is supposed to undermine the *Strong Objectivity* claim. In what follows, I hope to show that her dismissal of the view is based on the uncharitable assumption that there is only *one* metaethical version of Kantian Constructivism, namely the approach advocated by Korsgaard in *The Sources of Normativity*. Korsgaard’s argument in the *Sources* seeks to establish that, insofar as we engage in practical reasoning at all, we have reason to adopt a moral outlook, or, in other words, to value humanity in ourselves and others (1996b: 120-6, 132-45). It is from this basic substantive judgment that all further judgments about moral reasons are thought to derive and from which they are supposed to inherit their strong objectivity. In my view, this form of argument grants too much to the Humean because it relies on their account of what it means to make a judgment about reasons. But this Korsgaardian version of Kantian Constructivism is in fact not the only ver-

²⁸⁰ In short, the phrase “as such” indicates that the judgment is counterfactually robust. We have already touched on counterfactual robustness in ch. 3, sect. 2.2, and we will return to the points raised there later on.

sion there is. I will argue for an alternative version, which establishes the *Strong Objectivity* claim via a much more direct route.

In sect. 2, I will outline the view that Street ascribes to Kant. Her interpretation relies largely, if not entirely, on Korsgaard's argument for the value of humanity in *The Sources of Normativity*. Along the way, I will focus on the claims and passages that expose Korsgaard to Street's attacks. The resulting reading of the *Sources* is certainly not the most charitable one available, but I will take time to show how the text lends itself to the problematic interpretation that has invited so much criticism. I will then go on to present Street's objection to Korsgaard (2012) and supplement it with similar points of criticism raised elsewhere. For the most part, I will endorse her objection, but I will note that, given the arguments she offers, one of her conclusions concerning constitutive standards is too strong. Having shown that Kantian Constructivism as Street, following Korsgaard, conceives of it is prone to fatal objections, I will examine whether there is room for an alternative version of Kantian Constructivism.

This is the task I will tackle in sect. 3. Taking Street's taxonomy of constructivist positions as a starting point, I will expose her undefended move from the claim that metaethical constructivists must appeal to a purely formal procedure of scrutiny to the claim that they must appeal to *her* procedure (2008: 224-30). This claim saddles her with a rather implausible view: the view that the difference between Kantian and Humean Constructivism does *not* concern their procedures of scrutiny. I will argue that this is wrong and that Street neglects an alternative. This alternative position comes into view once we reflect on why Kantians go along with this: why Korsgaard opts for a Humean procedure of scrutiny, for example. In trying to answer this question, I will draw on a contrast that I highlighted earlier (ch. 3, sect. 1, sect. 2.5), namely the contrast between two concerns that might be taken to underlie Kant's insistence on universal validity: a concern for consistency and a concern for objectivity. Korsgaard seems to assume that his main concern is the former. This leads her to a reading of the Categorical Imperative according to which it is a mere coherence requirement. If we instead suppose that Kant's main concern is the latter – a concern for objectivity – then we arrive at a formal procedure of scrutiny that is distinctly Kantian. Unlike Korsgaard's, the corresponding version of Kantian Constructivism can secure the *Strong Objectivity* claim because it introduces the perspective of other agents into the procedure itself.

2. Street's Objections to Kantian Constructivism

Street reserves the label “Kantian Constructivism” for a metaethical theory that is concerned with the correctness of judgments about reasons in general. She contrasts such *metaethical* versions of constructivism with *restricted* versions that limit the scope of their claims to a specific set of judgments about reasons, e.g. judgments about justice or judgments about moral rightness and wrongness.²⁸¹ Even though this use of the term might strike some as inappropriately narrow (after all, it was Rawls, a restricted constructivist, who introduced the term “Kantian Constructivism”), it is a use that suits our purposes: the purpose of determining whether Kantian Constructivists can establish the *Strong Objectivity* claim. We can see this if we think about how restricted constructivists proceed. According to Street's taxonomy, restricted constructivists are constructivists who presuppose a “grounding set of judgments” about reasons (Street 2008: 210).²⁸² As such, they develop procedures that fail to address all those who do not share these judgments. The deliverances of their procedures are judgments that only *certain* rational agents, *as those who accept the judgments in the grounding set*, can agree to hold for them.

Consider Rawls' theory of justice as fairness, for example. If voters and political decision makers are to use his procedure of scrutiny (the so-called original position) when choosing principles of justice and assessing the basic structure of their society, they must already be convinced that they have reason to participate in a fair system of cooperation and to treat people as free and equal (Rawls 1999: 475). The original position doesn't have a grip on those who reject the two basic judgments that are embedded in it. For why should they regard its verdicts as binding? Take the judgment “The fact that this policy violates the second principle of justice is a reason not to implement it”. A political decision maker who rejects the normative conceptions of persons and societies embedded in Rawls' original position might grant the factual

²⁸¹ Street discusses the views of Rawls and Scanlon in particular (2008: 209-14).

²⁸² Because restricted constructivists do not say anything about what it is for the judgments in their grounding set to be correct, Street concludes that their theories are not metaethical theories proper. In fact, in her view, their accounts are compatible with different metaethical theories (Street 2008: 217-20). Of course, Street acknowledges that Rawls *does* address this question. His response to it changes over time. Some passages in his early work suggest that he takes a Kantian stance on it, assuming that the grounding set judgments are constitutive of what it means to make judgments about reasons. But other passages, especially those concerned with his notion of reflective equilibrium, make our commitment to the grounding set judgments look more contingent and retractable. In his later work, he takes an explicitly quietist stance, arguing that endorsement of the grounding set judgments must be regarded as a matter of overlapping consensus.

claim contained in this judgment but deny that it gives him any reason to stop the policy in question from being implemented. Because Rawls' constructivist theory can only establish what a certain limited set of rational agents can or cannot agree to hold for them, it cannot secure the *Strong Objectivity* claim. In what follows, I will focus on non-restricted, metaethical versions of Kantian Constructivism. Street considers only one view that falls into this category: Korsgaard's view in the *Sources*. However, as we shall see in sect. 3, there might be others.

2.1 Korsgaard's Argument for the Value of Humanity

In *The Sources of Normativity*, Korsgaard pursues two projects. On the one hand, she provides a constructivist theory of rational agency in general, or, in other words, an account of what constitutes the correctness of judgments about reasons in general. On the other hand, she offers an answer to what she calls "the normative question", the question of "what *justifies* the claims that morality makes on us" (1996b: 9-10). The answers to these questions cannot be entirely independent, of course, since to justify a claim just is to give a reason, but to the extent that there are two separate questions they can be treated separately – and that is exactly what Korsgaard seems to be doing. This, I will argue, is the main problem with her view and the reason why it is susceptible to the kind of criticism voiced by Street and others. So let us consider her answers to the two questions in turn.

Korsgaard's theory of rational agency is constructivist in that it conceives of the correctness of our judgments about reasons as a matter of withstanding a procedure of scrutiny from the standpoint of our other judgments about reasons. The procedure of scrutiny that she favours, the so-called "reflective endorsement method" (1996b: 50), is inspired by a range of historical and contemporary figures, including David Hume, John Stuart Mill, and Bernard Williams.²⁸³ What makes us rational agents, according to Korsgaard, is the fact that we are reflective beings. When we are confronted with an impulse, we can and must step back from it and ask ourselves whether we really have reason to do what, in virtue of the impulse, we are inclined to do. "Each impulse as it offers itself to the will," she writes, "must pass a kind of test

²⁸³ That Korsgaard's procedure traces back to Hume prefigures my criticism that she grants too much to the Humean.

of normativity before we can adopt it as a reason for action”.²⁸⁴ And that test is the test of reflective endorsement. Translating this thought into Kant’s terms, she rephrases it in the following way: before we can act on a desire, “we must *make it our maxim*” (Korsgaard 1996b: 94) by asking ourselves whether it can be a law for us. Below we will see that her understanding of this question deviates from Kant’s, but for now let us consider her account of how we answer it.

Part of what it is to be a reflective being, according to Korsgaard, is to have a normative conception of oneself or practical identities. A practical identity is 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 1996b: 101). These identities, e.g. the identity of a teacher, a lover or a citizen, give rise to reasons and obligations because it is in light of these various identities that we ask ourselves whether to endorse an impulse. If we abstract from the practical identities of the person whose reasons are at issue, the question whether a maxim can be a law for them has no answer; it is empty.²⁸⁵ In order to square her claim that the reflective endorsement test per se is empty with Kant’s Formula of Universal Law, Korsgaard is forced to introduce a distinction that does not appear in Kant. She distinguishes between the Categorical Imperative, which for her is another name for the test of reflective endorsement, and the moral law, which she views as the law of an agent with a particular identity: a citizen of the Kingdom of Ends (Korsgaard 1996b: 98-100). The moral law is not empty. It demands that we “act only on maxims that all rational beings could agree to act on together in a workable cooperative system” (Korsgaard 1996b: 99). But because this requirement is not part of Korsgaard’s conception of what it is to make a correct judgment about reason, it has to be established by means of a *separate* argument that this law, and the identity that lies behind it, is one that all rational agents as such have a reason to adopt. Before I explain how

²⁸⁴ Korsgaard 1996b: 91. She adds that “the reflective mind cannot settle for perception and desire, not just as such. It needs a *reason*. Otherwise, at least as long as it reflects, it cannot commit itself or go forward” (1996b: 93).

²⁸⁵ Korsgaard 1996b: 113. This, she notes, is the “element of relativism” that attaches to the reflective endorsement test and to Kant’s question of whether a maxim can be a law for us. When confronted with this question, we must always ask: a law *for whom*? It seems to me that there is a tension between Korsgaard’s claim that a maxim is an intrinsically normative entity (1996b: 108) and the view that a maxim’s endorsability depends on the agent’s identities. This tension leads to some of the difficulties discussed below.

this separation exposes her argument to Street's objection, let me comment on her second project: answering the normative question.

While most of our various practical identities are contingent, there is something about them, Korsgaard argues, that is not: we must adopt some practical identity or other.²⁸⁶ For without any such identity, a reflective being could not act at all. So we have a reason to embrace and conform to some practical identity or other, and that reason springs from our nature as reflective beings, or from our *humanity*. In contrast to our particular practical identities, our identity as human beings is one that we cannot reject. It is inescapable because we are the kinds of beings that we are, and it is inescapable in a normative sense: it provides us with reasons to identify ourselves in some way. Note the twofold relation that our identity as human beings bears to the particular identities and reasons we have. On the one hand, it is *supported* by the fact that we need some identities and reasons if we are to act at all.²⁸⁷ On the other hand, it also *supports* these particular identities and reasons insofar as "their importance is partly derived from the importance of being human" (Korsgaard 1996b: 121). It is due to this twofold relation that our identity as humans can stop the looming regress of reasons. It supplies a reason for embracing particular practical identities and does not require any higher-order identity to supply reasons in its support. In Korsgaard's view, our identity as human beings is at the same time our moral identity because it gives us a reason to value humanity as such.

In order to move from the idea that we have a reason to embrace some practical identity or other to the *Strong Objectivity* claim, Korsgaard relies on two assumptions, only one of which is explicitly argued for in the text. The first assumption is that our identity as humans comes with substantive reasons and obligations that are recognizably moral, i.e. that embracing this identity is tantamount to *valuing* humanity in a non-technical sense of the term. We might wonder why this should be so, but Korsgaard does not provide much of an answer²⁸⁸ – and indeed, we will see below that there isn't one. The second assumption is that the reason that one has for valu-

²⁸⁶ Korsgaard 1996b: 120. The argument I go on to reconstruct can be found in her 1996b: 120-125.

²⁸⁷ Perhaps it would be more appropriate to say that it is *self-supporting* because, technically speaking, it is not that our particular identities and reasons give us a reason to embrace our identity as humans. Rather, this traces to the fact that we are reflective beings who cannot act without reasons. On this point see Street 2012: 47.

²⁸⁸ Recall that my focus in this section is exclusively on the *Sources* because this is the text that Street engages with. I think that Korsgaard does offer an answer to this question elsewhere.

ing one's own humanity is in fact a reason to value humanity as such, and therefore also a reason to value other people's humanity. Korsgaard's argument in support of this second assumption draws on Wittgenstein's famous private language argument (1996b: 132-45). Since Street's objection, as well as the objections of many others, are directed at the first assumption, I will not go into Korsgaard's argument for the second here.²⁸⁹ Instead, let me return to the question of how the two projects pursued in the *Sources* are related to one another.

Now that we have seen how Korsgaard accounts for the correctness of judgments about reasons, it is in fact plainly obvious that her answer to the normative question and her argument for the *Strong Objectivity* claim must take precisely the form they do. If a judgment about reasons holds for an agent insofar as it can be derived from and does not conflict with their various practical identities, then, in order to establish that moral judgments hold for all rational agents as such, one needs to show that all rational agents as such must identify themselves as moral agents. To have a reason is to embrace an identity from which that reason arises; thus, to have a moral reason is to embrace a moral identity.

2.2 Street's Objection to Korsgaard's Argument

Ever since the publication of Korsgaard's Tanner Lectures in 1996,²⁹⁰ her arguments have been the subject of ongoing debate, and over the course of the years many readers have raised doubts about and objections to her argument for the value of humanity. Although my discussion will focus on Street's objection to the argument, I will draw attention to some of the links and similarities to criticisms and worries voiced by others. Street's objection turns on the observation that the conclusion of Korsgaard's argument for the value of humanity lends itself to two different readings (Street 2012: 51). The answer to the normative question that Korsgaard takes to emerge from that argument, and by means of which she takes herself to have established the *Strong Objectivity* claim, is that we *must* embrace our identity as human beings; we must value humanity in ourselves and in others. The "must" contained in

²⁸⁹ The argument against the notion of a private reason that Korsgaard offers in support of the second assumption suggests that Street's reading of the argument in the *Sources*, which I follow here, is not the most charitable one available. In fact, this argument points to the account that I will propose in sect. 3, as do arguments that she presents in other pieces (see e.g. Korsgaard 1996a: 275-310).

²⁹⁰ The Tanner Lectures, held in 1992, form the basis of *The Sources of Normativity*.

this conclusion, Street remarks, can be understood as a normative or as a constitutive notion.

If Korsgaard uses the term “must” to denote a *normative* requirement, then she is saying that we have a *reason* to value humanity in ourselves and in others. Her project could then be described as an attempt “to place morality within practical reason”²⁹¹ by showing how moral reasons follow from certain other branches of practical reason. Although Street herself does not comment on the plausibility of this reading, it is in fact supported by a considerable amount of textual evidence. In the second lecture, where Korsgaard retraces the history of the reflective endorsement method, she casts the task that is associated with the normative question in the following way:

The normative question is answered by showing that the points of view from which these different interests arise are congruent, that meeting the claims made from one point of view [e.g. the point of view of self-interest, or sympathy, or morality] will not necessarily mean violating those that arise from another.

(Korsgaard 1996b: 61)

According to this way of understanding the question, morality’s demands on us will be justified if and only if they can be vindicated from within the various standpoints that make practical claims on us – that is, from within our various practical identities. Now, of course, it might be objected that Korsgaard is not speaking with her own voice in the above passage. After all, this is said in the context of her engagement with Hume’s position, so it is perhaps unfair to take her at her word. I think this is true: for Korsgaard, the normative force of moral demands does not depend on particular practical identities in any straightforward way.²⁹² Otherwise, she would not regard herself as a *Kantian* Constructivist who is defending the *Strong Objectivity* claim.

²⁹¹ This formulation is from Hussain and Shah (2006: 267). In their paper, the two authors aim to show that, because Korsgaard’s project in the *Sources* is one of placing morality within practical reason, her position is not really a *metaethical* position that can be regarded as an alternative to moral realism. This conclusion is supposed to follow no matter how her project is understood – that is, whether it is understood as an attempt a) to show how moral reasons follow from other types of reasons or b) to show how moral reasons follow from a conception of practical reason as such. I agree that a) should be classified as an exercise in *normative* ethics, but the claim that b) is an exercise in normative ethics as well is at least disputable.

²⁹² If it did, if she was trying to offer non-moral reasons for moral conduct, then she would face what, following Robert Stern, we can call the “Prichardian Challenge” (Stern 2015: 76, referring to Prichard 2002: 7-20). She could be accused of trying to convince sceptical agents that morality has a grip on them by offering them the wrong kinds of reasons for moral behaviour.

However, there is nonetheless a sense in which Korsgaard presents morality as seeking approval from our particular practical identities. Recall that our identity as humans not only supports but is also itself supported by our particular practical identities. This need for approval “from below” is reflected in the deliberative situation of an agent who identifies, primarily, as a lover and a teacher, for example, and who asks themselves: should I do what morality demands of me? The answer at which our lover and teacher arrives, of course, is not the straightforward one that Humeans would expect. It is not that their love for their partner and their concern for their students, simply as such, give rise to reasons to be moral. Rather, they acknowledge that, qua reflective being, they could neither love their partner nor care for their students if they did not embrace their identity as a human. In order to do these things, they need reasons, and these reasons must bottom out in an appropriate way. So, in this limited sense, Korsgaard *does* take the demands of morality to be vindicated from within our particular identities, e.g. the identity of a lover or a teacher.²⁹³

In Street’s view, the project of establishing the *Strong Objectivity* claim by way of placing morality within practical reason is incoherent. It is based on a reading of the normative question that renders it *ill-formed*, at least in the mouth of someone who accepts metaethical constructivism (Street 2012: 49). To see why, we need to return to Korsgaard’s constructivist account of the correctness of judgments about reasons. On this account, a judgment about reasons is correct if and only if it withstands the reflective endorsement test: if it can be shown to be compatible with and derivable from (or, in short, to *cohere with*) the agent’s various practical identities. This *coherentist* characterization of the reflective endorsement method fits with Korsgaard’s claim that, in itself, the method is empty. It cannot be applied before we settle whose reasons are in question because otherwise there simply wouldn’t be any practical identities that a judgment candidate could (fail to) cohere with. But, as Street observes, this is exactly what we are being asked to do when faced with the question of whether we have any reason to embrace our practical identity as humans. For, given Korsgaard’s

²⁹³ Rae Langton argues that this makes it impossible for Korsgaard to accommodate Kant’s claim that the value of humanity is unconditional (2007: 177-81). Since, in Korsgaard’s view, our commitment to the value of humanity is conditional upon the fact that we value certain other things (our partner or our students, etc.), people who do not value anything (e.g. because they are too depressed to do so) lose their status as unconditionally valuable beings. Arguably, Korsgaard could respond that even very depressed people must value *something*, if only their own death. However, such an evaluative attitude does not seem to entail any moral constraints. We will come back to this problem shortly.

technical use of the term, this is just the question of whether we have any reason to identify ourselves as anything at all. We are being asked to “stand apart from all [our practical identities and] judgments about what count as reasons, and then to ask whether [we] have some further reason” (Street 2012: 49) – and that, in Street’s view, is a question that a metaethical constructivist should dismiss as *unanswerable*. They should dismiss it as unanswerable because it is asserted from nowhere (Street 2012: 50).²⁹⁴

So it is only if we interpret the “must” in Korsgaard’s conclusion as a *constitutive* notion that the conclusion as a whole can be taken to express something meaningful. That a rational agent must embrace their humanity, Street argues, can only mean that a rational agent is *defined* as someone who identifies themselves in one way or another and therefore takes something or other to be a reason (2012: 51). She suggests that we should read Korsgaard’s conclusion along the same lines as the claim that a parent must have a child: not as a claim about what a not-yet-parent has reason to do but rather as a claim about what it is to be a parent (Street 2012: 51). On the face of it, we wouldn’t expect Korsgaard to object to such an interpretation of her conclusion. For although she does not appeal to constitutive standards in the *Sources*, she has since become well known for advocating the claim that moral demands arise from what is constitutive of agency.²⁹⁵ But here we need to be careful, because what Street means by “constitutive” is not quite what Korsgaard means in her more recent writings. Before I explain what the difference amounts to, let me outline Street’s ar-

²⁹⁴ This is only one of many criticisms that the “regress of reasons” or “regress of identities” reading of Korsgaard’s argument has attracted. Stern, for instance, has argued that this argument goes wrong in assuming that we can fall into doubt about every single one of our particular practical identities in a way that would make it necessary to appeal to a more basic identity, our identity as humans (2015: 67). Some of our particular identities, he believes, are invulnerable to this kind of challenge from the point of view of other identities. However, in Stern’s view, there is a better reading available: the “regress of value” reading (2015: 68-72). According to this reading, Korsgaard’s argument proceeds roughly in the following way. We can only think of our identities as providing us with reasons if we *value* them, and (given Korsgaard’s opposition to value realism) we can only value them if we value our leading the kinds of rational lives that such identities make possible. In response, one could raise an objection that is similar to the second part of Street’s objection, which I go on to present in what follows. The crux of this objection, as I see it, is whether anything substantive follows from the idea of leading a rational life (or being a rational agent). Does this idea entail that I should not harm myself, for example? Something along these lines would seem to be required in order to get to *moral* demands once we move from the need to value our own humanity to the need to value the humanity of other people.

²⁹⁵ See Korsgaard 2008 and 2009. Part of my argument in this chapter depends on the interpretative claim that Korsgaard’s argument for the value of humanity, as presented in the *Sources*, is precisely *not* an argument to the effect that morality arises from what is constitutive of agency. That is what I see as its deepest flaw.

gument for why, on this reading too, Korsgaard's argument for the value of humanity falls short of establishing the *Strong Objectivity* claim.

According to Street, metaethical constructivists should regard the conclusion of Korsgaard's argument, *so* understood, as trivially true (2012: 51). Of course, if the correctness of our judgments about reasons is assessed on the basis of how these judgments cohere with our prior judgments about reasons, then only those who have prior commitments can count as making judgments about reasons – only they can count as exercising their rational agency. However, far from establishing the *Strong Objectivity* claim, this understanding of the term “rational agent” has no moral implications whatsoever. It says nothing about how we ought to treat ourselves or other people. As Street observes in the following passage, the only implication that it has is tautological:

Nothing substantive about the value of one's *self* is implied by the mere fact that one takes something or other to be valuable. At this point, the Kantian might grant that [we] need not take [ourselves] to be valuable in the ordinary sense of this expression (a sense we normally take to involve valuing one's own survival, bodily integrity, and so forth), but insist that what [we] *must* do is take [ourselves] to be valuable in the sense of valuing [our] own valuing. So far as I can see, however, *this* understanding of the “value of the self” is empty in the sense that it directs nothing substantive: it just says to value what you value. (Street 2012: 54-5)

If this observation is accurate, it follows that Korsgaard's argument for the value of humanity does not accomplish any of the goals that unite Kantian Constructivists. It doesn't answer the normative question (meaning that it doesn't establish morality's authority over us) and it doesn't establish the *Strong Objectivity* claim (meaning that it doesn't establish that moral demands are strongly objective). From the perspective of a Humean Constructivist like Street, it is precisely because Korsgaard is trying to do both that she is bound to fail: it is because she is trying to establish the correctness of our moral judgments independently of the judgments that express our contingent commitments and identities that she ends up with an empty set. Although Korsgaard does not spell out which particular moral demands her argument has shown to be binding,²⁹⁶ it is clear that a commitment to one's identity as a human is supposed to exclude certain specific practical identities and forms of conduct. Broadly speaking, it

²⁹⁶ She says that she will have “little to say about the content of ... [moral] obligations” (1996b: 92).

is meant to exclude all those identities and forms of conduct that tend to undermine or diminish someone's capacity for rational agency, e.g. by harming them, violating their freedom or ruining their self-respect.²⁹⁷ Humeans agree that the bindingness of such recognizably moral restrictions can be established; they agree that the normative question can be answered. But, in their view, it cannot be answered for all rational agents *as such*.²⁹⁸ This is because the mere fact that someone satisfies the definition of "rational agent" entails only that they have a reason to do whatever follows from the particular identities in virtue of which they satisfy this definition. There is nothing that representatives of the set of rational agents *in general* have reason to do.

What is easily overlooked, I think, is that this second part of Street's objection, which focuses on the constitutive reading, proves too much – and the way in which it does is quite surprising. In the recent literature on metaethical constructivism, it has become something of a convention to use the terms "constructivism" and "constitutivism" interchangeably, and this is not least because the notion of constitutive standards is so prominent in Korsgaard's more recent writings (Bagnoli 2011: §7.3). Street herself not only uses the term "constitutive" when outlining her procedure of scrutiny, but also explicitly recognizes that her thinking about what is constitutive of judgments about reasons "owes a great deal" (2008: 228) to Korsgaard. It is in light of this that Street's objection is somewhat surprising because she seems to agree with the idea, famously articulated and defended by David Enoch, that "normativity will not come from what is constitutive of action" (Enoch 2006: 171). This is an idea that Korsgaard would firmly reject. We can see how Street's account of the constitutive differs from Korsgaard's by comparing how they deal with violations. For Korsgaard, what is constitutive of a form of being or activity provides a standard for evaluation. So for her, an agent who violates a constitutive principle of agency is a

²⁹⁷ When we treat others in ways that neglect and thereby undermine their rational agency, we treat them as *mere means* rather than ends in themselves. Korsgaard elaborates on this idea in 1996a: 295-9.

²⁹⁸ Street expresses this idea as follows: "On a characteristically Kantian way of thinking about morality, it's part of the very idea of morality that its requirements are categorical – not something whose 'bindingness' one may escape merely by failing to care about it. ... The *Kantian constructivist's* mistake is the mistake of thinking that moral requirements must bind us independently of the *particular evaluative nature* with which we find ourselves – and, in particular, independently of whether we already have moral concerns as a deep part of our nature" (2012: 55-6). Korsgaard's stance on the modal status of our practical identities is hybrid: while she agrees with Humean Constructivists that our reasons could have been different had we adopted other identities, she shares the realist view that our moral reasons could not have been different, whatever choices we had made. The realist's objection to Korsgaard's view is accordingly the reverse of Street's: our particular identities, too, must be based on the reasons there are for adopting them. See e.g. Scanlon 2012: 238.

bad agent (Korsgaard 2009: 32-3, 159-76). Street, by contrast, treats what is constitutive of a form of being or activity as a classification criterion. In her view, an agent who consciously violates a constitutive principle of agency simply *drops out of agency*.²⁹⁹ It is not that, in such a case, they are doing something wrong. Instead, they simply aren't acting – they are shmacting, as Enoch would say (2006: 179).

I started out by saying that Street's objection proves too much, and with these two accounts of the constitutive in mind, I can now explain why. All she needed to show, to adapt Enoch's phrase, is that "morality will not come from what *Korsgaard* (in *The Sources*) takes to be constitutive of action". This, I think, is the claim she actually argues for, and convincingly so. The stronger claim, namely that "morality will not come from what is constitutive of action (period)," it seems to me, is simply being assumed. Below I will provide some reasons for thinking that the stronger claim might be false. It will in any case become apparent that it is no accident that Street conflates these two claims.

3. Formal Metaethical Constructivism: A Kantian Variety

The considerations presented in sect. 2.2 indicate that Korsgaard fails to show that "there are certain normative judgments [moral judgments] to which every reflective creature who accepts any normative judgment at all is committed" (Street 2008: 244). To be sure, one must accept some such judgments to count as a rational agent. But, according to Street, the mere fact that one is "playing the game of rational agency" and that one is subject to the demand to make coherent judgments does not by itself place any substantive restrictions on what one has reason to do. More specifically, she sees it as a distinguishing mark of metaethical constructivism that it appeals to a *purely formal* procedure of scrutiny, because this is what makes it a metaethical theory that can account for the correctness of judgments about reasons in general (Street 2008: 226). But, in her view, the hope that such a *formal* procedure, e.g. the reflective endorsement method, could entail a set of substantive judgments is in vain.

We agreed with Street that Korsgaard's reflective endorsement method does not entail any particular reasons or obligations, moral or otherwise. In what follows, we shall ask whether this is true of *all* formal procedures of scrutiny, and, specifically,

²⁹⁹ Street 2008: 228. This account of the constitutive has strange implications. These will be addressed in sect. 3.1 below.

whether it is true of the procedure that emerges from the Agent-Scope Reading of Kant's conception of universal validity. The first step is to see that this is a neglected alternative within Street's taxonomy of constructivist positions and that Korsgaard's interpretation of Kant's ethics facilitates this oversight.

3.1 Street's Taxonomy of Constructivist Positions

Restricted vs. Metaethical Constructivism

We began our discussion in sect. 2 by considering Rawls' theory of justice as fairness as an example of restricted constructivism. Such theories, we noted, cannot be regarded as full-fledged metaethical theories. Nor can they secure the *Strong Objectivity* claim, because the procedure of scrutiny that they appeal to has a so-called grounding set of judgments built into it. Rawls' original position, for instance, determines what is just or unjust on the basis of two first-order judgments about reasons that are built into the procedure, namely that persons are to be treated as free and equal and that societies are to be set up as fair systems of cooperation. The scope of his procedure is therefore limited to determining whether a judgment holds *for those who accept* (and *as* those who accept) these normative conceptions of persons and societies. The key to developing a constructivist theory that can at least potentially secure the *Strong Objectivity* claim is therefore to answer the following two questions separately:³⁰⁰

- a) Which procedure of scrutiny determines the correctness of judgments about reasons, and in light of which other judgments about reasons does it do so?
- b) Which judgments about reasons are correct?

Metaethical constructivists answer question a) without relying on anything that could be offered in response to question b) – that is, without incorporating any moral or other *content* into their procedure. In order to avoid contentful presuppositions, they characterize the set of judgments in light of which any proposed judgment is to be assessed not in terms of their content but in terms of *whose* judgments they are. The procedure itself captures the nature of the relationship that the proposed judgment is supposed to bear to these other judgments. The result is what Street calls

³⁰⁰ The first is a *metaethical* question, the second a *normative ethical* question. It is because restricted constructivists do not keep these questions apart, that their theories seem to fall between the cracks of metaethics and normative ethics.

“a purely formal characterization” (2008: 226) of what it means to make judgments about reasons in general. This is what all metaethical constructivist positions have in common.

Formal vs. Substantive Metaethical Constructivism

The disagreement between *formal* and *substantive* metaethical constructivists concerns the question of whether this formal characterization of the standpoint of rational agency entails any substantive judgments about reasons, i.e. whether the metaethical constructivist’s answer to question a) implies any answers to question b) (Street 2008: 243-4). According to Street, Korsgaard is a substantive metaethical constructivist.³⁰¹ As such, Korsgaard’s position differs from the position of restricted constructivists in that she characterizes her procedure of scrutiny without *presupposing* any first-order judgments about reasons. Instead, she argues that her formal procedure of scrutiny, the reflective endorsement method, *entails* some such judgments, namely judgments about moral reasons. In this way, she seeks to establish the *Strong Objectivity* claim.

**The* Formal Procedure of Scrutiny*

In her discussion, Street seems to assume that her formal procedure of scrutiny is the only game in town (2008: 224-30). Below, we will see that Korsgaard’s reading of Kant lends support to this assumption. But first we need to consider the details of Street’s answer to question a).

In reflecting on the set of judgments in light of which a proposed judgment is to be assessed, and in particular on the question of *whose* judgments they should be, Street reviews two possibilities: the correctness of judgments about reasons could relativize either to the standpoint of the speaker, i.e. the standpoint of the person judging, or to the standpoint of the agent. The first option is dismissed because it does not accord with how we speak. After all, on this account, “it [would be] impossible for you and me sensibly to disagree about whether X is a reason to Y for [agent] A, since the answer might be ‘yes’ for me but ‘no’ for you” (Street 2008: 224). By contrast, if the correctness of “X is a reason to Y for A” relativizes to A’s commitments, then there is something we can sensibly disagree about because then we have a common question: does the judgment at hand withstand the relevant procedure of

³⁰¹ Street 2008: 243. I hope to have shown that such a reading of the *Sources* is at least not implausible, even if it is not the most charitable reading available.

scrutiny from the standpoint of A's other judgments about reasons? Let us grant that this is right: the standpoint of the agent is a better fit than the standpoint of the speaker.³⁰² But is the standpoint of the individual agent really the only alternative? Before we address this question, let us review Street's account of the procedure of scrutiny itself.

To the extent that Street's account of the relation that a judgment must bear to the agent's other judgments is Humean at all, it is based on an instrumentalist reading of Hume.³⁰³ In her view, acting for reasons is a matter of acting in accordance with the instrumental principle and the basic principles of theoretical reason. Accordingly, her procedure of scrutiny is composed of the following principles (see Street 2008: 227-9):

Instrumental Principle: Whoever judges that she has conclusive reason to Y judges that she has reason to take (what she recognizes to be) the necessary means to Y.

Principle of Non-Contradiction:

- (a) Whoever judges that X is a reason to Y cannot also (simultaneously and in full awareness) judge that X is not a reason to Y.
- (b) Whoever judges that X is a reason to Y cannot also (simultaneously and in full awareness) judge that Z, which is identical with X, is not a reason to Y.
- (c) Whoever judges that only facts of kind X are reasons to Y, and recognizes that Z is not a fact of kind X, cannot also (simultaneously and in full awareness) judge that Z is a reason to Y.

These principles, Street argues, are *constitutive of* making judgments about reasons. For her, this means that someone who consciously violates these principles cannot count as making judgments about reasons at all (see sect. 2.2).³⁰⁴

³⁰² If Street had opted for the standpoint of the speaker, then she would not have been able to secure *any* objectivity. In opting for the agent's standpoint, she is able to secure *weak* objectivity (see sect. 1).

³⁰³ For a discussion of whether Hume is an instrumentalist or a sceptic about practical reason, see e.g. Hampton 1995.

³⁰⁴ Now that Street's procedure of scrutiny is in full view, we can see that, unlike Korsgaard's account of the constitutive, hers has rather counterintuitive implications. Consider an example that she herself discusses (Street 2008: 227). You judge that you have conclusive reason to go to Rome immediately and, as a matter of fact, flying is the only way to do so. If you then say that you have no reason whatsoever to get on a plane, this might be because you are just not aware that flying is the only means to your end. In that case, you are making a normal mistake. However, if you are aware that flying is the only way to get to Rome immediately and you nevertheless deny that you have any reason to get on a plane, then, on Street's view, you do not "count as genuinely making [a] 'normative judgment' at all"

To sum up, Street's answer to question a) is this: a judgment about reasons is correct if and only if it does not contradict any of the agent's other judgments about reasons and does not amount to a failure to take the necessary means to the various ends that these other judgments commit the agent to. For Street, this is not just *one* answer that a metaethical constructivist might give but rather *the answer* – the only one there is (2008: 224-30). Accordingly, she assumes that it must be the Kantian answer, too: Kantians who want to be metaethical constructivists must endorse this account of what is constitutive of agency. This explains an observation that we made when concluding sect. 2.2: that Street jumps from the claim that morality won't come from what Korsgaard takes to be constitutive of agency (which is the only claim that her criticism can actually establish) to the stronger claim that morality won't come from what is constitutive of agency, period. If we take into consideration that Street sees her account of what is constitutive of agency as the only account there is, then we can understand why, having argued that normativity does not come from what is constitutive of agency on *that* account, she feels entitled to conclude that it doesn't come from what is constitutive of agency at all. But is this the only option?

What Korsgaard and Street Agree On

Street's assumption that Kantians share her answer to question a) is confirmed in *The Sources of Normativity*. Recall our sketch of the reflective endorsement method in sect. 2.1. The reflective endorsement method demands that agents scrutinize their impulses in light of their various practical identities. An impulse passes the reflective endorsement test if and only if it proposes a judgment that coheres with the agent's practical identities. These identities are, at bottom, commitments to ends, both basic and specific. And it is in virtue of promoting some such ends and not undermining (important) others that the actions recommended by the agent's impulses are deemed worthy of pursuit or even obligatory. If this is right, then it seems that Korsgaard's reflective endorsement method is simply a more vivid representation of the *Instrumental Principle* and the *Principle of Non-Contradiction*, which make up Street's procedure of scrutiny. In a nutshell, Korsgaard seems to agree with both parts of Street's answer

(2008: 228). This assessment is counterintuitive, however, because surely what we would want to say is that you are being *irrational*. The only way to act irrationally, according to Street, is to act against one's own judgment: to act akratically (Street 2008: 228). I agree that akrasia is an instance of irrationality, but so is the Flying-to-Rome case.

to question a): the view that the correctness of judgments about reasons relativizes to the standpoint of an individual agent and that such a judgment is correct insofar as it stands in appropriate instrumental relations to their other judgments about reasons.

Why does Korsgaard opt for such a Humean account of rational agency? Here, we need to remind ourselves of some of the points discussed in ch. 3, sect. 1. In particular, we need to remind ourselves of what I called the “standing-commitment thread” of her argument against particularism: the idea that having a practical identity is a matter of having commitments that one stands by in every situation – a matter of being a principled person, for example, someone who is generally honest or generally interested in theatre. This way of thinking about practical identities, we said, is bound up with the subsumptive conception of rationality, according to which consistency in judgment and action is a matter of subsuming particular cases or actions under Case-Scope universal principles. Now if *this* is the set of concerns that one takes to lie behind Kant’s insistence on universal laws – if one thinks that his main concern is a concern with *consistency* – then it is to be expected that one’s attempt to interpret the Categorical Imperative as a procedure of scrutiny would lead to the kind of Humean account that we outlined above. To act on principles that one can will as universal laws, on this account, is to assess and embrace impulses on the basis of whether they cohere with one’s overall ends, projects and commitments, which one upholds over time and across changing circumstances.

That this is indeed how Korsgaard understands Kant’s underlying motivations is manifest in the fact that she distinguishes between the Categorical Imperative and the moral law – a distinction which she herself calls un-Kantian (1996b: 98). Recall that the former is understood as the demand to act only on maxims that one can will as universal laws *for oneself*, which is simply a different articulation of the reflective endorsement method. Korsgaard’s reading of the Categorical Imperative reveals that, for her, Kant’s concern for universal validity is in the first instance a concern for internal consistency – and this is the concern that informs her answer to question a). It is an implication of this reading that Kant’s moral law is *not* part of the Kantian story about the correctness of judgments about reasons, that it is not part of the formal procedure of scrutiny that Kantians rely on. Instead, it is a (set of) judgment(s) about what reasons we have, which Korsgaard seeks to derive from this formal procedure. It is an answer to question b), not a). So Street and Korsgaard

agree on this much: to be a metaethical constructivist is to adopt a Humean account of the correctness of judgments about reasons. Their disagreement is merely about whether this formal procedure of scrutiny entails any substantive judgments about reasons.

This result is certainly surprising, if not counterintuitive. After all, we are used to thinking that the dispute between Humean and Kantian Constructivists is a dispute *precisely* about which procedure of scrutiny is the right one, and only derivatively, if at all, about first-order judgments about reasons. In what follows, I will argue that we should not settle for this counterintuitive account of the dispute. Instead, I will put forward what I see as a more genuinely Kantian version of formal metaethical constructivism.

3.2 Kantian Formal Metaethical Constructivism: A Neglected Alternative

On closer examination, Street's taxonomy of constructivist views is not exhaustive. It is not exhaustive because Kantian Constructivists are free to reject the assumption that the concern for universal validity that underlies Kant's Categorical Imperative is, in the first instance, a concern for consistency, as Korsgaard seems to think. Instead, Kantians can agree with the conclusion that we reached in ch. 3: that Kant's concern, at this point, is a concern for *objectivity*. Recall where this insight led us: it led us to reject the Case-Scope Reading of Kant's conception of universal validity in favour of the Agent-Scope Reading. According to the latter reading, the Categorical Imperative is *not* tantamount to the coherence norm that is built into the reflective endorsement method. Instead, it points to an *alternative* formal procedure of scrutiny that Street overlooks and that emerges if we answer both parts of question a) differently: the standpoint question (whose other judgments matter?) and the relationship question (in which relationship should a given judgment stand to these other judgments?).

According to the Agent-Scope Reading, a principle of duty is universally valid if and only if it can be agreed to hold by all rational agents and for all such agents who are in the same circumstances (see ch. 3, sect. 2.5). If this is what Kant means when he speaks of universal laws in the practical domain, then the Categorical Imperative is the demand to only act on principles that all rational agents can will to be morally

viable for all rational agents who are in the same circumstances.³⁰⁵ In order to bring out how this requirement relates to the principles on Street's list, we can reformulate it as follows:

*Categorical Imperative*_(Const.): Whoever judges that X is a reason to Y presupposes that all rational agents can agree that all rational agents in the same situation would be allowed to treat X as a reason to Y.³⁰⁶

Before we consider how a procedure of scrutiny involving this principle would differ from Street's Humean procedure, let us look at an example in order to illustrate why one might think that *Categorical Imperative*_(Const.) is indeed a standard against which we measure our judgments about reasons.

Piano Lesson: My child needs a lift to their piano lesson and your child needs a lift to the hospital because they had an accident. You don't have a car, and I'm the only one who can help – but I refuse. “Why?!” you ask, in desperation. “Because I don't want my child to miss their piano lesson,” I say. I gave you a reason, or so it seems. In this kind of situation, it would be natural for you to object. “That's not a reason!” you might say.

On the face of it, you are rejecting my judgment that my child's needing a lift to the piano lesson is a reason to deny your child a lift to the hospital. However, according to Street, the final say is always mine. In her view, all you are trying to do is remind me that, given *my* other commitments, I should give priority to your child's needs. But I have made it clear that my child's musical education means more to me than your child's injuries, haven't I? What other commitments could you be referring to? Perhaps Street could say that you are reminding me of my own *moral* commitments. If this is what you are doing, then, according to Humeans, I can brush off your objection as follows: “When it comes to my children, I don't care about morality!” My identity as a mother is deeper than my moral identity, and that's that. Now, at this point, if you keep insisting that my judgment is invalid, the Humean must treat you as linguistically confused. Here one might beg to differ. Surely, one might say, you

³⁰⁵ In saying this, I am assuming that the various formulas of the Categorical Imperative are equivalent and that each refers to both the subjective aspect (can be willed by all) and the objective aspect (can be willed for all) of universal validity.

³⁰⁶ Strictly speaking, we should say: “would be allowed or obligated”. After all, the reason in question might be a moral reason. Below I will suggest how we might circumscribe the moral domain within this larger field.

are not linguistically confused.³⁰⁷ If Kantian Constructivists were to incorporate *Categorical Imperative*_(Const.) into their procedure of scrutiny, they could take this intuition seriously. They could say that my judgment is incorrect because it rests on a false presupposition: it is *not* true that all rational agents could agree that all rational agents in my situation would be allowed to treat their child's needing a lift to a piano lesson as a reason to deny someone else's child a lift to the hospital.³⁰⁸

A Kantian Formal Procedure of Scrutiny

How would a procedure of scrutiny that included *Categorical Imperative*_(Const.) among its principles compare to the Humean procedure that we outlined above? Both are *purely formal*. The Kantian procedure that we are considering does not presuppose any substantive judgments about reasons or make any assumptions about the content of our reasons. Nonetheless, it is very different from the Humean procedure.

Consider how a proponent of this procedure would answer the above *standpoint* question (the question of *whose* other judgments matter). They would say that it is the standpoint of all of us: of all rational agents insofar as they occupy the standpoint of rationality.³⁰⁹ This answer to the standpoint question is an important step on the way to establishing the *Strong Objectivity* claim. For if the failure of Korsgaard's argument has taught us anything, it is that, when it comes to constructivist accounts of rational agency, "you only get out what you put in". If one's procedure of scrutiny demands that the judgments of *individual agents* cohere with one another, then all reasons that arise from it, including the moral ones, will only apply to agents *as individuals*. So the only way to establish the *Strong Objectivity* claim is to build answerability to other agents into one's procedure of scrutiny itself: to make answerability to others part of one's understanding of what it is to act for reasons.³¹⁰

³⁰⁷ Of course, if you continue to insist, we might suspect that you are not so much linguistically confused as simply trying to manipulate me. In that case, I should regard your insistence not as a challenge to my reasoning but rather as a further fact, an obstacle, that I need to take into account. If you scream and shout, will it embarrass me so much that it would be better to give in? On this point see Korsgaard 2009: 193-4.

³⁰⁸ How do we know that they could not agree? Here we need to remind ourselves of our discussion of the Empty Formalism Objection in ch. 3, sect. 3.3, esp. fn. 192, 197 and 198.

³⁰⁹ The caveat that they have to occupy the standpoint of rationality follows from the "can" in the above rendering of the Categorical Imperative. This is a point that we touched on in sect. 1, when we commented on the qualification "as such" in the *Strong Objectivity* claim: Kantians are not interested in *actual* agreement since they acknowledge that real agents may go wrong.

³¹⁰ Carla Bagnoli makes a similar point when discussing Engstrom's interpretation of Kant (2012: 64).

The idea that reasons, qua reasons, ought to be public or shareable is an idea that many Kantians endorse (see e.g. O'Neill 1989: 28-50). In fact, Korsgaard herself is known for presenting arguments to this effect elsewhere.³¹¹ To flesh this out, Kantians often highlight that reasons are not only the subject matter of deliberation but also *objects of exchange*. We exchange them whenever we converse about what to do and how to live. As O'Neill notes, such exchanges are not "primarily expressive, something that can in principle be purely private, indeed solitary, ... [instead they are] primarily communicative".³¹² They aim at recognition or uptake by others and, in fact, at the possibility of recognition or uptake by *all* others.³¹³ Insofar as our interlocutors are rational, our attempts to give them reasons should not result in puzzled looks. If they do, we treat them as failed attempts. When discussing the *Piano Lesson* example, we saw that this is an idea that Humean Constructivists struggle to accommodate. Although they maintain that our judgments about reasons are weakly objective in the sense that we should be able to agree on what a particular agent has reason to do in a particular situation, given their individual commitments,³¹⁴ they deny that such judgments are strongly objective. As a result, they must cast your refusal to accept my reason for not taking your child to the hospital as a case of linguistic confusion, which is not a satisfactory account of what is going on in our example.

Let us turn to the *relationship* question. In which relationship, if any, should our judgments about reasons stand to the judgments of others? It is certainly a weaker relationship than the one characterized by the *Instrumental Principle* and the *Principle of Non-Contradiction*. After all, it is perfectly acceptable for you to judge that the sunny weather is a reason to hide away in your room, even though this "contradicts" your

³¹¹ In fact, one of these arguments is provided in the *Sources* (1996b: 132-42), though after the argument for the value of humanity. I've already mentioned that Street's reading of Korsgaard, which I have traced above, is perhaps not the most charitable one. But I do think that there are at least tensions between the passages of the *Sources* that we discussed above and the argument against private reasons presented later. The shareability or publicity of reasons is also discussed in other parts of Korsgaard's work, e.g. in her 1996a: 275-310 and her 2009: 177-206.

³¹² O'Neill 1989: 31. O'Neill makes this remark when discussing the public uses of reason. She argues that the toleration of such uses, on which Kant insists, is more than just a matter of non-interference.

³¹³ When commenting on Kant's distinction between private and public uses of reason, O'Neill notes that private uses of reason are uses that fail to address "the world at large" because they presuppose some contingent authority, e.g. the church. As such they are *deficient* uses of reason because they fail to provide a full justification of the claims made. O'Neill 1989: 33-7. The same issue arises for uses of reason in accordance with the procedures of scrutiny put forward by restricted constructivists.

³¹⁴ In this respect, they are in a better position than some others, e.g. emotivists. Emotivists cannot make proper sense of reasoned exchanges at all, since they regard judgments about reasons as merely expressive. They cannot explain why *any* interlocutor should do anything other than shrugging their shoulders.

neighbour's judgment that the sunny weather is a reason to go outside, or to judge that the summer sale is no reason to go shopping, even though this "undermines" your friend's end of buying a new summer dress. Others can agree that anyone in your situation would be *allowed to* treat X as a reason to Y, without having to treat X as a reason to Y themselves; they don't have to see Y as conducive to their ends or as compatible with all their commitments.³¹⁵ I think the appropriate term for the weaker relationship that *Categorical Imperative*_(Const.) picks out is "justifiability". Our judgments about reasons have to be *justifiable* in light of the judgments of other people, insofar as they take up the standpoint of rationality. But what does this mean, exactly?

We can answer this question by appealing to the framework developed in ch. 3. To say that taking yourself to have reason to Y or that treating X as a reason to Y is not justifiable to others is to say that, in Y-ing or treating X as a reason to Y, you would be making an exception *for yourself*. You would be allowing yourself to act and "reason" (or rather "rationalize" your act) in a way that you can't will to be available to other people with the same ends and projects who are in the same situation as you.³¹⁶ If this is what your Y-ing or treating X as a reason to Y would amount to, then the presupposition that all rational agents can agree that anyone in your situation would be allowed to Y or treat X as a reason to Y is false, and then you have a *moral* reason not to do so. According to the Kantian Constructivist account that I am proposing, moral reasons differ from other kinds of reasons in that they do not derive from or depend on the ends and commitments that we, as individuals, happen to have.³¹⁷ Instead, they derive from the *non-universalisability* of certain ends or policies that we might be tempted to adopt. Recall the *Piano Lesson* case. When I try to answer the question whether all rational agents could agree to my treating my child's musical education as a reason not to take your child to the hospital in this kind of situation, I do not need to think about any of my ends or commitments at all. In fact, it is crucial that I abstract from all the ends and commitments that distinguish me from others

³¹⁵ This is why Kantian Constructivists who endorse *Categorical Imperative*_(Const.) are not committed to the view that all reasons are moral reasons (see ch. 3, sect. 3.2) or, in more contemporary terms, that all reasons are agent-neutral reasons, which many people find implausible. See e.g. Nagel 1986: 152-6.

³¹⁶ You might deny this. The rich person who is enjoying certain privileges, for example, might say that they want everyone to have the same opportunities that they had in their life. But they cannot mean it, for if everyone had those opportunities, then they would not be able to enjoy the privileges that they are currently enjoying.

³¹⁷ This is a *formal* way of distinguishing the moral domain from other domains. Bagnoli endorses this way of circumscribing the moral domain when she says that "there is no one eminent domain of morality, because there is no one fixed set of interests that immediately qualifies as moral" (2012: 67).

and that I take up the standpoint of a rational agent as such. From this standpoint, I will be able to see that my judgment is incorrect. These reflections indicate that my version of Kantian Constructivism can secure the *Strong Objectivity* claim. But before I spell out how it does so, I want to cite some textual evidence to substantiate my hypothesis that Kant is a *formal* rather than a *substantive* metaethical constructivist.

Kant on Formal and Material Principles

Kant's formalism is one of the most heavily criticized aspects of his ethics. He is well known for insisting that we ought to adopt and reject principles of action solely on the basis of a purely formal consideration, namely whether we can will that they become universal laws. Whether a principle of action is a legitimate ground for the derivation of judgments about what we have reason to do on a given occasion is solely a matter of form, not content. This view, however problematic, is a view that Kant endorses time and again,³¹⁸ but on reflection it is difficult to see how Korsgaard can make sense of it. For if she is right – if the Categorical Imperative merely demands that I adopt principles that can be laws *for me*, given my various practical identities – then, in deliberating about the satisfaction of this allegedly formal requirement, I have to bring in substantive considerations. I have to see how the relevant principle squares with my ends, projects, and commitments, including the moral ones.

Unlike Street, Kant does not neglect this alternative view of morality. In fact, he uses it as a foil when developing his own view. Somewhat curiously, he would argue that endorsing a judgment solely on the basis that it withstands Street's and Korsgaard's "purely formal" procedure of scrutiny would mean grounding it on its matter, not its form. On their Humean view, to make a judgment about reasons is to claim that the action it recommends is appropriately related to one's system of ends: that it promotes some of these ends and does not undermine important others. Since such a judgment "presupposes an *object* (matter) of the faculty of desire as the determining ground of the will" (Kant CPrR 5:21), our endorsement of it is based on the way we affectively relate to the end that it promotes. According to Kant, to act on

³¹⁸ Kant says, for example: "If a rational being is to think of his maxims as practical universal laws, he can think of them only as principles that contain the determining ground of the will not by their matter but only by their form" (CPrR 5:27).

such a judgment is to act on a *material* principle of action.³¹⁹ Compare this to the version of Kantian Constructivism that we developed above. On this view, acting on a moral reason is a matter of acknowledging that alternative courses of action are not justifiable to others. So when we endorse the judgment that X is a moral reason to Y, our endorsement is not contingent on an affective relation to the object or end at which Y is directed.

Kant's reflections on why material principles of action "can furnish no practical laws" (CPrR 5:21) shed further light on the difference between the kind of universal validity that Street's metaethical constructivism can make room for and the kind that Kant is after. For Street and Korsgaard, the universal validity of moral demands is the result of an *overlap*. If there are judgments about reasons that can be agreed to hold by and for all rational agents, insofar as they are in the same situation, then this is because these judgments are entailed by the identities of each and every one of them, separately. (For Street this *may be* true, for Korsgaard it *must be*). However, following Kant, we may wonder whether such overlaps in what people value would really lead to anything that resembles moral conduct. Here is what he says:

For whereas elsewhere a universal law of nature makes everything harmonious, here, if one wanted to give the maxim [a material maxim] the universality of a law, the most extreme opposite of harmony would follow, the worst conflict, and the complete annihilation of the maxim itself and its purpose. ... In this way there results a harmony like that which a certain satirical poem depicts in the unanimity between a married couple bent on going to ruin: "O marvellous harmony, what he wants she wants too" and so forth, or like what is said of the pledge of King Francis I to the Emperor Charles V: "What my brother Charles would have (Milan), that I would also have". (Kant CPrR 5:28)

Now, of course, this is slightly unfair. Both Street and Korsgaard seek to explain why there may or must be a kind of overlap that differs from that depicted in this quote, an overlap of judgments about reasons that results in genuinely harmonious coexistence and recognizably moral behaviour. But given the failure of Korsgaard's attempt to secure the *Strong Objectivity* claim, it seems that any overlap can only ever yield *contingently* universal moral demands. In order to show that the universal validity

³¹⁹ Kant CPrR 5:22. In Korsgaard's terms, we could put it as follows: although we step back from our impulses, and ask ourselves whether to endorse them or not, the standpoint from which this question is asked is ultimately the standpoint of other impulses that we have decided to endorse in the past.

of moral judgments is *necessary* – that moral reasons bind all rational agents *as such* – we must build the demand to, as it were, aim at harmony right into our account of what it means to make judgments about reasons. This is precisely what my formal version of Kantian metaethical constructivism is designed to achieve. As such, it succeeds in establishing the *Strong Objectivity* claim.

4. Taking Stock: Conjuring Rabbits out of Hats

My aim in this chapter was to show that Street's criticism of Kantian Constructivism targets a view that is neither properly Kantian nor very convincing in its own right. Substantive metaethical constructivism, as defined by Street, tries to conjure a rabbit out of a hat. It tries to establish that the authority of moral demands extends to all rational agents as such, but then, in doing so, it appeals to a procedure of scrutiny that casts all normative demands as relative to the systems of ends and commitments of individual agents. As one might expect, this conjuring trick miscarries. Throughout my discussion, I have presented Korsgaard as an accomplice. Whether this assessment of the *Sources* is fair depends, in part, on whether she really does pursue the two projects that I distinguished above (sect. 2.1) *separately*: whether her general account of the correctness of judgments about reasons really is independent of her answer to the normative question in the way I have suggested.

The more important conclusion is, in any case, the following: if Kantian Constructivists aim to show not only that moral demands are internal to our standpoint as rational agents but also that they are strongly objective, then this aim – the aim of accounting for *objectivity* – should inform their thinking about the appropriate formal procedure of scrutiny. That this aim informed *Kant's* thinking about the supreme principle of morality and its content was one of my main claims in this thesis. In this final chapter, I hope to have shown that it helps to advance the debate between Humean and Kantian Constructivists.

Bibliography

- Allison, Henry. 1990. *Kant's Theory of Freedom*. Cambridge University Press.
- . 2011. *Kant's Groundwork for the Metaphysics of Morals: A Commentary*. Oxford University Press.
- Anscombe, G. E. M. 1957. *Intention*. Harvard University Press.
- . 1958. 'Modern Moral Philosophy'. *Philosophy* 33 (124): 1–19.
- Aune, Bruce. 1979. *Kant's Theory of Morals*. Princeton University Press.
- Bagnoli, Carla. 2011. 'Constructivism in Metaethics'. In *The Stanford Encyclopedia of Philosophy*, edited by E. Zalta, (URL = <<http://plato.stanford.edu/entries/constructivism-metaethics/>>).
- . 2012. 'Morality as Practical Knowledge'. *Analytic Philosophy* 53 (1): 61–70.
- Baron, Marcia. 1995. *Kantian Ethics Almost Without Apology*. Cornell University Press.
- Beck, Lewis White. 1960. *A Commentary of Kant's Critique of Practical Reason*. University of Chicago Press.
- . 1998. 'Did the Sage of Königsberg Have No Dreams?' In *Kant's Critique of Pure Reason: Critical Essays*, edited by H. Allison et al., 103–16. Rowman & Littlefield Publishers.
- Bennett, Jonathan. 1966. *Kant's Analytic*. Cambridge University Press.
- Bittner, Rüdiger. 1974. 'Maximen'. In *Akten des 4. Internationalen Kant-Kongresses Mainz*, edited by G. Funke and J. Kopper, 485–498. De Gruyter.
- Bojanowski, Jochen. 2012. 'Is Kant a Moral Realist?' *Kant Yearbook* 4 (1): 1–22.
- . 2015. 'Categories of Freedom as Categories of Practical Cognition'. *Kantian Review* 20 (2): 211–34.
- . 2017. 'Thinking about Cases: Applying Kant's Universal Law Formula'. *European Journal of Philosophy*. DOI: 10.1111/ejop.12307.
- Brandom, Robert. 2000. *Articulating Reasons: An Introduction to Inferentialism*. Harvard University Press.

- . 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.
- Carroll, Lewis. 1895. 'What the Tortoise Said to Achilles'. *Mind* 4 (14): 278–80.
- Cassam, Quassim. 2015. 'Persuasion'. Unpublished manuscript.
- 'Casuistry'. n.d. Merriam-Webster.com. Accessed July 3, 2017.
<https://www.merriam-webster.com/dictionary/casuistry>.
- 'Casuistry'. n.d. Oxforddictionaries.com. Accessed July 3, 2017.
<https://en.oxforddictionaries.com/definition/casuistry>.
- Cholbi, Michael. 2013. 'The Constitutive Approach to Kantian Rigorism'. *Ethical Theory and Moral Practice* 16 (3): 439–48.
- Constant, Benjamin. 1964. 'Des Réactions Politiques'. In *Écrits et Discours Politiques*, edited by O. Pozzo di Borgo. Pauvert.
- Cullity, Garrett. 2013. 'The Context-Undermining of Practical Reasons'. *Ethics* 124 (1): 8–34.
- Dancy, Jonathan. 1993. *Moral Reasons*. Blackwell.
- . 2004. *Ethics Without Principles*. Oxford University Press.
- Darwall, Stephen. 1983. *Impartial Reason*. Cornell University Press.
- Engstrom, Stephen. 2009. *The Form of Practical Knowledge: A Study of the Categorical Imperative*. Harvard University Press.
- . 2013. 'Constructivism and Practical Knowledge'. In *Constructivism in Ethics*, edited by C. Bagnoli, 133–52. Cambridge University Press.
- Enoch, David. 2006. 'Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action'. *Philosophical Review* 115 (2): 169–98.
- Ginsborg, Hannah. 2006. 'Kant and the Problem of Experience'. *Philosophical Topics* 34 (1/2): 59–106.
- . 2011. 'Primitive Normativity and Skepticism about Rules'. *Journal of Philosophy* 108 (5): 227–54.

- Hampton, Jean. 1995. 'Does Hume Have an Instrumental Conception of Practical Reason?' *Hume Studies* 21 (1): 57–74.
- Hare, R. M. 1978. 'Relevance'. In *Values and Morals*, edited by A. I. Goldman and I. Kim, 73–90. D. Reidel.
- Hart, H. L. A. 1948. 'The Ascription of Responsibility and Rights'. *Proceedings of the Aristotelian Society* 49 (1): 171–94.
- Hegel, G. W. F. 2011. *Elements of the Philosophy of Right*. Edited by A. W. Wood and H. B. Nisbet. Cambridge University Press.
- Herman, Barbara. 1993. *The Practice of Moral Judgment*. Harvard University Press.
- . 2006. 'Reasoning to Obligation'. *Inquiry* 49 (1): 44 – 61.
- . 2007. *Moral Literacy*. Harvard University Press.
- . 2011. 'Embracing Kant's Formalism'. *Kantian Review* 16 (1): 49–66.
- . 2013. 'Making Exceptions'. In *Kant und die Philosophie in weltbürgerlicher Absicht: Akten des XI. Kant-Kongresses 2010*, edited by M. Ruffing et al., 245–62. De Gruyter.
- Hill Jr, Thomas. 1973. 'The Hypothetical Imperative'. *Philosophical Review* 82 (4): 429–50.
- . 1991. 'Making Exceptions Without Abandoning the Principle'. In *Violence, Terrorism, and Justice*, edited by R. G. Frey and C. W. Morris, 196–229. Cambridge University Press.
- . 2000. *Respect, Pluralism, and Justice: Kantian Perspectives*. Oxford University Press.
- Holton, Richard. 2010. 'The Exception Proves the Rule'. *Journal of Political Philosophy* 18 (4): 369–88.
- Hooker, Brad. 2008. 'Moral Particularism and the Real World'. In *Challenging Moral Particularism*, edited by M. N. Lance, M. Potrč, and V. Strahovnik, 12–30. Routledge.
- Hume, David. 2007. *An Enquiry Concerning Human Understanding and Other Writings*. Edited by S. Buckle. Cambridge University Press.

- Hussain, Nadeem, and Nishi Shah. 2006. 'Misunderstanding Metaethics: Korsgaard's Rejection of Realism'. In *Oxford Studies in Metaethics 1*, edited by R. Shafer-Landau, 265–94. Oxford University Press.
- James, David. 1992. 'Twenty Questions: Kant's Applied Ethics'. *Southern Journal of Philosophy* 30 (3): 67–87.
- Kant, Immanuel. 1902–. *Gesammelte Schriften*. Edited by the Königlich Preussische (Deutsche) Akademie der Wissenschaften. G. Reimer (De Gruyter).
- . 1991–. *Cambridge Edition of the Works of Immanuel Kant*. Edited by P. Guyer and A. Wood. Cambridge University Press.
- Kitcher, Patricia. 2003. 'What Is a Maxim?' *Philosophical Topics* 31 (1/2): 215–43.
- Kleingeld, Pauline. 2017. 'Contradiction and Kant's Formula of Universal Law'. *Kant-Studien* 108 (1): 89–115.
- Korsgaard, Christine M. 1996a. *Creating the Kingdom of Ends*. Cambridge University Press.
- . 1996b. *The Sources of Normativity*. Cambridge University Press.
- . 2008. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford University Press.
- . 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press.
- Lackey, Jennifer. 2013. 'Lies and Deception: An Unhappy Divorce'. *Analysis* 73 (2): 236–48.
- Lance, Mark, and Margaret Little. 2006. 'Particularism and Antitheory'. In *The Oxford Handbook of Ethical Theory*, edited by D. Copp, 567–94. Oxford University Press.
- . 2007. 'Where the Laws Are'. In *Oxford Studies in Metaethics 2*, edited by R. Shafer-Landau, 149–71. Oxford University Press.
- Landy, David. 2015. *Kant's Inferentialism: The Case Against Hume*. Routledge.
- Lange, Marc. 2000. *Natural Laws in Scientific Practice*. Oxford University Press.
- Langton, Rae. 2007. 'Objective and Unconditioned Value'. *Philosophical Review* 116 (2): 157–85.

- Little, Margaret. 2000. 'Moral Generalities Revisited'. In *Moral Particularism*, edited by B. Hooker and M. Little, 276–304. Clarendon Press.
- Lyons, David. 1965. *Forms and Limits of Utilitarianism*. Clarendon Press.
- Mahon, James. 2008a. 'The Definition of Lying and Deception'. In *The Stanford Encyclopedia of Philosophy*, edited by E. Zalta, (URL = <<https://plato.stanford.edu/archives/win2016/entries/lying-definition/>>).
- Mahon, James. 2008b. 'Two Definitions of Lying'. *International Journal of Applied Philosophy* 22 (2): 211–30.
- McCarty, Richard. 1991. 'Moral Conflicts in Kantian Ethics'. *History of Philosophy Quarterly* 8 (1): 65–79.
- McDowell, John. 1994. *Mind and World*. Harvard University Press.
- . 1998. *Mind, Value, and Reality*. Harvard University Press.
- McGee, Vann. 1985. 'A Counterexample to Modus Ponens'. *Journal of Philosophy* 82 (9): 462–71.
- McKeever, Sean, and Michael Ridge. 2006. *Principled Ethics: Generalism as a Regulative Ideal*. Oxford University Press.
- McNaughton, David, and Piers Rawling. 2000. 'Unprincipled Ethics'. In *Moral Particularism*, edited by B. Hooker and M. O. Little, 256-275. Clarendon Press.
- Murdoch, Iris. 2001. *The Sovereignty of Good*. Routledge.
- Nagel, Thomas. 1986. *The View From Nowhere*. Oxford University Press.
- Newton, Alexandra. 2013. 'Hypothetical Imperative: Principle or Conclusion of Practical Reasoning?' Unpublished manuscript.
- . 2015. 'Kant on the Logical Origin of Concepts'. *European Journal of Philosophy* 23 (3): 456–84.
- Nussbaum, Martha. 1986. *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*. Cambridge University Press.
- O'Neill, Onora. 1989. *Constructions of Reason: Explorations of Kant's Practical Philosophy*. Cambridge University Press.

- . 2002. 'Instituting Principles: Between Duty and Action'. In *Kant's Metaphysics of Morals: Interpretative Essays*, edited by M. Timmons, 331–48. Oxford University Press.
- . 2004. 'Modern Moral Philosophy and the Problem of Relevant Descriptions'. *Royal Institute of Philosophy Supplement* 54: 301–16.
- . 2013. *Acting on Principle: An Essay on Kantian Ethics*. 2nd Edition. Cambridge University Press.
- O'Shea, James. 2012. *Kant's Critique of Pure Reason: An Introduction and Interpretation*. Routledge.
- Paton, H. J. 1953. 'An Alleged Right to Lie: A Problem in Kantian Ethics'. *Kant-Studien* 45 (1-4): 190–203.
- Peregrin, Jaroslav. 2014. *Inferentialism: Why Rules Matter*. Palgrave Macmillan.
- Rawls, John. 1955. 'Two Concepts of Rules'. *Philosophical Review* 64 (1): 3–32.
- . 1980. 'Kantian Constructivism in Moral Theory'. *Journal of Philosophy* 77 (9): 515–72.
- . 1999. *A Theory of Justice*. Revised Edition. Harvard University Press.
- . 2000. *Lectures on the History of Moral Philosophy*. Harvard University Press.
- Raz, Joseph. 2006. 'The Trouble with Particularism (Dancy's Version)'. *Mind* 115 (457): 99–120.
- Rippon, Simon. 2014. 'Were Kant's Hypothetical Imperatives Wide-Scope Oughts?' *Australasian Journal of Philosophy* 92 (4): 783–88.
- Ronzoni, Miriam, and Laura Valentini. 2008. 'On the Meta-Ethical Status of Constructivism: Reflections on G.A. Cohen's 'Facts and Principles''. *Politics, Philosophy and Economics* 7 (4): 403–22.
- Rosenberg, Jay. 2005. *Accessing Kant: A Relaxed Introduction to the Critique of Pure Reason*. Oxford University Press.
- Ross, W. D. 2002. *The Right and the Good*. Newly edited by P. Stratton-Lake. Clarendon Press.

- Russell, Bertrand. 1914. *Our Knowledge of the External World: As a Field for Scientific Method in Philosophy*. Routledge.
- Saul, Jennifer. 2012. *Lying, Misleading, and What Is Said: An Exploration in Philosophy of Language and in Ethics*. Oxford University Press.
- Scanlon, T. M. 2012. 'The Appeal and Limits of Constructivism'. In *Constructivism in Practical Philosophy*, edited by J. Lenman and Y. Shemmer, 226-242. Oxford University Press.
- Schapiro, Tamar. 2003. 'Compliance, Complicity, and the Nature of Nonideal Conditions'. *Journal of Philosophy* 100 (7): 329–55.
- . 2006. 'Kantian Rigorism and Mitigating Circumstances'. *Ethics* 117 (1): 32–57.
- Schroeder, Mark. 2005. 'The Hypothetical Imperative?' *Australasian Journal of Philosophy* 83 (3): 357–372.
- . 2011. 'Holism, Weight, and Undercutting'. *Noûs* 45 (2): 328–44.
- Schüssler, Rudolf. 2012. 'Kant und die Kasuistik: Fragen zur Tugendlehre'. *Kant-Studien* 103 (1): 70–95.
- Sedgwick, Sally. 1991. 'On Lying and the Role of Content in Kant's Ethics'. *Kant-Studien* 82 (1): 42–62.
- Sellars, Wilfrid. 1953. 'Inference and Meaning'. *Mind* 62 (247): 313–38.
- . 1968. *Science and Metaphysics: Variations on Kantian Themes*. Routledge.
- Setiya, Kieran. 2007. 'Cognitivism about Instrumental Reason'. *Ethics* 117 (4): 649–73.
- Sherman, Nancy. 1997. *Making a Necessity of Virtue: Aristotle and Kant on Virtue*. Cambridge University Press.
- Singer, Marcus. 1954. 'The Categorical Imperative'. *Philosophical Review* 63 (4): 577–91.
- Sinnott-Armstrong, Walter. 1999. 'Some Varieties of Particularism'. *Metaphilosophy* 30 (1-2): 1–12.
- Siyar, Jamsheed. 2013. 'The Conditionality of Hypothetical Imperatives'. *Kantian Review* 18 (3): 439–60.

- Stern, Robert. 2012. *Understanding Moral Obligation: Kant, Hegel, Kierkegaard*. Cambridge University Press.
- . 2015. *Kantian Ethics: Value, Agency and Obligation*. Oxford University Press.
- Stocker, Michael. 1976. 'The Schizophrenia of Modern Ethical Theories'. *Journal of Philosophy* 73 (14): 453–66.
- Stratton-Lake, Philip. 2000. *Kant, Duty, and Moral Worth*. Routledge.
- Strawson, P. F. 1966. *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason*. Methuen.
- Street, Sharon. 2008. 'Constructivism about Reasons'. In *Oxford Studies in Metaethics 3*, edited by R. Shafer-Landau, 207–45. Oxford University Press.
- . 2010. 'What Is Constructivism in Ethics and Metaethics?' *Philosophy Compass* 5 (5): 363–84.
- . 2012. 'Coming to Terms with Contingency: Humean Constructivism about Practical Reason'. In *Constructivism in Practical Philosophy*, edited by J. Lenman and Y. Shemmer. Oxford University Press.
- Tannenbaum, Julie. 2002. 'Acting with Feeling from Duty'. *Ethical Theory and Moral Practice* 5 (3): 321–37.
- Timmermann, Jens. 2000. 'Kant's Puzzling Ethics of Maxims'. *The Harvard Review of Philosophy* 8 (1): 39–52.
- . 2001. 'Alles halb so schlimm: Bemerkungen zu Kants ethischem Rigorismus'. *Ethik ohne Dogmen. Aufsätze für Günther Patzig*, edited by A. Stephan and K. P. Rippe, 58–82. Mentis.
- . 2009. 'Acting from Duty: Inclination, Reason and Moral Worth'. In *Kant's Groundwork of the Metaphysics of Morals: A Critical Guide*, edited by J. Timmermann, 45–62. Cambridge University Press.
- . 2013. 'Kantian Dilemmas? Moral Conflict in Kant's Ethical Theory'. *Archiv für Geschichte der Philosophie* 95 (1): 36–64.

- Timmons, Mark. 1997. 'Decision Procedures, Moral Criteria, and the Problem of Relevant Descriptions in Kant's Ethics'. In *Jahrbuch für Recht und Ethik*, edited by B. S. Byrd et al., 389–417. Duncker & Humblot.
- . 2001. *Moral Theory: An Introduction*. Rowman & Littlefield.
- Trampota, Andreas, Oliver Sensen, and Jens Timmermann (eds.). 2013. *Kant's Tugendlehre. A Comprehensive Commentary*. DeGruyter.
- Van Cleve, James. 1999. *Problems from Kant*. Oxford University Press.
- Väyrynen, Pekka. 2004. 'Particularism and Default Reasons'. *Ethical Theory and Moral Practice* 7 (1): 53–79.
- . 2006. 'Moral Generalism: Enjoy in Moderation'. *Ethics* 116 (4): 707–41.
- . 2008. 'Usable Moral Principles'. In *Challenging Moral Particularism*, edited by M. N. Lance, M. Potrč, and V. Strahovnik, 75–106. Routledge.
- . 2009. 'A Theory of Hedged Moral Principles'. In *Oxford Studies in Metaethics* 4, edited by R. Shafer-Landau, 91–132. Oxford University Press.
- . 2011. 'Moral Particularism'. In *The Continuum Companion to Ethics*, edited by C. B. Miller, 247–60. Continuum.
- . 2013. *The Lewd, the Rude and the Nasty: A Study of Thick Concepts in Ethics*. Oxford University Press.
- Walla, Alice Pinheiro. 2015. 'Kant's Moral Theory and Demandingness'. *Ethical Theory and Moral Practice* 18 (4): 731–43.
- Wallace, R. Jay. 2001. 'Normativity, Commitment, and Instrumental Reason'. *Philosophers' Imprint* 1 (4): 1–26.
- Ware, Owen. 2009. 'The Duty of Self-Knowledge'. *Philosophy and Phenomenological Research* 79 (3): 671–98.
- Wiggins, David. 1998. *Needs, Values, Truth: Essays in the Philosophy of Value*. Third Edition. Oxford University Press.
- Willaschek, Marcus. 1992. *Praktische Vernunft: Handlungstheorie und Moralbegründung bei Kant*. Metzler.
- Williams, Bernard. 1973. *Problems of the Self*. Cambridge University Press.

- . 1985. *Ethics and the Limits of Philosophy*. Harvard University Press.
- Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Wiley-Blackwell.
- Wood, Allen. 1999. *Kant's Ethical Thought*. Cambridge University Press.
- . 2008. *Kantian Ethics*. Cambridge University Press.
- . 2011. 'Review: On a Supposed Right to Lie from Philanthropy'. *Eidos: Revista de Filosofía de La Universidad Del Norte* 15: 96–117.