The Reflective Dimension of Practical Judgment: Kant, Critique, and the Revision of Moral Universals

Sabina F. Vaccarino Bremner

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2021

© 2020

Sabina F. Vaccarino Bremner

All Rights Reserved

Abstract

The Reflective Dimension of Practical Reason:

Kant, Critique, and the Revision of Moral Universals

Sabina F. Vaccarino Bremner

This dissertation defends a Kantianism that espouses an ethics of principles while emphasizing the role of *revising* moral principles. It claims that the resources to account for moral revision can be found in Kant's theory of *reflective judgment*, the cognitive power required for forming new concepts or principles on the basis of experience. Reflective judgment, on the interpretation elaborated here, involves both the incorporation of *new moral particulars* into a standing moral conceptual repertoire and the *revision of moral universals* on the basis of such alterations. The alteration of moral universals and accommodation of recalcitrant particulars are among the operations of what is termed the 'reflective dimension' of practical reasoning for Kant, alongside the moral dimension of application of *a priori* moral principle (most notably, the categorical imperative) typically taken to exhaust Kant's moral philosophy.

Incorporating a reflective dimension into practical deliberation calls for a broader virtue, namely the taking up of a critical stance to one's principles *in general,* as well as the cultivation of inventiveness towards the formation of new principles or revision of preexisting ones. Autonomy presupposes *heautonomy:* the practical requirement of universalizing a given maxim presupposes, first, that one create this maxim, with a critical awareness of its perpetual revisability.

Table of Contents

List of Figures ... iv

List of Abbreviations ... v

Acknowledgments ... viii

Dedication ... xii

Introduction ... 1

- Moral revisability and its history ... 1
- Dissertation structure ... 7

Chapter 1: Critical Theory and the Prospect of a Reflective Constructivism ... 12

- 1.1 In defense of normative instability ... 12
- 1.2 Immanent critique ... 18
- 1.3 Constructivism in metaethics ... 24
- 1.4 Reflective constructivism ... 32
- 1.5 Progress and 'partial transformations' ... 40

Chapter 2: On Moral Intelligibility ... 43

- 2.1 Obdurate particulars ... 43
- 2.2 The limitations of two frameworks ... 46
  - Hermeneutical injustice ... 46
  - Imaginative resistance ... 50
- 2.3 Intelligibility: moral and practical ... 52
  - An odd split between practices and actions ... 57
  - Moral intelligibility and critique ... 63
- 2.4 Beauvoir on moral intelligibility ... 66
- 2.5 Conclusion ... 72

Chapter 3: The Reflective Dimension of Practical Deliberation ... 74

- 3.1 Universalism and particularism ... 74
- 3.2 The problem of relevant descriptions ... 79

3.3 Responses to the problem of relevant descriptions 83
Onora O'Neill 83
Barbara Herman 84
Hannah Arendt 89
3.4 An alternate approach 91
The example of maxim construction in teleological judgments 93
The difference between 'mere' reflective judgments and the reflective dimension of theoretical and practical judgment 99
Technical practical principles and the technical activity of the power of judgment 101
Natural teleology in deriving FRE 105
3.5 On the role of maxim construction 107
Two kinds of maxims: a provisional distinction 110
I-maxims 113
P-maxims 116
3.6 Responding to the particularist 120
Chapter 4: Kant on Autonomy as Self-Making 125
4.1 The 'application presupposition' in Kantian conceptions of autonomy 125
4.2 Autonomy in the *Opus Postumum* 129
Self-making as a transition from metaphysics to physics 129
Autonomy and reflectivity in self-positing ideas and maxims of reason 132
Heautonomy 134
4.3 Self-making and heautonomy in the *Doctrine of Virtue* 139
From the denial of a practical schematism to a 'reflective faith' 143
Schematizing the anthropological and the moral 149
4.4 Reflective judgment and practical reason 151
Natural teleology in Kant's shift on self-perfection 152
4.5 Personal autonomy, moral autonomy, and self-critique 158
Chapter 5: Kant on Moral Ideas and Virtue 161
5.1 The right and the good: virtue ethics and Kantian ethics 161

5.2 Ideas of reason in the *Critique of Pure Reason* ... 165
Are regulative ideas principles of reason? ... 170
Is the content of regulative ideas relevant? ... 172
5.3 Ideas of reason in the *Critique of Judgment* ... 175
5.4 Moral ideas ... 179
Aesthetic ideas and thick concepts ... 179
The 'development' of moral ideas ... 181
Moral ideas as working notions guiding maxim formulation ... 184
Moral ideas and moral archetypes ... 186
Bootstrapping ... 189
5.5 Virtue as meta-idea ... 193
Virtue and moral revision ... 194
Conclusion: Virtue as Critique ... 197
A metanormative stance on normative revisability ... 197
The end of history ... 201
Moral regression ... 204
Systematicity and intelligibility ... 207
References ... 212

List of Figures

Figure 1: Transcendental and empirical maxims produced by teleological judgment 97

Figure 2: The 'transition project' in the metaphysics of natural science and the metaphysics of morals 142

Figure 3: The two dimensions of practical judgment 159

List of Abbreviations

Aristotle

References to Aristotle are to the standard Bekker pagination. English translations are drawn from *The Basic Works of Aristotle*.

NE: *Nicomachean Ethics*

Kant

References to Kant are to the volume and page of *Kants gesammelte Schriften* (except citations to KrV, which are to the A and B pagination) and accord with the following abbreviation scheme:

ApH: *Anthropologie in pragmatischer Hinsicht* / *Anthropology from a Pragmatic Point of View* (vol. 7)

B: *Briefwechsel* / *Correspondence* (vol. 12)

EE: "Erste Einleitung" in die *Kritik der Urteilskraft* / "First Introduction" of the *Critique of Judgment* (vol. 20)

GMS: *Grundlegung zur Metaphysik der Sitten* / *Groundwork of the Metaphysics of Morals* (vol. 4)

HN: *Handschriftlicher Nachlaß: Logik* / *Handwritten Manuscripts: Logic* (vol. 16)

IaG: "Idee zu einer allgemeinen Geschichte in weltbürgerlicher Absicht" / "Idea for a Universal History with a Cosmopolitan Aim" (vol. 8)

KaU: *Kritik der ästhetischen Urteilskraft* / *Critique of Aesthetic Judgment* (Part I of the *Critique of Judgment*) (vol. 5)

KpV: *Kritik der praktischen Vernunft* / *Critique of Practical Reason* (vol. 5)

KrV: *Kritik der reinen Vernunft* / *Critique of Pure Reason*

KaU: *Kritik der teleologischen Urteilskraft* / *Critique of Teleological Judgment* (Part II of the *Critique of Judgment*) (vol. 5)

KU: *Kritik der Urteilskraft* / *Critique of the Power of Judgment* (vol. 5)

LJ: *Logik Jäsche* / *Jäsche Logic* (vol. 9)

MAN: *Metaphysiche Anfangsgründe der Naturwissenschaft* / *Metaphysical Foundations of Natural Science* (vol. 4)

OP: *Opus Postumum* (vol. 21-22)

MS: *Metaphysik der Sitten* / *Metaphysics of Morals* (vol. 6)

RGV: *Die Religion innerhalb der Grenzen der bloßen Vernunft* / *Religion Within the Bounds of Reason Alone* (vol. 6)

TL: *Tugendlehre* / *Doctrine of Virtue* (Part II of the *Metaphysics of Morals*) (vol. 6)

TP: "Gemeinspruch: Das mag in der Theorie richtig sein, taugt aber nicht für die Praxis" / "On the Common Saying: That May Be True in Theory, But It Is of No Use in Practice" (vol. 8)

VA: *Vorlesungen über Anthropologie* / *Lectures on Anthropology* (vol. 25)

VE: *Vorlesungen über Moralphilosophie* / *Lectures on Ethics* (vol. 27)

English translations are drawn from the *Cambridge Edition of the Works of Immanuel Kant*, with some modifications by the author.

Other sources

All other sources are cited according to author name and initial date of publication (or composition, in the case of posthumous publication). The full citations can be found in the final List of References, where the initial date of publication or composition is followed by the publication of the cited edition in brackets. Texts divided into fragments, such as Wittgenstein's *Philosophical Investigations* or Nietzsche's *Gay Science,* are cited by section number.

Acknowledgments

This work represents my current state of thinking on topics such as Kant's ethics, normativity, critical theory, genealogy critique, lived possibilities for action, and their relations. That I was given relatively free rein to address the wide-ranging issues I do in what follows in the space of a dissertation is a testament to the open-minded environment of the Columbia University philosophy department, as well as the expert guidance of my committee—the best context for someone of my particular intellectual temperament to flourish.

I feel so fortunate to have been able to study under my two dissertation advisers, Patricia Kitcher and Axel Honneth. Pat was unwavering in her support, starting from the moment I first approached her about advising a decidedly un-Kantian topic. She left room for me to pursue my own interests and imbued me with confidence, while providing me with a model for how to undertake rigorous historical scholarship that I continue to follow. I couldn't imagine a more conscientious and attentive adviser; in fact, I find myself at a loss for words to adequately express my gratitude for the amount of care she has shown me, and how appreciative I feel. Axel has impressed on me that historical scholarship can, and should, *matter* to the contemporary world. He gave me space to allow my interests to grow while still prompting me to unify my often scattered thinking into a cohesive thesis. I am indebted to him for pushing me at precisely the times, and in precisely the ways, in which I needed it most. It is because of him that I feel able to say that this work represents *me*.

Wolfgang Mann took me under his wing from the time I arrived at Columbia. He had faith in me well before I had faith in myself, and I would not have succeeded without his guidance and advice. I always think of Wolfgang and I as sharing an intellectual wavelength, although perhaps it's more a testament to his open-mindedness: no matter what I was interested in at any given time, whether it was Kant, ordinary language philosophy, Aristotle, or critical theory, I could always count on Wolfgang to push my thinking forward and show interest in what I had to say. He made me feel that there was space for someone like me in philosophy.

I was immensely fortunate to get the chance to work with Amia Srinivasan at Oxford University last year, who welcomed me with open arms. Amia's meetings with me were, it's no exaggeration to say, transformative. I always felt like we were getting to the bottom of deep philosophical issues, during which Amia had a magical ability to articulate just what I was fumbling for the words to say. I credit her with helping me to find the hidden soul of the project (to the extent, of course, that I have succeeded in doing so).

Over the time I have spent writing this dissertation, I have been deeply moved by the willingness shown by others to comment on chapter drafts and engage me philosophically about my work. Often these individuals were incredibly busy, but still were generous enough to volunteer their time to help me improve. I would like to thank Taylor Carman, Fred Neuhouser, Bernard Harcourt, Katja Vogt, Christia Mercer, Angela Breitenbach, Michelle Kosch, Francey Russell (pre-committee!), Bernard Berofsky, Andrew Cooper, Billy Eck, Janum Sethi, and Sophie Cote. I was fortunate to present parts of this work at the University of Cambridge, University College London, Princeton University, New York University, and the Eastern Meeting of the American Philosophical Association, and I am indebted to audiences on each of those occasions for their helpful comments and questions.

One of the things I have loved most about my time in graduate school was the friendships it left me with: Laura Martin, Helen Zhao, Jake McNulty, Natalia Rogach, Richard Booth, Conor Cullen, Shivani Radhakrishnan, Thimo Heisenberg, and Connie Wang. My marathon discussions with Tuomo Tiisala helped me to clarify my philosophical questions from an early stage of graduate school. I also wish to thank those who encouraged me before I entered graduate school, and who helped to shape the philosopher I would become: Arnold Davidson, Candace Vogler, and Cullen Sacha. My greatest debt is to Daniele Lorenzini for his love and support.

This work is dedicated to my family: my parents Viola Vaccarino and James Douglas Bremner, and my brother Dylan. I am indebted to them for their sense of humor, their spirited political debates, and their many eccentricities. While they may have been apprehensive that their daughter and sister was studying something as impractical as philosophy, they never showed it. They never failed to encourage and believe in me.

The research that has informed this dissertation has been generously supported by funding from the Council for European Studies, the American Council for Learned Societies, and the Mellon Foundation. All errors, of course, are mine.

Ce qui m'effraie, dans l'humanisme, est qu'il présente une certaine forme de notre éthique comme un modèle universel valant pour n'importe quel type de liberté. Je pense que notre avenir comporte plus de secrets, de libertés possibles et d'inventions que ne nous laisse en imaginer l'humanisme. (Foucault 1982: 1601)

What worries me about humanism is that it presents a certain form of our ethic as a universal model valid for any given kind of freedom. I think our future holds more secrets, more inventions, and more forms of possible freedom than humanism allows us to imagine.

There is a place… for a sort of contemplation of the Good... [of] a source of *new* and quite undreamt-of virtue. (Murdoch 1970: 383)

To my parents, Viola Vaccarino and James Douglas Bremner,

and brother, Dylan Bremner

Introduction

Moral revisability and its history

This dissertation defends a Kantianism that espouses an ethics of principles, but also emphasizes the role of *revising* moral principles. It claims that the resources to account for moral revision can be found in Kant's theory of *reflective judgment*, the cognitive power required for forming new concepts or principles on the basis of experience. Reflective judgment, on the interpretation elaborated here, involves both the incorporation of *new moral particulars* into a standing moral conceptual repertoire and the *revision of moral universals* on the basis of such moral discoveries.

In so doing, I take up the longstanding problem of particulars and universals. In moral philosophy, this problem has generally been understood as a debate between two opposing historical camps. Universalism (or generalism), usually associated with Kant, holds that moral reasoning takes place through the application of *a priori* moral principles (universals) to cases in experience (particulars); particularism, most often associated with Aristotle, is the view that moral reasoning involves grasping moral particulars, making moral principles secondary, if not outright dispensable. The interpretation I develop in the ensuing pages stresses that Kant's position cannot be relegated to one faction among these two.

In this Introduction, I want to begin by advancing a broader claim regarding how contemporary philosophy comes to inherit its history. For Aristotle, too, emphasizes the role of

moral universals *along with* moral particulars, claiming that virtue involves acting "as the right rule [*orthos logos*] prescribes" (NE 1114b).[1] Aristotle's view of universals, however, eludes the terms of the contemporary debate. He claims, "All law is universal but about some things it is not possible to make a universal statement which shall be correct", namely when a case is found that does not fit the law as initially stated (1137b). In such a case, the universal must admit of "correction… where it is defective owing to its universality" (1137b). As a result, "the rule also is indefinite, like the leaden rule used in making the Lesbian moulding; the rule adapts itself to the shape of the stone and is not rigid, and so too the decree is adapted to the facts" (1137b). Aristotle sees universals as indispensable components of practical judgment, but ones whose very nature—their universality—can lend itself to 'defects' requiring correction by particulars. The right understanding of universals, then, is one on which they are 'indefinite', adapting themselves 'to the shape of the stone', or to the particular moral facts. Our (indefinite) universals guide us to consideration ("perception")[2] of particular features, but, in conjunction with this claim, "universals are reached from the particulars" (1143a-b).

[1] See also: "The brave man feels and acts according to the merits of the case and in whatever way the rule [*logos*] directs" (1115b); "the good-tempered man tends to be unperturbed and not to be led by passion, but to be angry in the manner, at the things, and for the length of time, that the rule [*logos*] dictates" (1125b-1126a); "the temperate man is… the sort of person that the right rule [*orthos logos*] prescribes" (1119a). Curzer (2016: 57n) notes the changes in translation trends of *orthos logos* from the 20th to 21st centuries; while this term was often translated as 'rule' in the 20th century, all but one translation in the 21st century uses 'reason' or 'reasoning' (see also Moss 2014). Indeed, the above citations are drawn from the Ross translation of the early 20th century. Curzer argues that Aristotle does provide rules, speculating that contemporary trends in particularism and virtue ethics may explain why 'rule' came to fall out of favor. Irwin (2000) provides another helpful defense of the generalist aspects of Aristotle's views.
[2] Though, here, the perception at issue is rather different than moral salience (an account of moral reasoning I discuss further in Chapter 3): Aristotle claims that practical knowledge does not consist in "perception of qualities peculiar to one sense but a perception akin to that by which we perceive that the particular figure before us is a triangle"—a mode of perception akin to the synthetic *a priori* knowledge Kant *also* attributed both to mathematical and to moral reasoning, rather than *visual* perception of empirical features (1142a).

This aspect of Aristotle's view of moral universals has attracted less attention than other facets, perhaps because it eludes the strictures of the universalism-particularism debate. Yet I want to say something similar about what Kant, on his mature view, held moral reasoning to be. Like Aristotle, Kant has a sophisticated view of how moral universals can take on the new 'shape' of particulars which resist assimilation into the discursive resources available; like Aristotle, Kant takes universals to ultimately be 'indefinite'.

As the examples of Kant and Aristotle show, how we appropriate history matters: what comes to be emphasized in historical interpretation lives on in contemporary philosophical endeavors. Consider Miranda Fricker's conception of 'hermeneutical injustice', or "the injustice of having some significant area of one's social experience obscured from collective understanding owing to a structural identity prejudice in the collective hermeneutical resource" (2007: 155). Fricker claims that hermeneutical injustice is neither strictly ethical nor epistemological, but straddles both domains. It therefore calls for a hybrid virtue, both intellectual and ethical, that is a "reflexive critical awareness" on one's own "patterns of credibility judgment" (92, 85). Insofar as prejudice gets in the way of gleaning truth, epistemic justice is an *intellectual* virtue; insofar as prejudice interferes with the good, namely pursuing justice, epistemic justice is an *ethical* virtue.

Fricker classifies the intellectual-ethical virtue of epistemic injustice as one best situated among the accounts of Aristotelian virtue and moral perception articulated by philosophers such as John McDowell, Martha Nussbaum, Iris Murdoch, and Jonathan Dancy, on which virtue requires a prior process of cultivation, or being brought up or socialized in such a way that the world acquires a kind of "moral colour" (Fricker 2007: 72). On these views, the virtuous individual should be able to intuitively *see* the right thing to do without protracted deliberation or applying rules, rendering virtue uncodifiable. The virtue of epistemic justice, then, consists in

recognizing the *particularity* of one's interlocutor and the information they impart. It takes up, then, the problem of particulars and universals, conceiving of practical reason in particularist terms.

While such views have rightly drawn attention to the necessary role of moral sensitivity or salience—that practical deliberation depends in part on what we *take* to be practically relevant, and thus on what we notice on the basis of the case to which we respond—one major deficiency is their attribution of virtue acquisition to upbringing and education alone.[3] First, counterexamples readily present themselves: individuals who would seem to lack the requisite moral education who turn out to be quite morally sensitive, and vice versa. Second, as Herman (2007) has also suggested, if virtue is largely the result of prior inculcation or socialization, it seems unclear where the *critical reflexivity* of the hybrid virtue Fricker describes is to come in. As Dancy puts it, "for us it is probably too late" to acquire the right kind of moral sensitivity: "As Aristotle held, moral education is the key; for those who are past educating, there is no real remedy" (1993: 64). There is no distance, then, between education and our considered moral outlook. Our capacity to be virtuous is the result of what we've been *taught* to believe virtue is and requires of us—and how well we've been taught it. But if our grasp of morality is primarily the product of prior social understandings of moral relevance, it seems unclear how we can obtain a *critical* stance on those social understandings and on our own constitution as agents in the world.

The beginnings of an alternate perspective can be found articulated in the ethical writings of Iris Murdoch. Murdoch puts emphasis not so much on education and upbringing *per se,* but

[3] See, e.g., McDowell on moral upbringing and *Bildung* (1978: 21, 1996: 84), and Nussbaum on moral education (1990: 102).

on the "deepening process" of experience (Murdoch 1970: 322). This shift in emphasis brings her account closer to my own view, particularly in her characterization of moral experience as "progress in understanding a scheme of concepts" in which "as we move and as we look our concepts themselves are changing" (1970: 325, 321). Importantly, however, this characterization moves Murdoch's view away from Fricker's claim that the virtue of epistemic justice is 'uncodifiable', which Fricker takes to mean a lack of reliance on advance moral rules.[4] Indeed, Murdoch's characterization rests on a questioning of the dichotomy between the codifiable, or predetermined and fixed general rules, and the uncodifiable, or moral sensitivity to the particular. Instead, Murdoch suggests that our moral experience, while shaped by the moral principles and conceptual schemes at our disposal, can nevertheless directly *inform* those conceptual schemes if we remain sufficiently receptive to new moral content. While she does not elaborate the point, these stray remarks on the potential for 'progress' in one's conceptual repertoire appeal to a two-way causal relation between rules and particulars, alluding to a position more complex than the particularist opposition between the particularities of experience and the generalities of a conceptual scheme.

The position Murdoch alludes to, which I will construe in terms of the operation of reflective judgment, also bears on contemporary accounts of critical theory. Here camps have

[4] See Fricker (2007: 73). Indeed, three pages later Fricker cites Murdoch approvingly as having elaborated a view of ethics as 'uncodifiable'. As I see it, however, Murdoch never claims that moral codes, or principles, are dispensable; instead, she makes reference to what I would term the *reflective* aspect of practical reason. As Bagnoli claims, Murdoch is no particularist: "Murdoch agrees with the particularist that we should not think of moral deliberation as a mechanical application of rules, but so might the universalist…. That moral reasons are not generated by the application of an algorithm does not imply that they are not universal, intelligible as reasons, and even shareable by others" (2011a: 223). Bagnoli concludes that Murdoch is a universalist, albeit one of a peculiar stripe: Murdoch affirms the process by which "abstract and empty universals are enriched, and therefore changed, by their embodiments, that is, by the ways in which they are expressed and manifested" (Bagnoli 2011a: 222).

been staked out between those who search for universalist 'normative foundations' to justify social critique, and those who reject foundationalism altogether, where the latter position has typically been taken as necessarily committed to particularism or, less charitably, relativism. Coloring the reception of such antifoundationalist views in critical theory has been the apparent untenability of their value-theoretical position; one of the major weaknesses of such views, it has often been held, is their underlying normative confusion. Even where Kantian reflective judgment has already been influential in such discussions,[5] it has been understood only in *particularist* terms, as bringing out a contrast with Kant's universalism in his moral philosophy.

Here, too, then, discussions about normativity have foundered on the particular-universal distinction. Without catching sight of the prospect that moral universals admit of *change*, as both Aristotle and Kant recognized, an important alternative has been missed. The alteration of moral universals is among the operations of what I will term the 'reflective dimension' of practical reasoning for Kant, alongside the moral dimension of applications of *a priori* moral principle (most notably, the categorical imperative) typically taken to exhaust Kant's moral philosophy.

In this dissertation, I concur that the intellectual-ethical virtue called for by the lacunae in our normative resources (in what I term 'moral intelligibility', or the current scope and depth of our moral conceptual repertoire) must take a critical, reflexive form. Yet I do not take this virtue to be limited, as in epistemic justice, either to instances of second-personal engagement with a structurally disadvantaged interlocutor or to 'patterns of credibility judgment' more generally. Instead, I argue that incorporating a reflective dimension into practical deliberation calls for a broader virtue, namely the taking up of a critical stance to one's concepts *in general*, as well as the

[5] See discussion of neo-Arendtian critical and political theory in Chapter 3.

cultivation of a kind of inventiveness vis-à-vis the creation of *new* concepts or revision of preexisting ones. Autonomy, I argue, presupposes *heautonomy:* the practical requirement of universalizing a given maxim presupposes, first, that one *create* this maxim and endorse it as fully one's own, with a critical awareness of its perpetual revisability.

Dissertation structure

In **Chapter 1**, I discuss the question of foundationalism in critical theory, or whether social critique needs normative foundations, and if so, how they might be conceived. I propose moving from a metaphor of normative *stability* to one of *instability*, or, better, *revisability*. I situate the genealogical tradition in post-Kantian philosophy as an important precursor to this way of thinking about normativity, an approach often rejected as confused, nihilistic, or relativist; instead, I take genealogists' efforts as directed towards the *revision* of moral universals. I conclude that this endeavor is consistent with the position of constructivism in metaethics, which takes normative facts to be constructed relative to the perspective of the valuing agent (rendering them metaphysically *subjective*, but epistemically *objective*), particularly when we consider that the coherence of a given evaluative perspective is contingent on how one's sense of the non-normative facts bears on one's value judgments. Since genealogy and ideology critique typically task themselves with redescribing the non-normative facts insofar as they bear on our values or practices (providing a new historical narrative regarding how a particular value or practice came about, or a redescription of a state of affairs in the current social order), they demonstrate how our evaluative perspective can be altered, and with it, our values themselves.

This discussion serves to set up my own position: I advance an account of a 'critical reflexivity' towards our own sense of moral relevance, one that can be taken to elaborate on

Murdoch's buried insight that moral progress consists in part in the alteration of our normative conceptual schemes. In **Chapter 2**, I term the target of this alteration 'moral intelligibility': what shows up to us as moral; what makes sense to us in moral terms. A particular which can be accommodated by an available universal is *morally intelligible*; one which resists subsumption and calls for revision is *morally unintelligible*. I raise three examples that evoke what it is like to be presented with a feature that is morally unintelligible, and that, accordingly, demonstrate when the revision of principle is needed. I take up two frameworks which might be brought to bear on this phenomenon, epistemic injustice and imaginative resistance, to demonstrate why I take moral intelligibility to elude either account. I discuss how accounts of social practices in social theory differ from accounts of practical reason in normative ethics, situating moral intelligibility at the intersection of these two ways of treating socially situated action. I then relate moral intelligibility to the project of *critique*, or more precisely, to the normative function of critique. To show this, I draw attention to Simone de Beauvoir's critique of the way in which moral universals are deficiently brought to bear on women as a secondary class in patriarchal society.

The account I develop of the critical reflexivity towards our sense of moral relevance is broadly Kantian, but it relies on a reading of Kant that departs significantly from the conventional reading of Kant's views. First, in the historically focused chapters of the dissertation, I do not limit myself solely to Kant's canonical moral texts, the *Groundwork* and the *Critique of Practical Reason*. Instead, I also take seriously later texts, such as the *Critique of Judgment* and the *Opus Postumum*, which I argue were broadly *transformative* of Kant's account of practical judgment. On my reading, Kant came increasingly to acknowledge that practical reason also involves a *reflective* dimension, by which maxims are formulated and reflexively posed to oneself ('self-given'). Moreover, I argue, his account of the formation of teleological judgment in

scientific inquiry provides us with a model of reflection as contingent and perpetually revisable.

In **Chapter 3**, I introduce this account of the reflective dimension by way of discussing why Kant's universalism has been rejected by 'critical' philosophers of race, gender, and class. Their criticisms, I claim, can be said to reiterate in part what Elizabeth Anscombe articulated as the 'problem of relevant descriptions': the permissibility of a moral maxim depends on how the agent describes it, such that, if there are moral features that escape the agent's notice (or her *discursive* resources), applying the categorical imperative to the maxim so described will not catch the moral deficiency of the action. I explain why I take several prominent responses to this problem, as advanced by Onora O'Neill, Barbara Herman, and Hannah Arendt, respectively, to fall short of resolving it. I then introduce the power of reflective judgment, the faculty needed when "only the particular is given, for which the universal is to be found" (KU 5:179). I focus on reflective judgment under its *teleological* guise, explaining how Kant takes teleological judgment to generate new, empirically sourced maxims to guide scientific investigation. In the moral domain as well, I claim that we generate practical maxims of two kinds, both of which, I argue, must be conceived as empirically informed: i-maxims, or intention maxims, maxims which pick out a determinate action in a determinate context, and p-maxims, or principle maxims, which articulate general moral rules that can be applied to many circumstances. Reflective judgment governs the *reflection on circumstance* that figures into i-maxims, as well as the *revision of moral principles*, or p-maxims.

In **Chapter 4,** I claim that Kant's account of reflectivity poses the need to incorporate his characterization of *heautonomy,* or the inventive formation and reflexive positing of principles (the giving of a principle to oneself), into his traditional conception of practical *autonomy,* by which a given principle is legislated to all rational beings. I show this by examining the presence,

or lack thereof, of 'autonomy' in two texts written after the *Critique of Judgment*: the *Opus Postumum* and the *Metaphysics of Morals*. While the *Opus Postumum* treats the autonomy of *reason* rather than moral autonomy strictly speaking, the *Metaphysics of Morals* does not invoke the notion of autonomy at all—despite the presupposition among scholars that the latter text simply has to do with Kant's *a priori* moral framework, including the principle of autonomy, to empirical cases. Instead, I put emphasis on Kant's claim in MS that a similar 'transition' to the one OP aims to secure between metaphysics and natural science is needed in the moral domain. On my reading, this transition should be conceived, in both domains, as being secured through reflective judgment and its distinctive, self-given normativity (*heautonomy*).

In **Chapter 5**, I return to the notion of p-maxims as I had earlier advanced it. I claim that Kantian principles, or p-maxims, must be reconceived in light of what Kant came to call 'moral ideas': where principles express *prescriptive* moral content ('act virtuously', 'be just', 'treat every human being as an end in itself'), ideas express *descriptive* moral content ('virtue', 'justice', 'humanity'), with the two bound up with each other (in what I term a 'bootstrapping' relation). As a result, the deontology/virtue-ethics distinction—as typically applied to Kant and Aristotle, respectively—is faulty. Kant, too, has a rich account of thick ethical concepts: a normative repertoire, he claims, which admits of 'development'. Kant also holds, accordingly, that moral principles, their prescriptive aspect, also admit of alteration. The relation between moral ideas and moral principles therefore comes to resemble the account Kant advances in KrV of regulative ideas 'guiding' empirical inquiry, a normative role that can ultimately be expressed through principles or maxims of thought—which, in KU, is accorded to the power of reflective judgment.

I conclude by drawing attention to Kant's account of virtue, which I characterize in terms of critique, and by returning to the question of moral progress. Rejecting the Hegelian 'end of

history', as I take the intrinsic contingency of reflective judgment to position us to do, heightens the importance of the *critical* aspect of Kantian virtue becomes all the weightier.

Chapter 1: Critical Theory and the Prospect of a Reflective Constructivism

1.1 In defense of normative instability

Much debate in critical theory has recently turned on the issue of *normative grounding.* What is requisite for social critique, it has been frequently suggested, is a stable ground or foundation from which normative claims can be validated. Jürgen Habermas stipulates a determinate procedure, that of a public discourse, to adjudicate common norms (1983, 1990), while Axel Honneth suggests the model of ascertaining that an ideal already immanent in a given social context "embodies progress in the process of the realization of reason" in order to show that "it can provide a justified standard in order to criticize the given social order" (2001: 9). As David Owen puts it, the common criticism stemming from the recent generations of the Frankfurt School takes the form of a shared assumption: that critique should involve "the articulation of normative criteria concerning the justification of moral norms and/or the legitimacy of social institutions and practices" (2002: 224). Social critique, on these tellings, must involve the stipulation of stable normative criteria if it is not to be nihilist, relativist, or irresponsible. To sum it up, the ontological picture of normativity relied on is one on which normativity is *fixed:* it provides us with a ground, with a foundation, with stability.

This dissertation fundamentally takes issue with this way of conceiving of the relation between normativity and critique, albeit by thinking along with a seemingly strange bedfellow: Immanuel Kant. On the view of moral normativity I will be presenting, normativity is not

inherently stable; normativity instead has to do with the tension between the ideal and the real, with a gap between the world as it is and how it ought to be. That normativity involves an inherent *tension* between the two, however, lessens the prominence in ethics of the ideal.[6] Instead, as social reality changes, so too do the normative notions with which we criticize the world, a process I term 'moral revision' and explain by recourse to Kant's conception of reflective judgment in practical reason. Normativity, then, on the view I present here, is better thought in terms of a perpetual process of *destabilization.*[7]

The second and third generations of the Frankfurt School's positioning of normative criteria as a precondition for critique stems in part from an interpretation of the first generation. Figures including Theodor Adorno, Max Horkheimer, and Hannah Arendt experienced National Socialism as demonstrating that "the meaning of normative ideals or principles proved

[6] See O'Neill (1987) and Mills (2004) on the problems with equating value theory with *ideal* theory. As Mills observes, the dominant emphasis in ethics on *ideal theory,* on accounts positing ideal reasoners and perfectly just societies, has obscured the *non*ideal from view, namely the experiences of oppressed and marginalized peoples. The problem, as Mills articulates it, is not with abstraction, generalization, or universalization per se, since recently coined morally weighted concepts, such as 'white supremacy' or 'patriarchy', serve as useful abstractions of real social phenomena, productively redirecting moral attention in kind. Instead, the problem comes when we conflate abstraction or universalization with *idealization*, abstracting away the empirical, frequently unjust features of the actual social world. While ethics has tied itself to ideal theory to the exclusion of contending with the ways in which social positioning complicates its picture of moral deliberation, 'critical' theories of race, gender, and class have distanced themselves from the resources of generalism in ethics, committing themselves to particularist or relativist strategies that complicate the possibility of criticizing conditions of oppression (Mills 2004: 173-4). One aim of my account is to carve out normative space for a generalism informed by particularity, resituating the force of critique in terms of the revision of principle.

[7] My account, which I situate below alongside the genealogical and 'hermeneutics of suspicion' traditions, might therefore be termed 'destabilizing', as distinct from the 'stabilizing' view of critical theory advocated by Habermas and taken up by many of his antecedents. The dispute here on the relation between normativity and critique could perhaps productively map onto current debates on the role of genealogy as *debunking* (as in Nietzsche and Foucault) and *vindicatory* (as in Edward Craig's (1990) state-of-nature epistemology, taken up also in Williams (2002) and Fricker (2007): 109-128) modes of genealogical argument. This debate has broadened out into larger questions involving the epistemological function of genealogy critique (Queloz 2018, 2019; Srinivasan 2019), while also giving rise to new camps, e.g. 'problematizing' (Koopman 2013, Allen 2016) and 'possibilizing' (Lorenzini 2020) interpretations of Foucault's genealogies.

themselves to be much more porous and open-ended—indeed, more vulnerable—than the original [Hegelian] program of critique envisioned. The consequence is the insight that a moral norm cannot prescribe how it may be put to social use" (Honneth 2001: 10). That is, this generation of thinkers noticed the phenomenon by which the professed application of a given moral principle could betray the content of this principle in its effects. Nazism was evil not just because of the horrors it inflicted, but also, for these thinkers, because it attempted to mask them behind the invocation in theory of moral principles which were openly contradicted in practice. In analyzing how social evils like those perpetuated by the Third Reich were able to achieve mass acceptance, social theorists needed to grapple with this apparent corrosion of moral value. Critique therefore tasked itself with pointing out tensions between the explicit content of values adhered to in social morality and the way in which they are applied in practice.

This process drew from the 'hermeneutics of suspicion' tradition traditionally identified with Marx, Nietzsche, and Freud, which viewed the stated meanings of social values 'with suspicion' by showing that they function antithetically to their surface meaning (Foucault 1964, Ricœur 1965). In the vein of thought represented by these thinkers, norms are seemingly hollowed out of their significance altogether.

As a result, the second and third generations of the Frankfurt School have frequently rejected the model of critique-as-suspicion as normatively unstable, ultimately bottoming out in the moral nihilism taken to be represented by the Nietzschean and Foucauldian genealogical approaches. Nietzsche and Foucault, on this appraisal, criticize a social order "by showing the means and processes whereby bounded ideas and norms become historically legitimated disciplining or repressive practices"—showing, in other words, that moral justification is merely a pretext for practices of domination (Honneth 2001: 7). While figures such as Marx and Adorno

at least "betray the retention" of certain ideals (Honneth 2006: 342), Nietzsche and Foucault are characterized as "consistent skeptics" about normativity in general, whose genealogical investigations into the origins of our values are taken to constitute a *reductio ad absurdum* argument against appealing to any values at all (Habermas 1983: 99).

Considering the genealogical tradition as a whole—under whose umbrella I would group not just the projects of Nietzsche and Foucault, but also Adorno and Horkheimer's genealogy of enlightenment rationality, Marx and Engel's genealogy of the German Ideology, and Simone de Beauvoir's genealogy of the second sex—I instead draw an alternate conclusion.[8] Rather than taking this tradition to represent an ethical deficiency, as on the normativity-as-stability model, we might consider it on an alternate picture: that of the *revisability* of normative understandings. When Foucault demonstrates that an Enlightenment value, such as treating prisoners humanely, is applied in a way that results in the self-constitution of prisoners (and eventually the broader social whole) as self-disciplining subjects, thereby undercutting their autonomy, this need not be understood as a *reductio* argument against the very application of the Enlightenment principles of

[8] Karen Ng distinguishes Adorno from Foucault on the basis of Foucault's apparent suspension of moral judgment ("Foucault consistently and deliberately abstains from moral prescriptions, and would not ever ground moral claims on the basis of bodies and how they are molded by specific power regimes", Ng 2020: 57). I think this is precisely the wrong reading of Foucault, for reasons I have begun to suggest here but cannot systematically defend. Foucault refrains from advancing explicit moral judgments not because he feels they cannot be drawn, but because he is suspicious of all attempts by figures of authority to draw exhaustive moral conclusions, given the ways in which such discourse can always be coopted; the moral function of his texts consists in their effects on readers' reasoning processes rather than in their explicit normative claims. Ng also argues that Adorno's moral conclusions are indexed to nature, and therefore normatively grounded in a way that Foucault's are not; she references Adorno's invocation of the destruction of nature by human activity, or of the natural needs of human beings. I think this is similarly insufficiently nuanced; a recurrent theme of Adorno's writings is that violence is already inherent in the natural order: female non-human animals, for instance, "undergo love in unfreedom, as objects of violence" (1951: §55). I mention these points because they are characteristic of the attempts to enlist Adorno in the foundationalist critical project (and Foucault to the antifoundationalist), obscuring more interesting possibilities.

autonomy or humanity.[9] Instead, we might understand genealogical argument to be directed towards the aim of *revising* our values, in this case broadening the typical scope of application of autonomy, or showing how the application of 'humanity' in practice undermines the content of this principle.

In this dissertation, I attempt to spell out what this means when it comes to Kantian principles, but we can, at a first pass, understand it to take several forms: the understood scope of a given principle; the entailment from principle to meaningful action; the possibility of *new* modes of meaningful action; making particular features of one's action-context newly morally relevant. The creative and historical efforts of the genealogist (or ideology critic)[10] are, in my view, directed at our sense of the *apparent necessity* of ingrained, habitual ways of understanding our values and realizing them in action. The approach undertaken is not that of a moral nihilist who aims merely to 'debunk' our values; it is, in itself, a moral task—what I will attempt to characterize as an instantiation of a virtue, the virtue of critique—whose aim is to productively

[9] Indeed, Foucault was quite explicit about his commitment to the value of autonomy, as inherited from the Enlightenment tradition. See Foucault (1984), Allen (2008), Tiisala (2017).

[10] The two categories of genealogy and ideology blur together, in my view, in the works of Adorno, Marx, Foucault, and other critical theorists. I would characterize genealogy as a method of argument, while ideology critique makes an epistemological claim: that a given picture of reality is a false one. Ideology critics can employ genealogical modes of argument (see Marx and Engels 1846, Adorno and Horkheimer 1944), while there can be notions of 'critique' that involve neither genealogical method nor an epistemological commitment to dispelling ideology (see Foucault 1984a on critique as philosophizing one's present moment, which would constitute one example). Foucault in particular was also often critical of the epistemological commitment to true and false consciousness presupposed by Marxist conceptions of ideology; see Foucault (1973). But since I take genealogy, particularly on Foucault's own conception of it, to nevertheless involve a pragmatist commitment to the possibility of a 'better' (more accurate, helpful, useful, or productive) description of the historical unfolding of our current representational practices, I take the two traditions of genealogy and ideology critique to be compatible with one another, and the site of disagreement to bottom out in Foucault's deflationary conception of truth; on these issues, see Horwich (1998) for skepticism towards the notion of truth in the analytic tradition. All of this is less relevant to the *normative* implications of genealogy and ideology critique. Given the prominence of Frankfurt School critiques against the genealogical tradition and comparative acceptance of the ideological tradition, I concentrate largely on genealogy in what follows.

subject our understanding of value to revision.

It is a philosophical commonplace that the nature of rule-following—of which intentional action is often taken to be an instantiation—gives way to a regress if merely explained in terms of the application of rules, and must therefore be situated within the context of a community or social context in which meaning is defined (Wittgenstein 1953, Winch 1965, Kripke 1982). Rule-following, that is, cannot solely be understood in terms of application; it must be contextualized within a collective normative practice.[11]

While Kant is often taken to be one of the main proponents of rule application in both concept use and moral action, in my view, Kant also came to recognize the limitations of this model in his account of *reflective judgment*. Kant's theory of reflective judgment was intended to explain how new 'universals'—empirical concepts as well as normative maxims of thought—are generated on the basis of particulars encountered in experience that fail to accord with existing conceptual schemes. But the theory showed, more generally, how the working concepts or heuristic principles we start out with are inevitably *informed* (justified, disconfirmed, fleshed out) by the particulars we come across in our ongoing practice of judgment. The normativity of judgment is therefore ultimately *subjective* and *reflexive:* this is a normativity we confer on ourselves (*self-given*), and therefore always defeasible, subject to revision. Our reception of a given particular is, of course, guided by the universals we start out with, but the universals are

[11] The dispute regarding *destabilizing* and *stabilizing* varieties of critical theory raises larger philosophical issues, as discussed in ordinary language philosophy, phenomenology, and other currents, around the role of skepticism and feeling at home in our practices: see Egan (forth.) on these themes in Wittgenstein and Heidegger, Davidson (1997: 1-20) on comparison of Wittgenstein and Foucault, and Crary (2018) for sympathetic discussion of Frankfurt School immanent critique in light of Wittgensteinian considerations. For my part, I see the project of figures in the hermeneutics of suspicion tradition, such as Marx and Foucault, as that of intentionally alienating ourselves from our practices, so that we are put in a position to newly see what is wrong with them such that we can subject them to change.

themselves often mutually informed by the new and unsubsumable particulars. The apparent stability of normativity, then, is a kind of social illusion (Harcourt 2020), one the critical theorist can productively break up.

1.2 Immanent critique

What I've said so far may seem no different than the Hegelian strand of contemporary critical theory, which has embraced the prospect of *immanent critique,* or criticizing social practices on the basis of the criteria already latent within them (Honneth 1992: 43, 2014; Celikates 2009; Jaeggi 2013). One might think that openness to moral revision is already implicit in the notion of immanent critique. And there is a sense in which this is right: the picture of normativity I am presenting here by way of a certain interpretation of Kant can be understood as a Kantian model of immanent critique—although one fundamentally different in method than the other two principal Kantian models currently operative in critical theory, which I term the *particularist* and *generalist* (or *proceduralist*) models and further disambiguate my position from below.[12]

On my view, moral revision works by way of what I call a 'bootstrapping' process, where initially simple resources can be employed to generate increasingly complex results. It is often thought that particular moral features must be interpreted as such in light of existing normative resources: it has been widely acknowledged, in a point I return to in Chapter 3, that knowing when moral action is called for depends on which features of an agent's circumstances are interpreted as morally relevant, and therefore on a faculty for moral perception that eludes the

[12] See discussion in §3.4 on particularist neo-Arendtian or neo-Kantian critical theories which emphasize the role of moral example. These accounts, in light of their particularism, might be contrasted with the proceduralist Kantian approaches I devote my attention to in this chapter.

mere application of moral rules. But, as I claim is suggested in Kant's account of reflective judgment, it must be possible for 'new' particulars, which have yet to be taken as morally relevant or interpreted in moral terms, to lead to the formation of new moral concepts and principles or newly inform preexisting ones. Thus, in an example presented by Miranda Fricker as a case of hermeneutical injustice, an experience initially understood by some as an instance of lighthearted office fun given existing normative resources can come to be reinterpreted as a new morally weighted concept: 'sexual harassment' (Fricker 2007).

As I construe it, this particular experiential feature which fails to fit into a preexisting moral conceptual repertoire does not merely aid in the birth of 'sexual harassment' as a new thick ethical concept. In so doing, it also deepens our understanding of more basic moral principles (allowing us to better understand the 'semantic depth' of our moral concepts: Platts 1988, Moody-Adams 1999), say of moral worth or respect for persons, or 'moral ideas' like autonomy, virtue, or justice (Chapter 5). If it can be shown, for instance, that the repetitive tasks required by the implementation of the division of labor and the extension of the working day corrode agents' possibilities for self-realization, our conception of the full conditions required for self-determination changes. I take this to be what the work of critique is all about: a critique of preexisting moral conceptual schemes can be brought about by a *crisis*—as the neologism 'sexual harassment' was coined by feminist lawyers in response to the case brought forward by Carmita Wood against her workplace (Fricker 2007: 150)—or it can be intentionally motivated by the efforts of the genealogist or ideology critic: comparing aspects of contemporary values with ones historically held; emphasizing tensions between the stated content of a given value and its function in practice; destabilizing a given set of normative claims by demonstrating their ignoble

origins.[13]

While my view of 'moral bootstrapping' can therefore be counted as a mode of immanent critique, the conclusions I draw about the prospects for moral revision are of a different nature than those typically advocated in contemporary critical theory. Both the prevailing neo-Kantian and Hegelian approaches appeal to a *stable normative principle* which, by being demonstrated as immanent to our preexisting social context (or presupposed in our current practices), can then be employed in the service of critique: for example, the principle of recognition (Honneth 1992), or the principle of dialogue (Habermas 1983) or justification (Forst 1994) as a procedure for settling substantive normative disputes.

On my view, the lesson to be learned from the prospect of critiquing currently held values is not the need to ward off normative skepticism or establish the validity of our normative claims, since for me, the genealogist or ideology critic is not to be interpreted as a skeptic, but as a moral revisionist. The implication of seeing our normative resources as perpetually revisable—as ontologically *unstable*—is that we need not defend the fixed nature of, say, a determinate procedure, such as the institutionalization of a common discourse that hinges on a fictive "right to universal access to, and equal opportunity for participation in argumentation [that] is en-joyed freely and equally" (Habermas 1983: 163). The appeal to a procedure, in Habermas as in Rawls, is intended to ward off substantive normative commitments (Rawls 1980: 540). But the prospect of

[13] Or, as Foucault (1971) would urge us to say, their 'beginnings'. Talk of origins in genealogy often brings up concerns about the genetic fallacy, or the argument from the illicit origin of a concept or practice to the impugning of its current validity (see Geuss 1981, Srinivasan 2019). On my view, genealogical modes of argument generally have more to do with emphasizing aspects of the *current* function of the genealogized concept, aspects that are hard to discern without the process of historical distancing from our enmeshment in habitual patterns of usage that genealogy helps to bring about. The 'illicit origin' of a concept is secondary for Nietzsche—and, in Foucault, usually plays no role whatsoever in the genealogical course of argument.

free and equal participation in language has been pointed out, by critical theorists such as Nancy Fraser as well as contemporary philosophers of language, to be an extremely demanding, if not unrealizable, condition to meet in practice.[14] Fraser points to the example of empirical research on patterns of communication in heterosexual couples which shows that men "tend to control conversations, determining what topics are pursued, while women do more 'interaction work' like asking questions and providing verbal support", contributing to a structural power dynamic of gendered dominance and subordination further reinforced by differing gestural, bodily, and vocal tendencies (1985: 108). The points raised against Habermas in the critical theory and feminist traditions have been reiterated in the philosophy of language literature, which has repeatedly drawn attention to the role linguistic phenomena, including the introduction of presuppositions (Lewis 1979), silencing (Hornsby and Langton 1998), and subordination (Langton and West 1999), play in perpetuating structural power imbalances.[15]

The cases raised in both continental and analytic traditions cast doubt on the prospect Habermas champions of an idealized public discourse by which common norms might be adjudicated. But they also demonstrate a narrowness in Habermas's conception of what discourse is, one that, if broadened, would encompass practices such as genealogy critique which he considers to be excluded from the purview of moral argument. Habermas claims that the 'consistent skeptic' represented by Nietzsche and Foucault's genealogical approach to normative questions "refus[es] to argue", instead taking "the attitude of an ethnologist vis-a-vis his own culture, shaking his head over philosophical argumentation as though he were witnessing the

[14] A point also noticed by Honneth (2000: 88) in relation to Habermasian discourse ethics.
[15] That said, we should not just pay attention to *who* can speak and who cannot—we should also examine what *is* said, what individuals are *brought* to say; defending this point here, however, would take me too far afield.

unintelligible rites of a strange tribe" (1983: 99-100). Instead, these rhetorical strategies attempt to break up preexisting discursive patterns with all of the power relations habitually bound up in them. Were the conditions for an idealized discourse to be institutionally secured—putting aside methodological questions of how all of those affected could each have a chance to participate equally, and be listened to equally, in the discourse, particularly for broad-scale issues affecting the global population at large—it does not follow that the historical contingency of the concepts available to be utilized in the course of dialogue would thereby be altered. Discursive strategies, such as genealogy, literature, or humor, can help here to distance us from our received categories and break up habitual patterns of moral application.[16]

Miguel de Beistegui raises a similar point about recognition: "There cannot be anything as straightforward as the recognition of what people 'really are.' The reason is that what people really are, the categories through which they identify themselves... of gender, ethnicity, nationality, race, sexuality, are generated not by the oppressed or the subordinated, but by the systems of power that generate the subordination" (2018: 187). Beistegui's criticism poses a problem for Honneth if he considers recognition to be a stable principle under which all other normative problems can be subsumed. Instead, if taken to be revisable, the principle calls for an altogether different attitude: one of epistemic humility towards the notion of recognition itself.[17] There is a

[16] See Iris Marion Young's critique of Habermas' position that "there appears to be no role for metaphor, jokes, irony and other forms of communication that use surprise and duplicity" (1987: 71), or Honneth: "Habermas must implicitly ignore all those potentialities for moral action which have not reached the level of elaborated value judgments, but which are nonetheless persistently embodied in culturally coded acts of collective protest" (2000: 83). Indeed, Honneth goes so far as to affirm the role of genealogy in social critique: "No social critique is possible today that does not use genealogical research to uncover shifts in the social meaning of their leading ideals" (2001: 11).
[17] As Beistegui puts it: "A critical attitude to recognition, then, would not mean one that rejects it altogether, but rather one that seeks to ask about the necessary limits, and even the operation of exclusion, that constrain it, and one that, in turn, demands its own transformation" (2018: 190).

sense, then, in which contemporary conceptions of immanent critique are not Hegelian *enough,* if Hegel advocates the idea that our philosophical vocabularies admit of continuous 'sublation' such that we should always hold open the prospect that more sophisticated terms might dynamically be made available.[18]

Whether the picture we assume is one of normative *stability* or *instability,* then, has important methodological implications. If we take our moral principles to be fixed, we will see genealogical attempts at moral revision as threats to those principles rather than as constructive attempts to reformulate our principles, pushing us to see their current applications in a new light. Because the problem is often that our habitual patterns of thought and action fail to rise to the level of conscious reflection or discourse—these being patterns wholly transparent to us when not made explicit as such—there is no one *procedure* that can be secured to exhaust all possible ways in which our principles might be productively revised. Indeed, I hold that our moral practices continue to evolve in contingent and often unintentional ways, a point I touch on again below (§1.5).

Honneth has become increasingly more explicit as to the role of open-ended moral revision: "The 'ethos' of ethical life [die 'Sitten' der Sittlichkeit] is not rigid and fixed once and for all but exhibits a certain flexibility and reflective corrigibility that leaves room for revisions in the light of new insights" (2014: 823). What I go on to say about the reflective dimension articulates what this process involves on the part of practical reason. Yet, it should also be acknowledged, coming to recognize the prospect of moral revision alters essential premises of the critical theory endeavor. The critiques of universal principles such as ideal discourse or recognition can no

[18] This puts aside Hegel's commitment to historical progress, which I do not adhere to; see discussion below.

longer be understood as inherently nihilistic or relativist attacks by moral skeptics. Normativity can no longer be pictured on the metaphor of ground or foundation, but comes closer to a relation of constant tension with the current makeup of the social world that remains open to alteration.[19] The critical task changes from stipulating normative criteria to assessing the right metanormative attitude to take towards moral revision.

1.3 Constructivism in metaethics

This picture may appear relativist, but it need not. The Kantian account of immanent critique I am proposing falls in line with a certain interpretation of Kantianism as a form of moral constructivism, a middle path between realism and relativism on which moral facts are constructed by ethical reasoners. This is a metaethical view that, in my estimation, can be productively brought to bear on much post-Kantian thought in the continental tradition, including critical theory (in part due to the influence of Kant himself, from whom the metaphor of construction originates—of which I say more below).[20]

Constructivism has traditionally been understood in terms of the stipulation of a *procedure*.[21] Indeed, the notion of a procedure is invoked, both in Rawls' procedure of the original position and in Habermas' procedure of an ideal discourse, in order to make sense of the mind-

[19] The picture of morality as in continuous tension with the current social order is one well-articulated by Adorno: "We all chafe at the narrow limitations imposed by prevailing ideas and existing circumstances and resent the assumption that these in some sense already embody the good life…. The illusion that culture and the adaptation of the individual to culture brings about the refinement and self-cultivation of the individual" constitutes a form of "ideology", "whereas culture stands opposed to moral philosophy and is actually open to criticism from that quarter" (Adorno 1963: 10-11).
[20] See Rawls (1980) and O'Neill (1989) for more on the metaphor of construction in Kant.
[21] "The constructivist is a hypothetical proceduralist. He endorses some hypothetical procedure as determining which principles constitute valid standards of morality.... [He] maintains that there are no moral facts independent of the finding that a certain hypothetical procedure would have such and such an upshot" (Darwall, Gibbard, Railton 1992: 140).

dependent objectivity of normative facts.[22] Both Rawls and Habermas explain the metaphysical status of substantive conclusions about what agents should do as following from the procedure they each stipulate of how these normative principles are *constructed*: through a procedure of choosing normative goods behind the veil of ignorance or pruning one's judgments in reflective equilibrium; or through an institutionally secured public discourse in which all those affected are permitted to speak. In each case, normative facts result from the stipulated procedure of construction: "Apart from the procedure of constructing the principles of justice, there are no moral facts" (Rawls 1980: 519).[23]

More recent attention in metaethics, however, has distanced the stipulation of a procedure from the core of the constructivist view. Instead, the notion of a procedure can be reconceived as merely a "heuristic device" meant to bring out what is truly distinctive about constructivism: the notion of the practical point of view and what follows from within it (Street 2010: 366). On this revised constructivist position, normative facts are constructed from within a practical point of view, that is, the point of view of a valuing creature insofar as it values anything at all.[24] That is,

[22] See Darwall, Gibbard, Railton (1992) and Street (2009) for critical discussion of whether Rawls' proceduralism amounts to a distinctive metaethical position.

[23] However, in interpreting Kant, Rawls distinguishes the procedure of the categorical imperative (CI-procedure) from the moral law and the categorical imperative, respectively: "Recall that the moral law, the categorical imperative, and the Cl-procedure are three different things. The first is an idea of reason and specifies a principle that applies to all reasonable and rational beings whether or not they are like us finite beings with needs. The second is an imperative and as such it is directed only to those reasonable and rational beings who, because they are finite beings with needs, experience the moral law as a constraint. Since we are such beings, we experience the law in this way, and so the categorical imperative applies to us. The Cl-procedure adapts the categorical imperative to our circumstances by taking into account the normal conditions of human life and our situation as finite beings with needs in the order of nature" (Rawls 1989: 82).

[24] Although Korsgaard's view is substantively different from Street's, Korsgaard holds a similar view regarding normative commitments as arising for any creature who takes up an evaluative standpoint, which may be expressed as, most basically, the ability to experience suffering, or of seeking to avoid pain; see Korsgaard (1996, 2018).

"any creature who takes at least some things in the world to be good or bad, better or worse, required or optional, worthy or worthless, and so on" adopts an evaluative standpoint: "the standpoint of a being who judges, whether at a reflective or unreflective level, that some things call for, demand, or provide reasons for others" (Street 2010: 366).[25] What dictates the fact of the matter regarding what such a creature *should* do is not a given procedure, but what follows from her evaluative standpoint, or her set of values, in combination with the non-normative facts.

Sharon Street's construal of constructivism tends to treat a given valuer's evaluative standpoint as fixed: "Contingencies have shaped what reasons we have in the same way that contingencies have shaped who we are. But now that we're here, we can shed our normative reasons no more easily than we can shed ourselves" (2008: 245). Implicit in the prospect of *critiquing* our values, however, is the idea that our evaluative standpoint can be changed, as when Nietzsche proclaims that the philosophical task is to "*create values*" (1882: §401, 1886: §311).[26]

Nietzsche's thesis of value creation may seem *prima facie* implausible, and has often been

[25] I would suggest that Street's reading of Rawlsian theory—situating the practical point of view as central and relegating the role of procedure to mere 'heuristic' —can also be applied to Habermas. Habermas, too, motivates his own account of discourse ethics on the basis of an argument of what follows from within the practical point of view. What makes his view distinctive is his argument that the phenomenology of *reactive attitudes* shows us that our moral psychology is embedded within the lifeworld of communicative participation with others—within practices of justification and reproach, or discursive practices which are themselves morally evaluable. That is, unlike constructivist views represented by Rawls, Korsgaard or Street in the analytic tradition, Habermas' proceduralism insists from the outset on the *shared* nature of giving and asking for reasons. However, for the reasons I indicated above regarding the inextricability of power from discourse, I don't think Habermas goes far enough in insisting on the sociality of practical reasoning.

[26] Nietzsche's claims along this vein, e.g., that "nature is always value-less, but has been given value at some time… and it was *we* who gave and bestowed it" (1882: §401)—have recently, and plausibly, been taken as an endorsement of the central constructivist thesis that values are constructed by valuers (Silk 2015). Much as Nietzsche's metaethical views can be productively interpreted in terms of constructivism, Street's argument against realism (Street 2006) could be helpfully construed in terms of a genealogy of the evolutionary processes of natural selection which have shaped value judgments, one which 'debunks' realist accounts of value.

treated as such in the literature: "The new philosophers create values not in the sense that they can invent or elicit new evaluative attitudes at will. This would be psychologically implausible and philosophically suspect" (Silk 2015). The thesis of value creation has therefore often been parsed in terms of philosophers' ability to instill in readers new affective responses, as artists do (Clark and Dudrick 2007, Silk 2015).[27] This mode of response seems to me to perpetuate a crude dichotomy between reason and feeling, thereby recalling Habermas' claim that Nietzschean genealogy is wholly outside the domain of philosophical argument, on the order of fiction or aesthetic production.[28]

While I don't wish to understate the aesthetic dimension and import of Nietzsche's writing, what seems to be missed in these understandings of value creation is that critiquing the current scope and application of moral concepts, as Nietzsche does, *alters the space of reasons*. That is, more fully understanding the current function of a given moral term does not merely make us *feel* a certain way, as if the role of feeling can be untethered from our sense of the justification of

[27] The burgeoning literature on Nietzsche's metaethics employs a distinction between 'value' in the normative sense (*endorsing* a value) and 'value' in the descriptive sense (the empirical study of what objects are *posited as valuable*, which does not involve the endorsement of the value in the question). This leads to formulations such as, "By coming to value new things in the descriptive sense, the new philosophers can thereby create new values in the normative sense" (Silk 2015). I fail to understand how one can 'value' a new thing descriptively: the original set of distinctions, posited in Richardson (2004) and Clark and Dudrick (2007), precludes using 'value' as a verb in the descriptive sense, since 'to value' can only mean 'to esteem, to endorse'. It therefore seems to me that Silk's thesis explaining how values can normatively be created trades on the two senses of 'value' at issue: the idea of 'valuing' new things descriptively mixes the two senses, even as they are purported to be disambiguated. A recent critical assessment of Silk's view, Lambert (2019), reiterates a similar ambiguity, equating value in the descriptive sense with social or anthropological facts about what values hold in a given culture (I'm with him so far), but dissociating this sense from value in the normative, 'metaethical' or 'metaphysical' sense—a distinction that presupposes that social facts have no metaphysical status (Searle 1995, Haslanger 2014, and many others, have emphasized that they do), or that values understood anthropologically cannot admit of a metaethical constructivist interpretation. The employment of this distinction, then, seems to create more confusion than it resolves.

[28] Of course, this claim itself implies that art and literature do not involve argument, but I leave this point aside.

and reasons for our behaviors and attitudes. Instead, it *changes* what reasons hold true for us. This appears as 'implausible' and 'philosophically suspect' in the literature only because of a misplaced application of the fact/value distinction, e.g. the assumption that the normative facts (as constituted by our evaluative perspective) can be held separate from the non-normative facts (both what we take them to be, and which ones we take to pertain to the situation in question). Instead, our sense of the normative and non-normative facts mutually inform one another. Consider the following scenario:

> Anne inhabits an evaluative perspective on which moral respect is a value to be accorded to all reasoners.
>
> Anne applies the concept 'reasoner' on the basis of what she takes to be a scientifically valid scheme of apportioning rational traits to different groups: physiognomy.
>
> Anne limits the scope of application of the value 'respect' to members of only those groups which exhibit the cranial structures with which traits associated with rationality are physiognomically identified. Those with smaller or bumpier brains are not accorded full respect.

While this example alludes to real historical instances of scientifically justified sexism and racism, there was more at work, historically speaking, in the *actual* relation between oppression and failed application of moral principle.[29] Nevertheless, the example serves to show how assuming a fact/value separation can lead us to overlook the possible relations between the two. It demonstrates that a misconception of the non-normative facts (or a *false belief,* of which, when it comes to the historical misapplication of moral principle, an endorsement of physiognomy as a

[29] One might therefore object that, given the many other bases on which to attribute rationality than physiognomy, Anne is mistaken, if she is truly committed to rationality and all that it entails, to fail to attribute it more broadly than those who lack the requisite physiognomic traits. That is, her commitment to rationality would compel her, were she coherent, to seek more evidence than physiognomy alone. Street (2009) suggests a line of response along these lines in discussion of 'ideally coherent' anorexics, Caligulas, etc. On the other hand, there were, of course, also more methods throughout history of misrepresenting the non-normative facts in the service of oppression than just physiognomy.

valid science is only one example) can affect the scope and application of our moral concepts, and with them, our evaluative standpoint more generally.

In other words, critique functions by giving better, more complete descriptions of the application of normative concepts: in the case of genealogy, offering an explanation of how a given concept or practice emerged, by offering a more accurate, helpful, or useful description[30] of the contingent alteration of the social practices in which a given normative concept (sexuality, discipline, bad conscience) came about; in the case of ideology critique, by offering a redescription of a contemporary institution or practice—one that often purports to realize a given value—in terms of the real functions it serves. In either case, the critique operates by altering our sense of the non-normative facts: it gives us a more accurate characterization of the workings of the social world.[31] But, in so doing, it also alters our understanding of our values, and with them, our evaluative standpoint. The relation here, as in the example of Anne's commitment to physiognomy above, consists in the fact that incomplete or false non-normative descriptions of

[30] Thus I do not entirely agree with Owen (2002), who claims that genealogy does not claim to offer a *true* description of historical events, but to broaden the domain of applicability of true-or-false claims (see also Queloz 2018, 2019, who reads fictional genealogies as a mode of conceptual engineering). While I agree with the latter clause, many genealogies, Beauvoir's and Foucault's among them, *do* make a claim to be better (more accurate, more productive) characterizations of historical alterations than other historical narratives; this is what separates, e.g., Foucault's genealogies from Edward Craig's (1992), William's (2002), or Miranda Fricker's (2007) state-of-nature genealogies, which are purely fictional thought experiments which make no claim to be truth-apt. The painstaking historical analysis Foucault undertakes matters; the force of the genealogy's conclusions would be altered if the historical narrative were simply made up. Nevertheless, the implications of any given genealogy remain even if a given historical detail is later found to be inaccurate. As Owen claims, genealogy *also* functions to broaden the domain of truth-or-falsity (not just altering the domain of what the non-normative facts *are*), drawing attention to what Hacking has termed the historical "conditions of possibility for ideas" or the emergence of "candidates for truth or falsehood" (2002: 76, 1996: 393). Broadening our sense of the scope and possible application of our normative concepts, bringing us to entertain their domain, definition, and current function, can still help us to productively interrogate our concepts and practices.

[31] In the rest of this dissertation, the role of the non-normative comes in at the level of reflection on context, or on empirical particulars which bear on our moral response.

social reality are situated as the locus of a genealogy when they bear on the historical application of normative concepts, rendering the application inconsistent, say, or failed. Where this is the case—where a mistaken view of the non-normative bears on the normative, or where there are tensions in our scheme of normative values—our evaluative standpoint is incoherent. To recall an example cited above, where our evaluative standpoint commits us to the value of autonomy, genealogical attention to the effects of our imprisonment and surveillance practices demonstrates that our attempted application of this value results in effects antithetical to it.

Genealogy critique can be conceived as a normatively motivated redescription of historical events, one which corrects the non-normative facts about an unjust or less than ideal state of affairs, or which introduces new possible concepts to our repertoire, ones which might enable us to alter this state of affairs. So conceived, genealogy critique can be both *normatively motivated*, but *non-normative* (in terms of the description it offers), and, in so doing, thereby work to resolve points of evaluative incoherence.[32] The effect of a genealogy, if successful, is therefore an alteration of our sense of the non-normative facts that also affects, by way of casting skepticism on the moral justification of our current practices, our evaluative standpoint, and with it the values (principles, judgments, and the relations between them) we construct.

I've so far talked of rendering our evaluative standpoint more *coherent*, which is consistent with Street's own commitments. Nevertheless, even this claim is incomplete, as is talk of correcting false beliefs (the non-normative facts) alone. It can indeed be the case that treating groups as not capable of rationality actually results in their constitution as 'less rational' on operative understandings of it—less educated, more childlike, and so on—than their oppressors.

[32] See Street (2009, 2010) on coherence as a criterion for normative objectivity in her brand of 'Humean' constructivism.

Mary Wollstonecraft, as part of her critique of eighteenth-century patriarchy, denounces the fact that "all the writers who have written on the subject of female education… have contributed to render women more artificial, weak characters than they would otherwise have been"; "she was created to be the toy of man, his rattle, and it must jingle in his ears whenever, dismissing reason, he chooses to be amused" (1792: 48, 59).[33] That is, it's not just that patriarchal writers *assume* women aren't rational; they actually, in part, *make* them less rational.[34] Thus, if an agent takes women to be irrational in such a context and therefore deserving of lesser treatment, there may be nothing *incoherent* about his evaluative standpoint, once it is fully spelled out. A successful critique must therefore advance not only what it understands to be the *right* description of a given state of affairs, but also that there is something *wrong*, something *immoral,* about what is occurring. What may be required, ultimately, is not just a better understanding of the state of affairs in question, but an alteration of the values we bring to bear in assessing it: our evaluative standpoint may, then, not be made simply more coherent, but may also be itself changed altogether.[35]

Changing our evaluative standpoint is a prospect presupposed by genealogy critique. But

[33] Of course, the critique is also advanced by way of Wollstonecraft's own instantiation of her own fine reasoning and argumentative abilities, a living counterexample to the sexist presumptions of the literature on female education she is here excoriating. But this point does not alter Wollstonecraft's central claim: that the conditions for rational cultivation have failed to be met for the vast majority of women under the operative conditions, and that this constitutes a massive social failing.
[34] As Haslanger claims, "'successful' ideology isn't always false" (2017: 4), a point also made by Jaeggi (2009: 69) and Srinivasan (2019: 145). Hacking's (1996) conception of the 'looping effect' of human kinds demonstrates that the mere application of a given category onto a human being can have effects on their own self-understandings and self-constitution, and therefore always has "moral connotations"—and implications (368).
[35] And this is why the role of history in genealogy does not merely show how a given practice *evolved* to be as it is today, but also to demonstrate new possible concepts, practices, or values—ones from other times and places, since forgotten—that might be incorporated into our contemporary repertoire, enabling us to make new features morally intelligible (Chapter 2).

this is not to be understood as the 'implausible' suggestion that this can occur 'at will', nor as a merely affective or psychological process that works on the psychic drives, as Silk (2015) and Clark and Dudrick (2007) have articulated it; nor is it the result of estranging oneself from practices of giving and asking for reasons, or suspending normative judgment altogether, as Habermas (1983) charges. Instead, it is one produced through the particular mode of philosophical argumentation genealogy represents, and the alteration of the space of reasons it therefore motivates. By putting us into a position to question the justification of our practices and to entertain new possible concepts or practices, it alters the reasons we take to hold true for us: it *eliminates* certain reasons that have motivated a given practice, while also introducing *new reasons*, say by suggesting the possibility of a new practice or making a new morally weighted concept intelligible. When successful, it gives us an altered conceptual repertoire by which our reasons can be formulated.[36]

1.4 Reflective constructivism

The prospect of altering our evaluative standpoint is not one currently entertained in the metaethics literature: as we saw, Street takes one's evaluative standpoint to be fixed, and similarly, proceduralist Kantian varieties of constructivism take fixing the evaluative standpoint to be the goal of procedure stipulation. While the constructivist tradition therefore accepts one of the guiding premises of immanent critique—that values emerge only out of a given normative standpoint—it has not yet investigated the prospect of *transforming* that standpoint from within, the central aim of immanent critiques throughout the post-Kantian tradition.[37]

[36] See §3.2 on the problem of relevant descriptions and its relation to reason formulation.
[37] Here, I also include Nietzsche, Foucault, and the project of genealogy more generally within the scope of immanent critique, insofar as these approaches share the aim of transforming a given normative

This feature of the constructivist position as fixed might derive from a certain way of understanding the central metaphor of construction, a metaphor originally stemming from its invocation in Kant's *Critique of Pure Reason.*[38] Kant claims in the Preface that in order to leave "the ground of experience, we should, through careful enquiries, assure ourselves as to the *foundations* of any building that we propose to erect" (A3/B7, my emphasis). The 'foundation' of the building at issue has been taken to refer to the categorical imperative: the notion that both thought and action is to have a public, universal character (O'Neill 1989), or a "suitably constructed social point of view that all can accept" (Rawls 1980).

O'Neill (1989) rightly notes that "a critique of reason must be a *reflexive* and *political* task" (9). But what has been neglected is that the process of construction, on Kant's conception, goes on to alter our sense of the composition of the foundation itself. This is, in part, what I take Kant to have realized in discovering the principle of reflective judgment, which adds a new cognitive power to the 'building materials' first set out in the first *Critique.* But this prospect is already alluded to in the Architectonic of Pure Reason, where Kant claims that scientific inquiry proceeds on the basis of initially indistinct ideas that at first seem to be only a 'mere confluence of aggregated concepts, garbled at first but complete in time':

> Nobody attempts to establish a science without grounding it on an idea. But in its elaboration the schema, indeed even the definition of the science which is given right at the outset, seldom corresponds to the idea [*entspricht das Schema… sehr selten seiner Idee*]; for this lies in reason like a seed, all of whose parts still lie very in-voluted [*eingewickelt*] and are hardly recognizable even under microscopic observa-tion…. The founder and even his most recent successors often fumble around *with an idea that they have not even made distinct to themselves* [*um eine Idee herumirren, die sie sich selbst nicht haben deutlich machen*] and that therefore cannot determine the special content, the articulation (sys-tematic unity) and boundaries of the

standpoint from within it. The second and third generations of the Frankfurt School, however, generally excise these figures from the immanent critique tradition.

[38] See discussion in Hacking (1999) on Kant's metaphor of construction and its impact in the philosophy of mathematics, ethics, and in the notion of social construction.

> science. It is too bad that it is first possible for us to glimpse the idea in a clearer light [*die Idee in hellerem Lichte zu erblicken*] and to outline a whole architectonically, in accordance with the ends of reason, *only after we have long collected relevant cognitions haphazardly like building materials* [*rhapsodistisch viele dahin sich beziehenden Erkenntnisse, als Bauzeug, gesammelt*] and worked through them technically with only a hint from an idea lying hidden within us. The systems seem to have been formed, like maggots, by a *generatio aequivoca* [spontaneous generation] from the *mere confluence of aggregated concepts, garbled at first but com-plete in time* [*aus dem bloßen Zusammenfluß von aufgesammelten Begriffen, anfangs verstümmelt*, *mit der Zeit vollständig*], although they all had their schema, as the original seed, in the mere self-development of reason. (A834-5/B862-3, my emphasis)

The way Kant describes the historical exercise of reason, then, is as a process on which the *practice* of reason informs our sense of the 'building materials'. The initially garbled aggregate of concepts that guide reason's inquiries reveal themselves to be regulative *ideas*.[39] The course of exercising reason informs our sense of the concepts—moral and theoretical—that we take to guide us, finally making them 'distinct to ourselves'. Moreover, Kant urges on us a self-imposition of *epistemic humility* regarding the process of normative construction (epistemic and practical): our evaluative standpoint, as Kant characterizes it here, is *not* fixed, but must remain open to change. Kant reinforces this point in the concluding section of KrV, "The History of Pure Reason," whose "title stands… only in order to indicate one remaining division of the system, which future workers must complete", since "those who have labored in this field" in the past have managed to construct "buildings, to be sure, but in ruins only" (A852/B880).[40]

[39] See Chapter 5 on Kant's eventual specification of this notion into 'moral' and 'rational' types.
[40] As Wood claims, "There is nothing ahistorical about Kantian ethics. It has a historically situated understanding of itself, and is addressed to the specific cultural needs of its own age" (1991: 336). Korsgaard (1996) paints Kant's position as the necessary endpoint of a historical progression in moral philosophy (from Aristotle, to Hobbes, to Hume, among others), to which Williams raises the question of what justifies situating Kantianism thusly; we should not "be so impressed by the onward march of the historical process that we feel we must accept its latest stage. (That line of argument is anyway open to the unanswerable objection… that if we have doubts about the 'latest stage', then it cannot be that stage, but rather our doubts, that constitute the latest stage)" (Williams 1996: 218). While many Kantians have positioned Kant as Korsgaard does, Kant himself leaves open the possibility that the questions Williams raises could *themselves* be included as integral building materials of the historical construction of reason.

There is one articulation of constructivism that reflects some acknowledgment of the need for epistemic openness Kant expresses in these passages. In "Realism and Constructivism in Twentieth Century Philosophy", Korsgaard (2003) suggests that normative concepts should be considered as "solutions" to "practical problems", such as the coordination or cooperation of a plurality of agents (115). She suggests that on a constructivist view of normativity, moral concepts are made "like the concepts of of artifacts. This doesn't make them arbitrary or relative, for there are kinds of artifacts—'chair' is an example—that all human beings in all human cultures have some version of, and that have to have certain features given the problems that they solve" (117). However, Korsgaard takes issue with Bernard Williams' view of values as socially contingent "artifacts" that agents "have not consciously built", arguing that, for the constructivist, "practical philosophy is all about" the conscious building of our social world (Williams 1985: 147; Korsgaard 2003: 115).

There is an interesting parallel here between Korsgaard's account of normative reasoning as problem-solving and Rahel Jaeggi's suggestion that we consider culturally specific 'forms of life' as "problem-solving strategies" or learning processes, which involve an inherent claim to the good "even where their manifestation is tied to a particular history or a particular place" (2013: 33). And, indeed, the epistemically objective but metaphysically subjective status constructivism assigns to normative facts aligns them with a certain view of the construction of *social facts* as the mind-dependent result of collective decision *by fiat* (Searle 1995).

Much turns, then, on whether we feel, in Korsgaard's optimistic spirit, that our social world can be consciously and collectively built, or in a skeptical, Williamsonian mood, that we are confined to our own contingent historical-geographical normative outlook: on whether it is *we* who construct our social world, or we who are instead socially *constructed*. If there can be a middle

path carved out here, that is where I would situate my own view: constructivism should be considered as *reflective* in method. The construction of the normativity of social life cannot always be assumed to be intentional and cooperative,[41] but aspects of it can be *made* conscious, particularly when we are intentionally distanced from our ordinary practices. The attempt to render the construction of value transparent—to open it up to *re*construction—is a moral task, a kind of virtue.

On the reflective constructivist picture, the 'contingencies' involved in the shaping of our normative standpoint referred to by Street (2008) are recast not merely as the result of forces of evolution, but as *social* forces: since the latter involve habituation into a second nature (McDowell 1996) as well as a 'first nature', they do not merely supervene on the former.

Furthermore, this approach to constructivism emphasizes that the operation of social forces are often not transparent to agents enmeshed within them. In this sense, Williams (1985) is right to claim that, given that any singular agent has done little to 'consciously build' the norms she inherits, the construction of value cannot be reduced to such a process alone. To be an agent at all is to be initiated into a given culture—to be, as McDowell insists, *gebildet*—where a culture, as Moody-Adams (1994) claims in her appraisal of its role in moral experience, is a "way of life of a given social group, that will be shaped by more or less intricate patterns of normative expectations about emotion, thought, and action" (295). Much individual behavior is therefore

[41] That Korsgaard takes the construction of normativity to normally function in a conscious, transparent fashion is brought out by her later claims in *Self-Constitution*: "When we create values, we invite others to share them, not just in the sense of helping us to promote them, but in the sense of interesting themselves in the valued object too. And, because we share a nature, the invitation is often accepted, and then people begin to explore the possibilities, and a tradition begins to take hold…. And we all live inside of such traditions of value, traditions of value that we hope and expect others around us to share" (2009: 209). As I gestured to at the outset, the critical theory tradition does not share the premise that traditions of value are ones others are transparently 'invited' into, or ones they can freely 'accept'.

only intelligible in light of the cultural practices into which one has already been initiated: the act of marrying, for example, is "conventional", as J.L. Austin (1962) famously put it, in that two people saying 'I do' while exchanging rings has no meaning outside of the preexisting social convention of marriage.

By extension, where a practice does not yet exist, possibilities for individual behavior may be foreclosed, or *practically unintelligible* (Chapter 2). This is due to the fact that practical possibilities show up to us as intelligible only within what Habermas (1981a, 198b) terms the 'lifeworld' or Gadamer (1960) a 'horizon of meaning': within a particular historical, social context which forms the backdrop of individual action. That is, what we do has to make sense to us, and this dimension of 'sense-making' is contingent on our enmeshment in a system of practices—discursive or otherwise—whose content we are habituated into, whose rules we have to learn, rather than being individually defined by any of us in isolation.[42] Ideology and genealogy critique work to remind us of the frequent operations of power in moral discourse and of the ever-present possibility that moral principles can be misapplied—but also to remind us of the role *we* play in perpetuating these phenomena. That is, they show that there is a way in which we each contribute to 'building' or 'constructing' the phenomena in question even when doing so has become habitual, seems necessary or self-evident, or when the true function of our constructions eludes our conscious grasp (Geuss 1981).[43]

[42] Not only Korsgaard, but also Searle seems to understate the obstacles involved in the 'conscious building' of social practices (or, as Jaeggi might call them, helping to bridge Korsgaard and Searle, 'concretizations of value'; 2013: 14). Since social facts are, on Searle's account, merely the result of collective agreement, the possibility of altering them is construed as transparent in nature. Indeed, this is one of the weakest point of Searle's analysis, as when he claims that the collapse of the Soviet regime was simply a matter of "the system of status-functions" no longer being "accepted" (1995: 92; see also Searle 2009). Needless to say, if things were that simple, social critique would not be necessary.

[43] As Margaret Urban Walker claims, "moralities, like the social lives from which they are not separable,

To acknowledge the relation between broader social forces and agency is not to excuse socially licensed individual actions that warrant moral reproach, but to situate it as a substantive normative problem regularly grappled with in everyday practical deliberation. Indeed, in his lectures on moral philosophy, Adorno situates the relation between society and the individual as perhaps the central question of moral philosophy in general:

> Fichte's famous assertion that 'morality is self-evident' cannot be upheld… Morality may very well appear to be self-evident in a world in which people feel themselves to be the exponents of a class in the ascendent… as was the case with the great bourgeois thinkers around the turn of the nineteenth century. The situation is very different when every important practice whose theory one tries to grasp has the unfortunate and even fatal tendency to compel us to think in a way that conflicts with our own real and immediate interests. (1963: 5)

The problem Adorno articulates of morality's lack of self-evidence presents itself whenever a social practice appears antithetical to our own needs—whenever our social world appears out of step with our own sense of the good. The problem of a live tension in contemporary life between our values and the social practices we often find ourselves (often in spite of ourselves) upholding is one driving much continental moral philosophy and critical theory. Many of us who identify as philosophers and scholars can easily affirm the good of what we do, yet to carry out a successful career as an academic philosopher in the current state of university life requires most of us to take many trips by plane over the course of any given year, flying to give transatlantic talks or attend discipline-wide conferences from coast to coast. The actions of many apparently 'radical' climate activists, such as committing oneself to only travel by bus or transatlantic ocean liner, would be incredibly difficult, if not outright impossible, for anyone who identifies with and participates in

are *collective more by accretion and concurrence than by concerted effort or design.* They are collective works sustained by their reproduction in many activities of many people who are only sometimes aware that they are sustaining something at the level of 'society' or "morality'" (2007: 237; my italics).

the social practice of academia—not to mention for others in many other industries.[44] Yet, in the wake of the current climate crisis—and bracketing all our other commitments including those values held by many scholars, such as the value of education, dissemination of scholarship, and intellectual inquiry—such a response is arguably the only rational, or moral, one.

The point of this example is not to claim that there is one correct moral path that those of us who participate in the practice of international conference-going have failed to take. Instead, it shows that the complexity of the social world makes it such that none of us can consider ourselves exempt from these kinds of moral dilemmas (albeit ones that, given our own needs and desires, we generally ignore rather than reflectively scrutinize), and with them, from complicity with the broader social forces instantiated in the particular practices we take up. We derive meaning from participation in a given practice, such as being a philosopher, whose content is not completely up to us, as international conference-going is not completely up to the more environmentally conscious in this group. At the same time, any given social practice is only in force to the extent to which it is perpetuated by individual agents, which always involves modification, reshaping, and revision.

What this interplay calls for, or so I will conclude, is a continuous reflective scrutiny of the principles that guide us, one that I will characterize as a kind of *virtue*: the virtue that Kant characterized as 'heautonomy', the normativity involved in revising or creating principles guiding judgment, and which motivated Kant in characterizing moral virtue as the "indefinite progress of one's maxims" (5:33). Defining this ongoing practice of reflective scrutiny and revision (that is, of critique) as a virtue enables us to excise it from the domain of moral blameworthiness or

[44] The choice of this example is an interesting one, given that, since it was initially written, the practice of conference-going has changed radically in the course of the global coronavirus pandemic.

responsibility, permissibility or impermissibility. As feminist moral philosophers have often acknowledged, this domain is one ill-suited for thinking through agents' apparent complicity in broader structural forces, since the exercise of agency here is often constrained by social gaps in the new moral knowledge that critique pioneers (hooks 1984, Calhoun 1989, Card 1996, Walker 2007).

1.5 Progress and 'partial transformations'

Before concluding, I will voice a final point regarding my commitment to progress as a normative notion. Immanent critiques invoking Hegel and Marx are often considered to be bound up with a commitment to a teleological conception of historical progress: the internal contradictions found within a given social order are to be worked out over the course of history, culminating in the realization of Spirit or the founding of a communist utopia. But, as the first generation of the Frankfurt School realized, such a conception of history can longer be presupposed, and—as becomes painfully ever more clear—appears to constitute a philosophical fiction. Moreover, as Amy Allen (2016) has importantly emphasized, building on a rich literature in post-colonial and decolonial theory, the general strategy of positioning progress as a normative notion—one still employed in both the Hegelian and Kantian strand of critical theory—risks falling into imperialist and Eurocentric tropes that ignore the enduring legacy of colonialism.

The notion of progress I am therefore committed to, to the extent I am committed to any such conception, is what Allen terms a 'forward-looking' conception of progress, one "oriented toward the future" as "a moral-political imperative… a goal that can be captured under the idea of the good or at least of the more just society", rather than a 'backward-looking' conception that treats progress as a developmental story, as a fact about the past (2016: 12). Any moral or

historical progress that can be ascertained, on my view, is to be understood as partial and fragmentary in nature. While, given our current evaluative standpoint, we might say that progress has been made on certain questions, such as construing the scope of the categorical imperative to admit groups previously excluded (Chapter 3), this does not preclude us from acknowledging that there have been other domains—say, the emergence of the prison-industrial complex, or the skyrocketing gap between the richest and poorest members of society in the United States—where moral revision might better be understood as moral *regression,* even where it may have once been understood as advancement. Indeed, any current interpretations of past moral progress must remain revisable in light of new understandings.

In this respect, while my reading of Kant might be seen by some as leading in a more Hegelian direction—at least insofar as it responds to some of Hegel's more pressing critiques of Kant's ethics—it does not see progress as any necessary or teleological feature of our historical reality. Conclusions about progress, to the extent that any are implied in what follows, are limited to particular, localized questions, and relativized to our historical moment—bracketing issues of the incommensurability of radically distinct historical (or, in certain cases perhaps, geographical[45]) periods. As Foucault (1984a) claimed about his own approach:

> I prefer the very specific transformations that have proved to be possible in the last twenty years in a certain number of areas that concern our ways of being and thinking, relations to authority, relations between the sexes, the way in which we perceive insanity or illness; I prefer even these partial transformations… to the programs for a new man that the worst political systems have repeated throughout the twentieth century. (316)

It seems to me that the epistemic humility Foucault expresses regarding what is realizable

[45] See Williams (1985).

through philosophy, and through critique in particular, has become increasingly attractive in the intervening years.

By retraining the object of critique to 'partial transformations' from broadscale institutions, such as the three institutions of recognition of the family, the market, and the state (Honneth 2015), critique, on my view, becomes an *ethical* task as much as a political one. The ethical and the political cannot be neatly pulled apart here, since, as I have suggested here, our individual reasoning about what to do is conditioned by historical and social context, while the normative force of social practices and institutions may very well need to be called into question. The instances of reasoning I draw attention to are variously individual or collective, but when undertaken collectively, I do not want to take for granted the stability and determinacy of the collectivity at issue: "The 'we' must not be prior to the question; it can only be the result—and the necessarily temporary result—of the question" (Foucault 1984c: 1413). If we find contemporary practices to fall short of our needs and aspirations, if they invite skepticism as to whether individual flourishing is encouraged or even made possible by them, this puts a responsibility on *us*—a collectivity not to be predetermined, but defined *post facto* on the basis of those who find one or more of those practices to be intolerable—to act, to the extent to which it is within our power to do so.

Chapter 2: On Moral Intelligibility

2.1 Obdurate particulars

What does it mean to speak of particulars that resist subsumption under available universals? To make a start at this question, I first want to put Kant aside. I will consider three examples of particulars that are variously *unintelligible* within their operative moral schemes. Each example involves a particular that, in its immediate context, could not fit into an available universal, yet could not readily result in the revision of its scheme, either.

1. Gayatri Spivak gives an account of the suicide of a young Bengali woman, in a time and place when women's suicide was considered noble under certain conditions (illegitimate pregnancy, being widowed):

> A young woman of sixteen or seventeen, Bhuvaneswari Bhaduri, hanged herself in her father's modest apartment in North Calcutta in 1926. The suicide was a puzzle since, as Bhuvaneswari was menstruating at the time, it was clearly not a case of illicit pregnancy. Nearly a decade later, it was discovered that she was a member of one of the many groups involved in the armed struggle for Indian independence. She had finally been entrusted with a political assassination. Unable to confront the task and yet aware of the practical need for trust, she killed herself.
>
> Bhuvaneswari had known that her death would be diagnosed as the outcome of illegitimate passion. She had therefore waited for the onset of menstruation. While waiting, Bhuvanesari, the *brahmacarini* who was no doubt looking forward to good wifehood, perhaps rewrote the social text of *sati*-suicide in an interventionist way.… She generalized the sanctioned motive for female suicide by taking immense trouble to displace (not merely deny), in the physiological inscription of her body, its imprisonment within legitimate passion by a single male. In the immediate context, her act became absurd, a case of delirium rather than sanity. (1988: 307)

Had Bhuvaneswari's act been committed for the 'right' reasons, it would have been socially and morally legible, ready for subsumption under governing moral universals such as 'purity', 'honor',

or ‘*sati*’, the archetypical ‘good wife’. But, Spivak emphasizes, Bhuvaneswari took pains to *resist* such ready subsumption. By waiting for the onset of menstruation to take her life, she transgressed the two norms for *sati*-suicide at once: first, she revealed that she had not dishonored herself by getting pregnant; second, she violated the injunction to be ‘pure’—to not be menstruating—before claiming the “dubious privilege” of feminine self-sacrifice (1988: 308). This act of resistance, however, could not be understood as such; it could only be made comprehensible as a ‘case of delirium’. Spivak has since remarked that Bhuvaneswari’s example shows “that whereas the British Indian reform of *sati* is much celebrated, when a young, single girl attempted to write resistance in her very body, she could not be read” (2006: 3). The singularity of Bhuvaneswari’s act remained hermeneutically impermeable, *morally unintelligible*: the particular opposed to the universal.

2. Judith Butler recounts her upbringing:

> I grew up understanding something of the violence of gender norms: an uncle incarcerated for his anatomically anomalous body, deprived of family and friends, living out his days in an ‘institute’ in the Kansas prairies; gay cousins forced to leave their homes because of their sexuality, real and imagined; my own tempestuous coming out at the age of 16; and a subsequent adult landscape of lost jobs, lovers, and homes.… It was difficult to bring this violence into view precisely because gender was so taken for granted at the same time that it was violently policed. It was assumed either to be a natural manifestation of sex or a cultural constant that no human agency could hope to revise. (1999: xxi)

Butler lists a litany of wrongs that eluded acknowledgment as such. Being forced out of one’s home, deprived of family and friends, incarcerated because of one’s body: without legible sexual preferences or gender (or gendered body), such suffering was tacitly rendered admissible. The particularity of these modes of experiences resist subsumption under the operative universals: of gender, of sexuality, and ultimately, of *humanity*, given its apparent narrowing of scope to only

certain gendered and sexed bodies, those identified as legible instances of the human. Yet, once again, the injustice inflicted remained unintelligible—unintelligible, that is, in moral terms.

3. Betty Friedan offers the case of the 'problem that has no name':

> Just what was this problem that has no name? What were the words women used when they tried to express it? Sometimes a woman would say "I feel empty somehow... incomplete." Or she would say, "I feel as if I don't exist." Sometimes she blotted out the feeling with a tranquilizer. Sometimes she thought the problem was with her husband, or her children, or that what she really needed was to redecorate her house, or move to a better neighborhood, or have an affair, or another baby. (1963: 16)

These women—notoriously upper middle class, white, and American—could each discern a problem in the structure of their everyday lives, but were unable to *name* the problem, to articulate the nature or source of the wrong. If we think of practical deliberation, on the conception I will argue for in the next chapter, as incorporating *both* reflection on particular features of one's circumstances *and* a connection to one's sense of the good (or right, or just), these women come up against a seemingly unbridgeable gap between the two. A particular feature of their experience feels wrong, but cannot be discursively assimilated as such. As a result, they name only false culprits—their husbands, their houses, their neighborhoods—and remain entrenched within the same constrained patterns of action.

There are different ways to account for the events described in each case, but one way to characterize what they have in common is that each involves a lack of fit between moral universals and a particular empirical feature: another's 'delirious' action, a systematic mode of treatment, a feeling of unease about one's pattern of life. The first two cases might be explained away as irrelevant to practical reason because they involve the *subsumption* of others' actions under

categories of moral evaluation (presupposing the stance of spectators; see O'Neill 2018).[46] However, the third demonstrates that the same phenomenon also manifests in the first-person phenomenology of practical deliberation: at the level of *application* of moral principle to experience, rather than subsumption alone.

The fact that Bhuvaneswari was menstruating constituted a morally salient feature of her action-context, but this merely rendered her behavior all the more strange. It was intelligible only as the consequence of insanity; it did not, could not, result in the revision of the moral universals whose applications she was resisting. The 'problem that had no name' was already a salient feature of the appraisal of the life circumstances and self-understandings of the women Friedan interviewed but, once again, was somehow blocked from entering into deliberation about how to respond to it. Similarly, Butler describes broad swathes of agents' lives being determined against their will, but in a way 'difficult to bring into view' because the source of the wrong, while salient as a 'constant' or 'natural' fact, was not taken to be open to human revision.

2.2 The limitations of two frameworks

Hermeneutical injustice

It has become customary to describe cases such as these as instances of epistemic injustice, particularly *hermeneutical injustice.* After all, in each of these cases, part of the problem was that epistemic resources were lacking to make sense of a significant patch of agents' experience, rendering them 'hermeneutically marginalized', subject to "subordination and exclusion" from participation in an (interpretive) practice they would have valued (Fricker 2007:

[46] As I will argue in Chapter 3, however, this reflection on context is a prerequisite for any kind of moral response.

153). This can happen, according to the epistemic injustice framework, because "relations of unequal power can skew shared hermeneutical resources so that the powerful tend to have appropriate understandings of their experiences ready to draw on as they make sense of their social experiences, whereas the powerless are more likely to find themselves having some social experiences through a glass darkly, with at best ill-fitting meanings to draw on" (148).

While each of the aforementioned examples involve aspects of experience open for interpretation only in terms of 'ill-fitting meanings', the epistemic injustice lens entails certain limitations. For instance, it presupposes prior knowledge of who the hermeneutically marginalized *are*. As Fricker puts the point:

> Different groups can be hermeneutically disadvantaged for all sorts of reasons, as the changing social world frequently generates new sorts of experience of which our understanding may dawn only gradually; but only some of these cognitive disadvantages will strike one as unjust. For something to be an injustice, it must be harmful but also wrongful, whether because discriminatory or because otherwise unfair. (151)

Thus, in only certain instances will hermeneutical disadvantage count as a form of injustice. To constitute unjust hermeneutical marginalization, the disadvantage in question must be 'discriminatory' or 'unfair'. But adjudicating this question is much easier to do after the fact, in analyzing historical cases where much of the work of recodifying first-person gaps in understanding as instances of injustice has already been done. The women Friedan described did not understand their inability to articulate the nature of what they were feeling to be the product of discrimination; nor did the women's colleges whose mission was to prepare them for their future lives as wives and mothers; nor did their husbands, who understood their task as materially providing them for lives of relative comfort at home; nor did society at large, which understood working outside the home to actually constitute a *dis*advantage facing only those women forced

to do so, out of material necessity. Similarly, in Spivak's telling, *sati*-suicide, far from a source of disadvantage, was coded as a *privilege* for women, the only way to release oneself from one's female body in the cycle of rebirth (1962: 303). In defining hermeneutical marginalization in terms of 'what strikes one as unjust', the epistemic injustice framework already bypasses much of the moral and political innovation involved in redefining as a wrong what previously seemed to constitute fair, even generous, treatment.[47]

Thus, part of what is missing here is a more basic understanding of the role *we* play in creating and maintaining social reality. The question of marginalization is usually not easy to answer, if it can be answered at all, at the moment in which the resources necessary to interpret a given issue are found to be lacking. For example, the act of designating 'homosexuality' a disorder in the Diagnostic and Statistical Manual of Mental Disorders (DSM) was widely considered a way of *helping* individuals understood at the time to be psychiatrically disturbed—not *marginalizing* them. While some of the individuals afflicted by this designation understood it to be a form of injustice, many others did not. Rather than turning to a collective fight against collective injustice, many individuals labeled 'homosexual' instead turned inward, understanding their duty to be that of ridding themselves of 'disturbance' or 'sin'. Once these agents could redefine their experience not as personal deficiency, but as marginalization—once, in other words, the available universal had been redefined to accommodate the particularity of their experience—overcoming their unequal hermeneutical participation was already possible.

An implication of Fricker's model of epistemic resources is that virtue is construed as

[47] Congdon (2017) helpfully raises the question of the moral framework the theory of epistemic injustice presupposes; as he notes, Fricker herself combines virtue-theoretic, Kantian, and social contract elements in her initial account, and these divergent—and, arguably, incompatible—approaches have also been taken up in the ensuing literature on epistemic injustice.

extending primarily to the *hearer* of the testimony of a socially marginalized agent. Even hermeneutical injustice, which does not principally involve an interlocutor but instead characterizes the victim's own individual attempt to make sense of her own experiences, calls for a virtue derivative of the virtue of testimonial justice: "The form the virtue of hermeneutical justice must take, then, is an alertness or sensitivity to the possibility that the difficulty one's interlocutor is having as she tries to render something communicatively intelligible is due… to some sort of gap in collective hermeneutical resources", such that "the virtue in question is like the virtue of testimonial justice" (Fricker 2007: 169). Thus, the ethics of epistemic justice is construed as primarily the responsibility of the socially advantaged, the powerful, with respect to the experience in question. It appears to leave little room for the victim's own response to her own experience: how to determine when she herself is marginalized with respect to her experience; how to notice when she lacks adequate epistemic resources; and what she can do about it, besides giving testimony to a (hopefully epistemically virtuous) interlocutor.[48] On this model, Bhuvaneswari is not, herself, a locus for virtue; there is no differentiating her act of resistance to the norm of wifely suicide from her peers' compliance to it. Instead, it is the class of those who witness this act—patriarchal men, British colonialists, and the relatively more privileged interlocutors to whom Spivak relates the story sixty years later—who are called upon to be virtuous. But this seems backward.

While the notion of epistemic injustice gives us a helpful way of understanding how morally relevant particulars might resist subsumption, it presupposes a fixed moral framework, neatly circumscribing distinct social groups, the 'powerful' and the 'powerless', adjudicating who

[48] While offering a systematic response to these questions is beyond my scope, it is an important element of a theory not to *foreclose* such questions.

counts as hermeneutically disadvantaged in a way that is 'unfair' or 'discriminatory'. The discovery of an epistemic injustice has the effect of ameliorating or rebalancing these disparities (helping to expunge instances of unfairness, discrimination, or marginalization); it does not, that is, reconfigure the governing terms of moral adjudication—what *counts as* unfairness, discrimination, marginalization.

Imaginative resistance

Another way to make sense of the 'obdurate particulars' I referenced at the outset might be in terms of *imaginative resistance*:[49] the puzzling phenomenon whereby we can freely suspend our belief when it comes to entertaining fictional worlds of wizards and monsters, castles cast under spells and self-teleporting extraterrestrial life—but find this willingness to reach a limit when it comes to *morally deviant* worlds. We resist being directed to imagine that murdering an infant or inciting a genocide is the correct thing to do, even if we can perfectly well imagine those actions taking place; we distance ourselves, then, only from the evaluative components of the fictional world rather than the descriptive. Gendler (2000) posits that this happens not because we are *unable* to follow authors in imagining such things, but because we are *unwilling* to do so; these are cases that occur "whenever we feel that we are being asked to add to our repertoire of schemata a way of looking at the world which we prefer not to have available" (79). Gendler's explanation posits that because there are actual instances of moral disagreement in the real world, while moral claims are instead often taken to hold categorically, it remains ambiguous whether deviant moral evaluations in fictional worlds demand importation into the real world, or

[49] I am indebted to Michelle Kosch for suggesting the relevance of the puzzle of imaginative resistance.

whether they are posited merely within the bounds of the imagined scenario. Our resistance stems from this inherent ambiguity.

The puzzle of imaginative resistance has generated a vast volume of literature questioning every aspect of Gendler's initial analysis. Without wading into the nuances of these discussions, I want to put pressure on the apparent intransigence of imaginative resistance. On Gendler's account, imaginative resistance is fixed—which is to say, fixed from within a subject's individual point of view: she gives the example of "a parent concerned with gender equality" who resists "calling the strong chair the 'papa chair' and the weak chair the 'mama chair'", or "the advocate of abortion rights speaks of himself as 'prochoice' and not 'antilife'" (2000: 80).

Against this picture, I want to question whether the bounds of imaginative resistance are alterable, whether by way of fiction, through social critique, or other means. There seem to be instances in which the aspects of fictional worlds that initially elicit imaginative resistance may no longer do so over time—where, in fact, overcoming the disposition to resist is part of the *point* of the way in which the fictional world has been set up. A work such as *Brokeback Mountain* (2005), to name a widely known example, is filmed with the cognizance that some viewers will meet some of the scenes depicted with disgust, resisting inhabiting the states of mind of the characters; but this is not an aesthetic *failure* of the kind Gendler emphasizes when it comes to the imaginative resistance elicited by the racism of Rudyard Kipling's poem "The White Man's Burden" (Gendler 2000: 61). Where Kipling seems to presuppose that his readers will *agree* with his moral point of view, thereby not anticipating any imaginative resistance, *Brokeback Mountain* *invites* such resistance in an attempt to do away with it, to *transform* one's moral imagination. The narrative succeeds to the extent to which some viewers initially resist incorporating its way of looking at the world into their conceptual repertoire, but come to do so anyway, due to the

vulnerability of its characters and the relatability of their situation.

Brokeback Mountain is not special in this regard; there are many fictional worlds constructed with a similar aim in view. (Nor is it the only possible strategy when it comes to narrativizing the experiences of the socially marginalized: we can think of feminist separatist or Afrofuturist literatures, whose critical function arguably does not rest in aiming to overcome the resistance of readers who reject their starting premises.) And, of course, it is unlikely that any one fictional world can alter imaginative resistance by itself. One can also think of more sinister examples, perhaps along the lines of *The Birth of a Nation*, where the aim of overcoming spectators' resistance seems objectionable rather than laudable. The point is just that we need not model the moral imagination as fixed; that the bounds of our resistance—the dividing line between what elicits it and what does not—can be expanded or restricted; and that fiction, among other discursive endeavors, can task itself with reshaping this imaginative capacity.

2.3 Intelligibility: moral and practical

To anticipate what I will develop in more detail in Chapter 3, reflective judgment can enter into practical reasoning at several junctures:

1) reflection on the circumstances in which we intervene in action;
2) reflection on the degree of 'fit' between the discursive resources available to us in stating intentions and our recognition of particular features of our circumstances;
3) reflection on the degree of 'fit' between our action characterizations (what I will term 'i-maxims') and moral principles ('p-maxims').

This list is not exhaustive; it leaves aside other possible junctures outside of my scope in this

project, notably the role of feeling and the moral emotions in practical reasoning.[50] But what unites each of these features is that each instance involves reflection on the fittingness of particulars to universals in the course of practical deliberation.

The employment of reflective judgment in these three aspects of practical reason demonstrates that we can *make* our own course of judgment—as I will discuss in Chapter 4—and, by extension, that our judgment can be *unmade,* or *remade.* The flexibility of practical reasoning, if attributed to the use of reflective judgement, offers a useful model for demonstrating that the bounds of imaginative resistance can be resituated and our moral imagination altered.

Korsgaard (2009) defines action as "an essentially intelligible object that *embodies* a reason, the way a sentence is an essentially intelligible object that embodies a thought" (14). Intentional action must be 'essentially intelligible' in that it is subject to how we make sense of it, how we describe it to ourselves; it is subject, that is, to the *problem of relevant descriptions* (Chapter 3): "The reason for an action is… the action itself, described in a way that makes it intelligible" (Korsgaard 2008: 227). What makes intentional action distinct from a set of physical movements is that it is reasoned, is made intelligible.[51]

---

[50] Kant emphasizes the role of feeling in aesthetic judgment from which further 'universals' can be derived, particularly the universally communicable pleasure we feel in judging an artwork to be beautiful; see discussion in §3.4. This strikes me as a helpful model for understanding the role of moral emotion in mediating the availability of new moral concepts, but I cannot develop the point here.
[51] Hacking remarks on the implications of this intelligibility constraint for discursive changes in history: "What is curious about human action is that by and large what I am deliberately doing depends on the possibilities of description…. Hence if new modes of description come into being, new possibilities for action come into being in consequence" (1985: 108). For example, it is only possible, Hacking argues, to be a *garçon de café* "at a certain time, in a certain place, in a certain social setting", for example, in a café in 1940s Paris (108). Brandom (2007) makes a similar point: "Essentially self-conscious creatures are (partially) self-*constituting* creatures… an alteration in self-*conception* carries with it an alteration in the self of which it is a conception. Essentially self-conscious creatures accordingly enjoy the possibility of a distinctive kind of self-*transformation*: *making* themselves be different by *taking* themselves to be different" (128). The rapidly proliferating literature on conceptual engineering (see, e.g., Chalmers 2018; Cappelen 2018; Burgess and Plunkett 2013a, 2013b) has recently taken up these points from a different

The examples I emphasized at the outset of this chapter, however, show that not all morally relevant features are intelligible as such; they show that it can take effort, an explicit intervention, to render them *morally intelligible*. The moral undertaking in such cases consists not in steeling oneself against the inclination to do what is already known to be wrong, but in creating *new* moral knowledge: attempting to demonstrate to others—or work out for oneself—the limitations of the available moral repertoire, making intelligible what cannot currently be comprehended as morally relevant (or what is taken up as relevant to an ultimately noxious moral universal, such as 'purity').

Yet, to render explicit a claim first suggested in Chapter 1 (to be elaborated in the following section), traditional treatments of morality have been too restricted to recognize the *moral* dimension of the efforts described in these examples. This failure has also been the source of a persistent criticism of anti-foundationalist critical philosophers (of which Spivak and Butler, not coincidentally, are contemporary exemplars) as amoral, anti-Enlightenment, or nihilistic.[52] I propose, however, that we resituate these endeavors as calling on us to utilize our reflective judgment in practical reasoning, and, by implication, to expand our conception of practical reason to encompass the reflective dimension alongside the moral. The employment of 'reflective judgment', however, is understood not just in terms of successfully recognizing salient features or

angle than I do here.

[52] Some of the confusion over these issues is also to be attributed to the theorists involved; the term 'non-normative', for example, is sometimes used to mean 'having no ethical commitments' (usually by detractors), and at other times (as self-ascriptions by 'non-normative' theorists themselves) to mean 'anti-foundationalist', skeptical of the prospects for devising exhaustive moral frameworks. Of course, it is not logically possible to lodge a critique without any normative preference for what would be better than the object of one's criticism, so the first reading is uncharitable. On the other hand, the anti-foundationalists have sometimes been unduly wary of saying anything positive by way of their commitments, given their general skepticism to talk of normativity, morality, or ideal theory in general.

eschewing universals altogether, but as a persistent critical attention to the respective connections between particular features of one's action-context, the discursive resources available to one in articulating what one is doing, and the 'semantic depth' of the moral principle one takes oneself to be actualizing (or judging in light of): between particulars, universals, and intentional actions as the 'schemata' mediating between the two.

There has been considerable debate in the fields of epistemology, philosophy of language, and history and philosophy of science over the extent to which we can talk of differing conceptual schemes.[53] Ian Hacking (1982, 2012) proposes the notion of historical 'styles of reasoning', which pick out what counts as a candidate for truth or falsity in a given scheme of thought. That is, a style of reasoning is defined not in terms of what counts as *true* or what counts as *false*, but what counts as *true-or-false.* A statement which constitutes a candidate for truth-or-falsity in one scheme or style might not be meaningful or comprehensible (thus, not the sort of statement which could be judged to be *either* true or false) in another.

It is interesting that the theorists who have devoted the most attention to modes of epistemic change come so prominently from the fields of art history and the history of science, representing the two modes of 'merely' reflective judgment, aesthetic and teleological, which, as I will argue in the following chapter, Kant took to exemplify the activity of the power of reflective judgment in isolation. Within this literature are seminal conceptions which emphasize the *normative* force of a given historical point of view. In particular, historians of art and science have claimed that empirical observation is guided by a given historical *normative orientation* or *principle,* one prone to change over time. Lorraine Daston and Peter Galison, for example, posit

[53] For classic—and still influential—reference points, see, e.g., Quine (1960), Kuhn (1962), Davidson (1973); see Rescher (2019) for a recent account of 'conceivability' in epistemology and logic.

distinct epistemic virtues that have historically guided scientists in their empirical inquiries: truth-to-nature, mechanical objectivity, and trained judgment, where "each successive stage presupposes and builds upon, as well as reacts to, the earlier ones" (2007: 19); Erwin Panofsky explains the 'discovery' of perspective in Renaissance art by appealing to the arrival of a new ideal, that of *unifying* and *systematizing* space, rendering it 'measurable' where previously it had been conceived as discontinuous (1927: 47-49);[54] Thomas Kuhn famously speaks of incommensurable paradigms guiding scientists' research, which normatively structure their investigations without always being formulable into distinct bodies of rules (1962: 45-7). While each of these theorists posit a distinct model of historical change in art and science, they each construe epistemic transformation in terms of a shift in the general normative orientation structuring historical observation, one sometimes thought to be brought about by empirical particulars, anomalies or counter-examples, impinging on that orientation, either from the outside or from within.[55]

In this dissertation, I have posited a similar, albeit less radical, picture when it comes to moral reasoning. Forming an intention, an i-maxim, requires reflection on particular features of context, but should also be congruous with a given p-maxim. When confronted with anomalous morally salient particulars—that is, ones that seemingly can't be incorporated into our action-responses, or resist assimilation into available moral universals—we can be led into moments of revision, moments where we attempt to render the particular newly intelligible.

When I speak of *moral* or *practical* intelligibility, I intend, more specifically, what counts as a candidate for *moral* or *practical* evaluation. Practical and moral intelligibility involve "ethical

[54] See also discussion of this theme in the history of art in Danto (1997): 36.
[55] I am indebted to Helen Zhao for helpful discussion on these issues in the history and philosophy of science.

phenomenology", or how possibilities for action and morality *seem* to us (Fricker 2000). Practical intelligibility is broader than moral intelligibility in that many actions may not explicitly enter the realm of moral evaluation. If I get myself a drink of water, in many circumstances this action will be undertaken without explicit moral evaluation. But it may come within the domain of moral intelligibility if, say, the drink I am taking is the last source of water available to a group of people dying from dehydration on a deserted island. What is practically intelligible should, in general, also be morally intelligible: it should be fit to come under moral evaluation. As I will argue in the following chapter, on the classic model of intention (one prefigured by Kant's considered account of the systematic orderability of higher- and lower-order maxims), in undertaking an action, I generally should be able to relate it to some conception of what it makes *sense* for me to do, what I have *reason* to do, of what is *good* to do—some conception, that is, of the *right* thing to do (as encapsulated in a given moral universal). Instances where this connection can no longer be affirmed are ones that can create crises or occasion instances of critique, since what I am doing lacks *moral intelligibility*.[56]

An odd split between practices and actions

The dimension of practical intelligibility I have just emphasized has been more readily thematized in philosophical reflection on the nature of social practices. Social philosophers often

[56] In particular, being socially marginalized can in itself take the form of 'bad moral luck'—for instance, being socially defined as perverse, criminal, or irrational—such that "our moral lives can fail because they are characterized by abnormally frequent *unintelligibility* to others or abnormally frequent inability to defend one's actions in terms that others find meaningful" (Calhoun 2016: 30). Calhoun (2016: 37), as well as Spivak (2006: 3), also speaks of moral 'legibility', of being 'read by others'; moral 'intelligibility', which has the same root, therefore encompasses both the degree to which moral particulars are taken up by others (are 'legible') and the degree to which we ourselves can make sense of them (are 'intelligible').

take group action to rest on a background condition of collective intelligibility, on which such action, which must necessarily be taken up individually as well as collectively, is in part carried out unthinkingly, implicitly, or not wholly intentionally.[57] Moral philosophers, on the other hand, often treat intentional action as subject solely to individual endorsement, and as thereby constitutive of individual rationality as such; the way in which such actions are collectively, historically, or socially shaped, as it is characterized in the first set of conversations, is generally not thematized as a practical concern.[58]

The two conversations thus tend to rest on different presuppositions about the 'practical' domain. This split is encapsulated in Rawls' dichotomy between practices and actions. Rawls claims that "the rules of practices are logically prior to particular cases", such that a particular

---

[57] For attention among philosophers, see Taylor (1985), Descombes (2014), Gooding-Williams (2017), Haslanger (2018). There has also been explicit attention in sociological and anthropological theory (as part of the so-called 'practice turn'; see Ortner 2006) to practical intelligibility in particular, although the term is used somewhat differently than how I employed it above. See Schatzki (2001), who refers to 'practical intelligibility' as "the state of affairs that action makes sense to someone to do", since "people almost always, I contend, do what makes sense to them to do" (55). See also Turner (2001), who refers to the 'mutual intelligibility' of actions or practices. While my use of the term overlaps with these theorists', I am more interested in its relation to what I call *moral intelligibility,* which has tended to escape notice in these fields.

[58] One helpful exception here has been the Wittgensteinian tradition in moral philosophy, which has stressed that moral concepts only show up as intelligible in light of their social context—as Wittgenstein put it, their "inherited background", "bedrock", or "form of life" (Wittgenstein 1953: §94, §317, §341). For elaborations of this way of understanding moral intelligibility, see Anscombe (1958: 6), Winch (1987), Diamond (1988), and Conant (1995). On these conceptions, once the relevant background is lost, a given moral term can cease to make sense, even if it remains in use. These approaches can manifest as a rather determinist way of treating moral intelligibility, where the world would need to change altogether before a recalcitrant particular could result in a new or revised universal. This comes out sharply in Diamond's (1988: 269) rejection of Alison Jaggar's (1974) argument for doing away with gender concepts, on the grounds that to lose these terms would require us to lose the context by which we come to learn them, and this is not something we can easily do. Instead, I find it more helpful to take in view, not just how the world shapes our concepts, but how our conceptual repertoire shapes the world in turn. As Taylor puts it, "There is no simple one-way dependence here [between vocabulary and social practice]. We can speak of mutual dependence if we like, but really what this points up is the artificiality of the distinction between social reality and the language of description of that social reality. The language is constitutive of the reality, is essential to its being the kind of reality it is" (1971: 24).

action falling under the rules of a given practice "would not be described as that sort of action unless there was the practice" (1955: 25). In other words, the difference between actions and practices is that practices rest on a precondition of *collective* intelligibility, one that surpasses what an individual reasoner can secure. On Rawls' picture, there is no contesting individual actions constituting a practice without questioning the practice itself. Moral questions therefore only arise on the level of the practice itself, not on the level of individual action within a given practice. The singular instantiations of practices in individual actions are by definition outside the domain of moral justification; moral inquiry extends only to the dimension of individual actions, or a given practice as a whole. But individuals have little input in the questioning of a whole practice; to the extent to which this can be done, it can only occur on the level of public discourse about a given practice. In order for public discourse of this kind to take place, a given domain of action must already be collectively recognized as a discrete and explicit social practice—otherwise, there would be no common object of discussion.

The practice/action dichotomy hides the fact that many, if not all, of our individual actions fall under the domain of some practice or other. Being an academic philosopher guides many of my actions, yet what it means to be a philosopher is not wholly up to me—it is a role that has predated my existence and one which will (hopefully) outlive me. Arguably, being a philosopher falls under Rawls' definition of a practice, "a sort of technical term meaning any form of activity specified by a system of rules which defines offices, roles, moves, penalties, defenses, and so on, and which gives the activity its structure" (1955: 3).[59]

[59] Calhoun (2016) raises similar worries regarding the way in which the categorical imperative procedure seems unable to rule out already pervasive, yet unjust, social practices: "The argument [of false promising] works independently of any moral assessment of the practice of promising itself. At no point is the question raised whether rational beings must, or could, will the practice of promising itself—that is,

Korsgaard's introduction of 'practical identity' into her account of normativity invokes a similarly Rawlsian split, separating ordinary moral choice from the domain of social practice. Practical identities refer to the roles, typically the *social* roles, to which an agent subscribes, such as being a student or a citizen. As such, practical identity ranges over the domain of *normative contingency*: how moral judgment can be arrived at in contingent circumstances, among differently situated agents, and given individual variation in self-constitution, as distinct from the normative objectivity that derives from the more fundamental value of humanity (Korsgaard 1996: 241-2). Adhering to a practical identity, on Korsgaard's view, requires a certain degree of suspension of reflective questioning. For example,

> A good citizen cannot pay her taxes because she thinks the government needs the money. She can *vote* for taxes for that reason. But once the vote is over, she must pay her taxes because it is the law. And that is again because citizenship is a form of practical identity…. To be a citizen is to make a certain set of decisions in company with the other citizens—to participate in a general will…. And for exactly that reason, in so far as you are a citizen, you aren't free to act on your own private reasons any more. (1996: 106)

Like Rawls on practices, Korsgaard holds that an individual's relationship to her practical identity is either one of wholesale endorsement or rejection. Since it is practical identity that *gives rise to* normativity, or at least a significant source of it ("your reasons express your identity, your nature; your obligations spring from what that identity forbids"), it is not possible in any straightforward sense to question actions constitutive of a given practical identity—such as

whether the practice is either morally required or morally permissible…. [If an agent] proposes keeping her promise[,] this maxim will sail right through the CI procedure. It will do so not because of any moral feature of the practice of promising. Rather, it passes simply because her proposal does not involve making an exception for herself to a practice that she assumes others will continue upholding. This is worrisome. The CI procedure appears tailor-made to catch a particular form of moral failing in its net, namely the temptation to make an exception of ourselves to generally followed social practices that are assumed to be morally acceptable ones. Because it is not designed to assess the practice itself, all maxims of compliance with generally followed practices pass it" (160).

paying taxes for the identity 'citizen' (101). It is only in the case that "practical conceptions of your identity… are fundamentally inconsistent with the value of humanity" that they "must be given up" (130). Thus, constitutive aspects of a given practice, or practical identity, are decided 'behind our backs', and can only be questioned either within the procedures established within the practice itself (such as voting), or by throwing out the whole practice, or identity, altogether.

Rawls emphasizes, more than Korsgaard does, the extent to which the constitutive rules of a given practice are 'logically prior' to their performance by particular agents in a particular circumstance. Korsgaard, on the other hand, emphasizes how pervasive these roles are in establishing what we take to be normative *for us,* for ordinary action as such. To punish or promise, Rawls suggests, we must first learn the actions collectively taken to constitute promising or punishing; we cannot ourselves *decide* which actions will be constitutive of the practice. But, as Korsgaard's conception of 'practical identity' already implies, it is hard to cordon off a domain of discrete social practices from the domain of ordinary moral action. Being a student or a citizen, much like promising or punishing, requires habituation to how these roles are constituted, and what they require in order to be fulfilled—roles that can greatly differ depending on cultural or historical circumstance, and which therefore cannot be entirely subject to individual choice or endorsement.

Perhaps Korsgaard might say that agents nevertheless can choose between adopting certain practical identities rather than others. Becoming a student, mother, or a philosopher can all, to a certain degree, be up to me. After all, 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", a sort of description that is inherently particularized and self-defined (1996: 101).

Yet there is a danger inherent in collapsing the notion of who *I* am with the social roles or 'identities' I inhabit. As Butler (1989) emphasizes, the normativity issuing from identity can cut both ways. For those within the bounds of those identities that are already collectively legible and intelligible, the categories at their disposal prove self-affirming, giving meaning and value to themselves and their life projects. But for those who fall outside these bounds, this *very same normativity* can prove exclusionary, 'do violence'.[60] Korsgaard writes that "to no longer be able to think of yourself under the description under which you value yourself and find your life to be worth living and your actions to be worth undertaking… is to be for all practical purposes dead or worse than dead" (102). But what to do when the conceptual ingredients for appropriate self-description are not available, when the descriptions operative in the dominant conceptual repertoire exclude the very possibility of who one feels oneself to be?

The prospect of critique therefore rests on a different entailment from Rawls' observation that reflection on the morality of agents' individual actions constituting a practice always entails reflection on the nature of the general practice they thereby undertake. Individual moral actions are 'practices' insofar as they depend in part on a background condition of intelligibility that operates to some extent 'behind the backs' of individual agents. But social practices are also moral 'actions' in that they can also constitute possible objects of moral commendation or reprobation. They can therefore be altered in part through everyday reflection on the justification of those practices; indeed, they can only exist to the extent that they are instantiated and perpetuated by individuals.

[60] Geuss notes in his reply to Korsgaard that "it is striking… that Kant himself doesn't talk about 'identity' in ethical contexts, and notoriously Kant thinks that 'rational psychology', the metaphysical discipline purportedly studying the underlying bearer of personal identity, is a pseudoscience" (1996: 191).

Moral intelligibility and critique

To comprehend one's action as part of a collective practice does not suspend moral scrutiny, but extends it to the social reality that renders that action intelligible as such. Likewise, to construe a social practice in terms of the individual actions that make it up resituates it as the potential object of moral endorsement or disapprobation. One way to view the function of critique is that critique *just is* the project of construing practices as individual actions, while situating individual actions as collective practices. The practical force of critique rests on the presupposition that practices are always reflectively endorsable, and that interrogating the contingency of a given set of practices cannot be situated outside the domain of everyday processes of giving and asking for moral reasons.

Celikates (2006) argues that accounts of critique have foundered in the debate between those, as in the Marxist tradition, who view critique as a matter of diagnosing ideology, or 'false consciousness' that must be 'unmasked', and those, such as Pierre Bourdieu, who reject the notion that critics can occupy an objective standpoint outside the practices they evaluate, situating themselves within the same hermeneutical perspective of the agents they observe.

On Celikates' diagnosis, theorists in the vein of Bourdieu have tended to define the practices they analyze as the 'social games' we participate in intuitively, such as "being a student in a classroom, a whitecollar worker in a cubicle, a visitor to the museum", where "what is characteristic of our engagement in those practices is a constitutive misrecognition of their real nature, an incapacity to critically distance ourselves from what we are doing and to question it" (24). But, Celikates insists, this is not in fact *true* of our participation in practices; we *do* in fact regularly step back from our practices and reflect on their justification. While we often participate in them unthinkingly or unreflectively, agents can in fact "relate to what they and

others are doing and thinking either critically or affirmatively", and commonly do so, particularly "in everyday situations of crisis and conflict" (30).

We always preserve the possibility, then, to stand back and reflect on the nature of what it is we are doing. And it is insofar as critical theory serves the function of drawing our attention to those practices that might particularly merit reflection that it can be integrated into these everyday practices of critique and justification, which are themselves part and parcel of "the practice of morality" (33).

In other words, the possibility of immanent critique only comes into view when we remember that 'ideological' *practices* are still *actions*—that is to say, when we remember that there is *no* practice that agents cannot take a moral point of view on. Yet, here again, the same point could be made from the standpoint of normative ethics, where, one might say, practices tend to *only* be evaluated as individual actions. The two standpoints, social theory and normative ethics, must be met in the middle.

As I suggested in Chapter 1, a prominent strand of the post-Kantian tradition recognized the importance of critical reflection on the morally intelligible. Foucault's analyses of processes of discipline, surveillance, normalization, and exclusion, for instance, are frequently interpreted as opposed to moral discourse. Yet, in an interview, Foucault instead describes this project as, in part, a critique of the current scope of application of our normative notion of 'humanity':

> Through these different practices—psychological, medical, penitentiary, educational—a certain concept, the model of humanity, took shape; and this notion of the human has now become normative, obvious, and is taken to be universal.... That doesn't mean that we have to reject what we call 'human rights' and 'freedom', but it entails the impossibility to say that freedom and human rights should be circumscribed within certain boundaries. For example, if you had asked, eighty years ago, if feminine virtue was part of a universal humanism, everyone would have said yes. (1982: 1601)

Foucault affirms, then, a *reflective* approach to the moral concepts 'human rights' or 'freedom'. That is, he does not reject the value of employing such terms wholesale.[61] Instead, he affirms it, albeit by advancing certain criteria for how to employ any given moral universal:

1) A willingness to subject their current scope and 'semantic depth'—their entailment to given action characterizations—to critique, and this on the basis of recalcitrant empirical particulars: a counterexample, say, to the model of 'feminine virtue'. Foucault's own critical works show that practices such as institutionalization, incarceration, and normalization (the process of discriminating the 'normal' from the 'abnormal' on the basis of ever more minute details of personality), while *justified* as applications of moral concepts—'humane punishment' (Foucault 1975), 'good governance' (Foucault 1976)—in fact constitute instances of *dehumanization*, contravening the dignity of subjects relegated outside the scope of the fully human.
2) An openness to invent, imagine, or notice new possible ways in which to define and apply these concepts. In doing so, these concepts retain their normative force in guiding our moral attention and practical reasoning, but remain *malleable,* open to alteration.

I have to be necessarily schematic about this: I lack the space here to lodge a full defense of this way of reading the way a rich and diverse tradition has approached ethical deliberation. But to show how I take this reading to be spelled out in at least one other case, I will close this chapter by considering Beauvoir's understanding of morality.

[61] See Golder (2015) for an important analysis of Foucault's nuanced stance on human rights discourse.

2.4 Beauvoir on moral intelligibility

What is interesting for my purposes is that Beauvoir's critique is cast in explicitly *moral* terms, for which it has come under much criticism: Beauvoir has been accused of a "classical form of voluntarism" derived from her adherence to the existentialist dichotomy between authenticity and bad faith, on which "every feeling of inferiority derives from a free choice" (Le Dœuff 1980: 280). Consequently, Le Dœuff takes the limitations of Beauvoir's analysis to demonstrate that the problem of oppression requires "another perspective than that of ethics or ethical inquiry" (1980: 288). Butler concurs that Beauvoir's invocation of the "doctrine of existential choice" is "assuredly insidious" (1986: 40), while Sara Heinämaa attempts to defend Beauvoir on the grounds, not of her self-professed ethical aim, but of her "phenomenological aim", namely that of describing the "sexual difference" (1997: 30, 22). Other studies of Beauvoir have tended to follow suit, focusing on her contributions to metaphysics, epistemology, phenomenology, or political philosophy rather than her ethics—reinforcing Le Dœuff's claim that Beauvoir's ethical intervention was largely a failure.[62]

Instead, in this section, I suggest that Beauvoir wants to make newly morally intelligible features that have so far escaped the purview of moral attention or practical deliberation. Her intervention should be conceived as operating on what I will go on to call the *reflective* dimension of practical reasoning, rather than the moral dimension alone.

In the *Second Sex*, Beauvoir suggests that women, as the 'second sex', have nearly always

[62] Commentators such as Bauer (2001), Bergoffen (2008), Moi (2008), and Kruks (2012) note in passing Beauvoir's characterization of 'existentialist morality', but refrain from structuring their analysis in terms of Beauvoir's self-professedly *ethical* project; two exceptions are Arp (2001) and Garcia (2018), who, however, fail to draw the distinction I do between reflective (or, in a broader idiom, epistemic, or virtue-epistemic) and moral dimensions of ethical deliberation.

been in a position of subordination to men. This marks a key difference between women and other oppressed groups: while enslaved or colonized communities throughout Western history have typically risked their lives in revolt or rebellion and can collectively remember a time prior to the event of their enforced subordination, the history of women's oppression has featured very few instances of matriarchy, and virtually no instances of collective female rebellion. On her account, there is no singular event in human history during which men gained the upper hand over women: across cultures and across times, women's subordination has been seemingly universal.[63]

As Beauvoir acknowledges, it has generally been presupposed that women's submission is attributable to natural or essential facts about women, specifically their apparent biological, intellectual, or psychological inferiority. In Chapters 1 and 2 of the *Second Sex* (Book 1), she carefully considers these arguments. If one looks at the taxonomy of species in the animal kingdom, those biologically closest to us also have the most rigid sex roles, and the females also appear to be the most biologically disadvantaged. Moreover, Beauvoir claims that, among all mammals, it is women who appear to bear the physiological burden of their reproductive capacity most heavily: "This is the most striking conclusion of this study [of biological sex in the animal kingdom]: she [woman] is the most deeply alienated of all female mammals" (1949a: 44).

Yet while Beauvoir holds that female physiology is important to explain women's current situation, she rejects the view that it determines women's destiny or renders inevitable and unalterable women's status as inferior to men: "I deny that [biological considerations] establish

[63] On this point, I am following Beauvoir's analysis of the anthropological and historical evidence available to her at the time of the *Second Sex.* There are doubtless inaccuracies in her discussion as a consequence; nevertheless, her general point—namely, the separability of biology and history from ethics and human agency—still stands.

for women a fixed destiny. They are not sufficient to establish a hierarchy of the sexes, they fail to explain why woman is the Other, and they do not condemn her to retain this subordinate role forever" (1949: 73). This is because the female subject is "like all humans an autonomous freedom": women are existentially free in a sense that precludes any material fact about them as determining the nature of their existence (1949a: 34).

If women are free in the manner Beauvoir outlines, one might still wonder why she feels that women have virtually never collectively revolted against male dominance. Why does the fact of male superiority appear to be natural, ahistorical, and necessarily true? On Beauvoir's view, the near-universal constancy of women's secondary status is attributable to the fact that it is only in part externally imposed; it is also a situation that, she claims, women accept, self-perpetuate, and in which they are complicit. That is, women in part *choose* to take the "easy path": that of resisting the "ethical claim" of one's self-assertion as a subject, and thus of accepting a limited domain of possible life-shapes and of possible human activity (1949a: 23-24). While both children and women inhabit an "infantile world", "cast into a universe which [they have] not helped to establish", the difference is that "the child's situation is imposed upon him, whereas the woman (I mean the Western woman of today) chooses it or at least consents to it" and thereby manifests a "deep complicity with the world of men" (1947: 47-51). Thus, Beauvoir's indictment extends not just to men, but also to women: "Once a possibility for freedom appears, it is a resignation of freedom not to exploit this possibility, a resignation which implies bad faith and which is actively a fault" (1947: 51).

As I noted above, claims like these have informed a widespread characterization of Beauvoir as committed to an undue moralism when it comes to agents' responses to structural oppression. Beauvoir is taken to have committed the fatal error of mistaking problems systematic

in nature for individual moral failings. Yet this view has missed another possible reading of Beauvoir's insistence on understanding these problems in ethical terms. Indeed, her general emphasis, even when directing her attention to individual behaviors, is to show the many ways in which discrimination against women currently lacks collective moral intelligibility; its particular manifestations in the grain of women's life escape collective notice.

In the section "Childhood", for example, Beauvoir catalogs the many ways in which women are developed to be passive; yet "it is false to claim that this is a biological given; in fact, it is a destiny imposed on her by her teachers and by society" (1949b: 28). The problem, of course, is that it is not *understood* to be a reality imposed on women by others. By understanding the trait of passivity as natural or biological rather than socially inculcated, it is cast outside the domain of human agency, outside the domain of collective choice. Our normative terms are too restricted to accommodate such realities; they therefore call for alteration.

And this is just what Beauvoir's critique consists in, showing how the current scope of moral talk excludes the injustices inflicted on women:

> A gentleman decorated with the Legion of Honor remains an honest man while deflowering a little girl; he has his weaknesses, but who does not? However, the little girl who has no access to the ethical region of the universal—neither a judge nor a general nor a great Frenchman, nothing but a little girl—gambles her moral value in the contingent region of sexuality: she is perverted, corrupted, depraved, and good only for the reformatory. (1949b: 498)
>
> We have seen the hypocrisy of men decreeing abortion to be criminal when every year in France a million women are put by men into the situation where they have to abort. Very often the husband or lover imposes this solution on them, and often these men tacitly assume that it will be used if necessary. They openly count on the woman to consenting to making herself guilty of a crime: her 'immorality' is necessary for the harmony of a moral society that men respect. (1949: 498)

Women, according to Beauvoir's diagnosis, are relegated outside of the moral frame altogether, aside from instances in which the manifold ways in which oppression textures women's everyday

lives is comprehended as evidence of their 'immorality'. It is important to note that Beauvoir's critique cuts deeper than mere hermeneutical injustice: in such a context, the opacities in women's self-understanding cannot even show up as the result of wrongful treatment, since the latter, in the case of women, has not yet been thematized as such.

Beauvoir also emphasizes the ways in which *critiquing* the moral unintelligibility of women's situation has been foreclosed:

> In men's hands reason becomes an insidious form of violence.... They want to confine her in a dilemma: either you agree or you don't; she has to agree in the name of the whole system of accepted principles. If she refused to agree, she would reject the whole system. Yet she cannot allow herself such a dramatic move, since she does not have the means to create another society. (1949b: 496)

In such an oppressive context, when the 'system of accepted principles' is taken to be the pure issuance of reason, it cloaks itself as being unavailable to question. Where an obdurate moral particular might result in the revision of moral principle, the workings of patriarchy deny the need for what I will characterize as reflective judgment. The 'universality' of the operative moral scheme is taken to show that only the moral dimension of practical reasoning is needed; the reflective dimension is rejected as irrelevant for moral deliberation.

These considerations go some way to showing why Beauvoir frames the central intervention of the *Second Sex* as an explicitly *moral* one (1949a: 33). Freedom, according to Beauvoir's existentialism, is to be described in terms of the "expansion" of present existence "toward an indefinitely open future", of "perpetual surpassing toward other freedoms" (33). That is, what autonomy requires, on Beauvoir's analysis, is an attitude of *conceptual openness* towards new possibilities, moral and otherwise, as well as the agent's active involvement in opening up new possibilities for both oneself and others.

Beauvoir herself carries out this interrogation of taken-for-granted concepts, a mode of critique both epistemic and normative. She insists repeatedly on the lack of naturalness of structural social inequality: "A situation of oppression… is never natural"; "one of the ruses of oppression is to camouflage itself as a natural situation since, after all, one cannot revolt against nature" (1947: 102, 104-5). That is, Beauvoir's intervention is one of interrogating oppression as an apparently *natural* state of affairs, revealing it to be dependent on collective *activity*, and thus as a particular salient to practical evaluation.

While commentators have tended to contrast Beauvoir's views from Kant's (see, e.g., Bauer 2001), bringing into view the 'reflective' aspects of Beauvoir's ethical project can, first, demonstrate the relevance of the reflective dimension of practical reasoning; and second, help us to better understand the respective roles of normativity and ethics in the critical tradition. Indeed, one of Beauvoir's central interests was her insistence on the necessity of an *ethics* for existentialism, a project which preoccupied her from her early works *Pyrrhus et Cinéas* and the *Ethics of Ambiguity* to the *Second Sex*, and which had been one of the central oversights of Sartre's conception of existentialism.[64] Beauvoir did not conceive of herself as a Kantian and is at times critical of Kantian morality (the topic of the remainder of this dissertation) as being "abstract and formal" (1949b: 474).[65] This criticism originates in part, I think, from her discovery that the inability to acknowledge the reflective dimension entails the failure to make morally intelligible many frequent forms of injustice, such as women's oppression. While for Kant the expansion of

[64] See discussion in Bergoffen (2008).

[65] However, in articulating the *practical* dimension of her ethics, Beauvoir often invokes Kantian language. For example, she defines her "morality of ambiguity" as "one which refuses to deny *a priori* that separate existents can, at the same time… *forge laws valid for all*" (1947: 24-5, my emphasis; see also references to universality in 1949: 33, 498, 501). In these passages, the moral force of Beauvoir's critique derives in part from its Kantian appropriation of universalizability. See helpful discussion in Webber (2018: 185-7).

one's capacity for reflective judgment is not primarily moral in nature, Beauvoir's analysis of women's oppression as being both practical *and* reflective begins to give us grounds to acknowledge the moral standing of this capacity.

The *reflective* dimension allows us to recast the apparent voluntarism of Beauvoir's appeals to complicity and consent. Her apparently moralistic critiques of individual comportment and actions, both women's and men's, should not be conceived in terms of individual moral reprobation, but as an intervention in the current scope and depth of our moral universals. The effect of Beauvoir's description of these types is *collective* rather than individual, *reflective* rather than narrowly practical: it alters the morally intelligible. The conclusion Beauvoir aims her readers to take away from the discussion is that they need not—and indeed *ought* not—accept the limits imposed on them without creatively experimenting with them, broadening them, and opening them up to question.

2.5 Conclusion

In this chapter, I have drawn attention to how and why we might conceive of a 'morally unintelligible' particular, or one resistant to the universals available; I have argued that this notion evades the established frameworks in philosophy of hermeneutical injustice and imaginative resistance. The morally intelligible, I have claimed, is open to critique, and the ethical project of Simone de Beauvoir provides one example in which this has been thematized as a problem.

Of course, I have had to put aside other complex issues that arise when making sense of moral intelligibility: first and foremost, the role of power. That is, what shows up as morally intelligible is not arbitrary; it reflects the operations of relations of power in society, albeit in

sometimes unexpected ways. Unlike in the epistemic injustice conception, power can at times result in *excesses* of knowledge rather than hermeneutical 'gaps' and 'opacities'. So I do not want to be read as denying the role of power in the picture I have just been presenting. My aim here has been circumscribed to the metaethical commitments involved in thinking about what I have attempted to pick out as 'moral intelligibility'; my view, generally speaking, does not excise the import of power, even if I have had to momentarily leave it unaddressed.

Chapter 3: The Reflective Dimension of Practical Deliberation

3.1 Universalism and particularism

In this chapter, I defend a reworked Kantianism that espouses an ethics of principles, but also bestows a role to the *revision* of moral principles. I claim that Kant has the resources to account for the role of moral revisability in his theory of *reflective judgment*, the cognitive power required for forming new concepts or principles on the basis of experience. On the view I develop here, reflective judgment involves both the incorporation of *new moral particulars* into a standing moral conceptual repertoire (as we already began to see in the previous chapter) and the *revision of moral universals* on the basis of such discoveries (Chapter 5). Below, I set up these discussions by considering Kant's account of teleological judgment in the *Critique of Judgment*, where the power of judgment's function in constructing maxims on the basis of experience is isolated from the guidance of the moral law. The example of teleological judgment enables us to isolate a *reflective dimension* of practical reason operative alongside the moral dimension generally taken to exhaust Kant's practical philosophy. I show that reflective judgment is implicated in the formulation and revision of maxims, understood in both senses of 'maxim' predominant among commentators: both in terms of moral principle and intention in action.

How do these scholarly discussions relate to the concerns in critical theory I referred to in the prior two chapters? If one of the historical tasks of critical theory has been to identify and explain how oppressive social practices shape our concepts, principles, and modes of reasoning, Kantianism instead takes these to be universal and independent of experience. The categorical imperative is understood to be unchanging and accessible to all agents, no matter their social

circumstances. So there would seem to be little point of contact between the two perspectives; indeed, one tenet characterizing the post-Kantian European tradition has been the objection that Kantian universalism cannot account for the influence of social factors on moral phenomenology.[66]

More recently, critics have brought attention to the ways in which universalist rhetoric, as much as it has contributed to the struggle for human rights and other important advancements, has also historically been used to prop up scientistic racist ideology, colonial policies of assimilation, and authoritarian political regimes (Foucault 1984a: 1393-4; Balibar 1994: 195; Golder 2015). And both contemporary 'critics' of race and gender and Kant scholars have noted that the employment of the categorical imperative has been productively broadened since the 1700s to include agents, such as women and minorities, that Kant himself precluded as fully 'rational' (Butler 1996: 46; Mills 2005: 126; Allais 2016: 3; Kleingeld 1993, 2007).

Do these charges undermine the force of universalism? Many among Kant's detractors have taken such criticisms as grounds for rejecting universalism altogether, while Kantians have largely left them unaddressed. Against both of these camps, I find such criticisms to show that moral imperative, rather than being rejected outright, should be extended beyond the application of principle to a regular reexamination of the presumed scope of principle, as well as the presupposed entailment from principle to maxim (intention) of action. As Louden writes, "Often, of course, we are not even aware of all of the different ways in which our own attempted articulation of a universal ideal fails to escape its local cultural encodings" (2000: 179), such that, as Butler argues, universality should be considered a "contingent and necessarily incomplete

[66] See discussion in Chapter 2 and the Conclusion. By 'moral phenomenology' I mean not only moral judgments or moral reasoning, but also moral psychology and the moral emotions; see Bagnoli (2011b).

notion" whose further articulations remain indeterminate (1996: 45-6).

As I attempt to show, particulars can directly *inform* moral principle. The relation between universal and particular can be understood as a two-way dependence rather than in terms of mere application.[67]

Objections to universalism are relevant not just to critical theory, but also to moral theory, which has generated longstanding debates on moral particularism and generalism. Those who characterize virtue as sensitivity to circumstance or context (McDowell 1978, 1998; Nussbaum 1986, 1990), mode of perception or way of seeing one's situation (Murdoch 1970), or responsiveness to novel combinations of reasons in context rather than applications of principles (Dancy 1993, 2004), each charge Kantians with failing to capture these aspects of morality. As Nussbaum characterizes Aristotle's practical philosophy, "Excellent choice cannot be captured in general rules, because it is a matter of fitting one's choice to the complex requirements of a concrete situation, taking all of its contextual features into account" (1990: 71-73). What generalist approaches have failed to address is the moral role of grasping particulars to which no general moral rule or principle definitively seems to apply. In such cases, what is needed is the "ability to recognize the salient features of a complex situation", such that the fitting moral response can spontaneously be "improvised" (1990: 74).

I contend that the choice between moral sensitivity and moral principle is a false one, and I show below why I take Kant's theory of reflective judgment to bring this out. This position is

[67] On this point, I am in agreement with a claim advanced by Amia Srinivasan: "The most compelling philosophers… recognise the reciprocal nature of the relationship between the universal and the particular. They do not seek merely to organise the particulars under various universal concepts. They also ask themselves how the universals can be changed and shifted to better fit with the particulars" (2018: 1415).

one I take not only on historical grounds, but also on methodological ones. In embracing particularism over universalism, one also rejects the moral universals, such as 'racial justice', which have been so helpful in fighting oppression. Instead, one risks relativism, forfeiting the ability to "affirm that, objectively, women and people of color are indeed oppressed—not merely that they believe they're oppressed" (Mills 2004: 173-4). There is a third option: such struggles need not involve the affirmation or rejection of moral universals as such, but instead the *revision* of universals on the basis of particulars not adequately captured by the terms available. If we can show this, we can reject the operative dichotomy between universalism and particularism in normative theory.

I suggested in Chapter 1 that there is a way in which the version of Kantianism I am proposing amounts to a form of *immanent critique,* with social experience prompting reflection on the status and formulation of our moral principles and normative concepts, guiding our attention to where they fall short. However, unlike predominant Hegelian models of immanent critique, the conception I develop is one not committed to a teleological conception of history, since the revisions at issue are only taken to be partial and fragmentary. The prospects for a similar conception of 'immanent critique' are also pressing for moral particularists, albeit articulated from a different point of view. Herman charges particularists with being unable to address how one's own sense of moral sensitivity can be critiqued: "What raises a question… is the apparent absence at the limit of a way to criticize [moral] sensitivity itself—for it to take itself as the object of its own critical regard" (2007: 26). Herman terms the new frontier for virtue ethics "the problem of *new* saliences: how to see that the moral facts have changed" (28). That is, the most pressing issue for virtue ethicists and particularists is how to account for something like a self-directed immanent critique: how to bring about a transformation of one's own normative point of

view, from *within* that point of view.

A reason this remains an open problem for particularism can be traced to the weight it gives to moral receptivity to the exclusion of considering how moral receptivity *reinforms* our initial normative concepts and principles. Since particularist approaches reject the role of moral principle altogether, the latter concern is, from the outset, positioned outside their purview. Yet, as I claimed in the previous chapter, it is the latter—the contestation of our available moral universals—which is, in many instances, the aim of a critique.

In what follows, I start by taking up an influential objection to Kantian ethics: the 'problem of relevant descriptions'. I claim that this problem is not merely a defect in Kantian approaches, but captures an important feature of practical deliberation. I turn to several predominant Kantian strategies for responding to this problem proposed by Onora O'Neill, Barbara Herman, and Hannah Arendt, showing why I take them to fall short on both interpretive and ethical theoretical grounds. Next, I flesh out an account of the reflective dimension of practical deliberation: while interpreters have tended to accord a reflective dimension to theoretical judgment rather than practical, I point attention to passages where Kant's view allows for extending a role to reflectivity in practical judgment. I suggest that the operations of teleological judgment, particularly its role in forming maxims to guide empirical inquiry, enables us to isolate the reflective dimension, since, unlike other judgments, teleological judgment lacks a corresponding determining or moral dimension. I conclude by introducing a distinction between two senses of 'maxim': 'i-maxims', or particular intentions of action, and 'p-maxims', or general moral principles. I suggest that the reflective dimension pertains to both domains: the accommodation of resistant particulars (to which I return in Chapter 2) and the revision of moral principle (Chapter 5).

3.2 The problem of relevant descriptions

One major criticism of Kantian universalism stems from what Onora O'Neill (1975) terms 'the problem of relevant descriptions', a charge initially formulated by Elizabeth Anscombe. Anscombe contends that since the permissibility of maxims depends on the description of the action one employs, and since any given action admits of innumerable possible descriptions, the universalization procedure for determining permissibility is "useless" (Anscombe 1963).[68] If an agent is poisoning a group of Nazis by putting toxic chemicals into their water supply, Anscombe argues that he can justifiably describe his action as 'moving his arm', 'operating the pump', 'replenishing the water supply', or 'poisoning the inhabitants'. The same action admits of all four possible descriptions.

The problem of relevant descriptions poses a concern for Kantian ethics, because the results of the categorical imperative procedure differ depending on the description supplied. The specific action description 'I am moving my arm' can be universalized across all agents, but the more general description 'I am fatally poisoning the inhabitants of this house' results in a contradiction, and therefore cannot be universalized.[69] In Anscombe's example, the agent in question is not just poisoning arbitrary individuals, but Nazis, "party chiefs… engaged in exterminating the Jews and [who] perhaps plan a world war" (37). Thus, the agent's most general action description is broader than 'poisoning the inhabitants' (or 'killing'); his action can be redescribed as 'saving six million people from extermination' or 'averting a world war'. Yet

[68] See also Anscombe (1957).

[69] That is, it fails the *contradiction in conception test*; see O'Neill (1975): 143-173.

redescribing the action at this even more general level arguably results in a moral conflict between two competing duties: the duty not to end human life against the duty to stop mass genocide if doing so is within one's power.

Anscombe objects that Kant's moral philosophy cannot tell us how to *formulate* our maxims, but only how to adjudicate the morality of already formulated maxims.[70] Since the same action admits of many possible descriptions—in other words, many possible maxims can apply to the same action—Kantian moral theory cannot speak to important morally relevant content that pertains to the very question of moral admissibility the categorical imperative was intended to adjudicate.

Moreover, the problem of relevant descriptions is not just theoretical; it bears directly on practical judgments that arise in ordinary deliberation, since not only is it the case that agents *can* offer varying descriptions of the same action, but it is also phenomenologically the case that they often *do.* As Allen Wood notes, "My intentions can be described—not artificially but quite accurately—in greater or lesser detail, at higher or lower levels of generality, by mentioning or omitting this or that feature of the situation" (1999: 103). Indeed, the problem of relevant descriptions can be taken to capture an ordinary moral dilemma: what is it, really, that I am doing (or propose to do)? And what is the most accurate description of what it is that I am doing?

While this problem is one posed to Kantian approaches in ethics, I take the problem to be more general. Take Dancy's characterization of the following groupings of reasons for acting:

1. If you are causing someone pain, you are doing something wrong. ($p \rightarrow q$)

[70] Anscombe's charge may be seen as a way of restating the long-standing empty formalism objection famously posed by Hegel to Kant: namely, that Kant's moral theory is too abstract and formal to have anything to say about the content or substance of practical deliberation, and thus cannot guide action (Hegel 1821: §134).

2. If *p* and the pain is a statutory punishment for a recognized offence, you are not doing something wrong. ($(p \,\&\, r) \rightarrow -q$)

3. If *p & r* and the punishee was unjustly convicted, you are doing something wrong. ($(p \,\&\, r \,\&\, s) \rightarrow q$) (Dancy 2008: 8-9)

Dancy takes reasoning chains (1-3) to tell in favor of particularism and against universalism, since, on his view, a particular feature (such as *p*) can count in favor of an action in one context, such as (1), and against in another, such as (2). Another way of putting the point, however, is in terms of the problem of relevant descriptions: if I choose to describe my action as 'I am causing someone pain', I will, following the reasoning in (1), conclude my action to be morally wrong; if I choose to describe my action as 'I am inflicting statutory punishment for a recognized offense', I will, following (2), take it to be morally justifiable.

In an important sense, the nature of my reasons depend on how I describe them—namely, on generals, not particulars.[71] If I contest the concept 'statutory punishment', for example, I may not recognize the reasoning chain in (2), even if another agent in the same situation would take (2) to accurately describe the pain to be inflicted; I may only recognize a description of the sort given in (1). Similarly, if I am aware of the striking racial disparity in criminal convictions, this

[71] Thus, relevant description is a feature of moral deliberation in general—it is not just an objection to Kantianism. As Peter Winch claims, "Suppose it is said [by an observer, O] of a certain person, N, that he voted Labour at the last General Election because he thought a Labour government would be the most likely to preserve industrial peace… then it should be noted that the force of O's explanation rests on the fact that the concepts which appear in it must be grasped not merely by O and his hearers, but also *by N himself*. N must have some idea of what it is to 'preserve industrial peace' and of a connection between this and the kind of government which he expects to be in power if Labour is elected…. The acceptability of such an explanation is contingent on N's grasp of the concepts contained in it. If N does not grasp the concept of industrial peace it must be senseless to say that his reason for doing anything is a desire to see industrial peace promoted" (1965: 43-44). A similar example is invoked by Charles Taylor: "An activity of marking and counting papers has to bear intentional descriptions which fall within a certain range before we can agree to call it voting…. That some practice is voting or negotiation has to do in part with the vocabulary established in a society as appropriate for engaging in it or describing it" (1971: 26).

may play a role in whether I ultimately take the 'recognized' nature of the offense to be transparent or the 'just' nature of the punishee's conviction to be self-evident. The particular features that one might take to count *for* an action in one circumstance and *against* in another are contingent on the discursive resources available, themselves often subject to live contestation.

A particular feature of a moral situation is never *just* particular; it is always subject to the nature of the concepts, the universals, in terms of which a subject represents a particular feature of her context of action.[72] The upshot here is that the description one formulates of one's action (or one's reasons for action) bears on the nature of one's moral conclusions. Indeed, the problem of relevant descriptions constitutes, in my view, one way of rearticulating an aspect of Williams' (1985) account of the role thick ethical concepts play in practical reasoning. The concepts elaborated in reasoning chains (1-3)—'statutory punishment', 'recognized offence', 'unjust conviction'—are both truth-tracking and normatively weighted: they not only pick out a given feature of the world, but carry an implicit ethical appraisal of it, where these two functions cannot be neatly prised apart. Invoking such concepts in the formulation of a given reason therefore already implies a given moral evaluation. But, if this is the case, the discursive articulation of particularity cannot be represented merely as iteratively introducing new reasons, new 'particulars': reformulation alters the nature of our reasons; it does not multiply the discrete number of reasons.[73]

---

[72] The nature of the description given, as the critique literature has recognized, is therefore itself normative, a potential target for ethical evaluation: "Ideology critique… challenges not only the truth/falsity of our beliefs about 'the facts,' but also the terms used to describe the facts. For example, redescribing meat as the flesh of tortured animals, or racial profiling as racial targeting, matters" (Haslanger 2017); "Ideology critique… is not least of all a matter of unmasking the normative character of certain descriptions" (Jaeggi 2009: 72).

[73] In redescribing 'gay marriage' as 'sin' or 'wifely duty' as 'marital rape' (Srinivasan 2018), the redescription does not amount to a new reason to be added cumulatively to my reason as I first described

3.3 Responses to the problem of relevant descriptions

Onora O'Neill

O'Neill (2018) has attempted to respond to the problem of relevant descriptions by objecting to the invalidity of its point of view: she charges particularists with taking up the stance of spectators or observers of action rather than that of *doers* of action. Thus, such accounts rest on a confusion between theoretical and practical cognition, since morally 'perceiving' or evaluating situations involves theoretical judgment rather than practical: "A fair amount of ethical writing… has tried to construe ethical judgement as reflective. This is plausible only for ethical judgement about existing or past cases, where the particular to be judged can be given. It is not possible for practical judgements about what to do, which do not seek to 'appraise' or 'attend to' or 'evaluate' existing situations" (124n).[74] Thus, practical judgment is "neither determinant nor reflective", since it involves "indeterminate 'universals' (concepts, rules, principles, laws), but here no particular is 'given'" (2018: 123). Instead, practical judgment "must be undertaken *before* the relevant particular exists: we cannot pick out future act-tokens" (2018: 123).

As I elaborate below, this approach is ultimately untenable, since the role of evaluating context cannot be bracketed from practical judgment. In fact, this position is in tension with O'Neill's influential early characterization of maxims, which can "be individuated only by referring to a person, and so to a place and a time" (1975: 62). A maxim, on this definition, necessarily includes the situation in which one acts, entailing the possibility that one's maxim can be "inappropriate to the situation" (247). If sensitivity to circumstance no longer enters under the umbrella of practical judgment, maxim formation risks appearing as what O'Neill terms a merely

it, but fundamentally alters the reasons I take myself to have in the first place.

[74] O'Neill (2002) gives a much more detailed counterargument to the role of reflective judgment in ethics.

reflective "preliminary" to action, an "early warning system"—even though O'Neill takes a maxim, which she understands to constitute an action description, to be an indispensable component of the moment of action itself (1989: 184; 2018: 18).

By O'Neill's own lights, then, it seems unpromising to abstract practical judgment from an evaluation or appraisal of the context in which the action is to be performed. It remains unclear what it would mean to act in abstraction from consideration of context; indeed, without any sense of one's situation, acting on an intention would seem to no longer be possible.

Barbara Herman

Herman (1993, 2007) has employed the opposing strategy: rather than deny the role of discerning salience or evaluating context altogether, she claims that Kantians must explicitly incorporate the notion of salience into their accounts of practical reasoning by way of what she terms 'rules of moral salience' (RMS). Herman argues that the results of the categorical imperative are subject to what agents take to be relevant content to be incorporated into their deliberative procedure, or into their RMS. In order to be able to apply a given moral principle, agents have to know which features of their proposed action are of moral import—or, even more basically, they must first know whether a given situation calls for action or not. Thus, pure moral principle (including the categorical imperative) is not enough; agents must also be able to discern "the moral marks of their actions and circumstances" (1993: 78; see also 1996). Yet RMS are "not themselves moral rules"; they are rules for judging context inculcated through socialization and moral training, and as such are fundamentally contingent (1993: 75).

While the strategy of embracing salience fares much better when it comes to broadening Kantianism to account for moral responsiveness, it lacks much, if any, textual basis in Kant's own

expressed views, posing a problem for the Kantian grounding of RMS: "If they are independent of the CI, the unity of the Kantian system is compromised, consisting partly of formal procedure and partly of something like preprocedural intuition or convention" (84). That is, Herman maintains that RMS must have a theoretical grounding in Kant's moral theory; otherwise, our interpretation of his views will start to look incoherent, taking the form of a hodgepodge of Kantian and non-Kantian elements.

Since this grounding cannot be located in the categorical imperative procedure, which applies only to already formulated maxims, Herman traces RMS to the moral law as a fact of reason. Kant formulates the fact of reason in KpV as follows: "It is the moral law… of which we become immediately conscious as soon as we construct maxims for the will… Reason exhibits it as a ground of determination not to be outweighed by any sensuous condition" (5:29). From this, Herman contends that she can derive two claims: first, that we are susceptible to moral experience; and second, that we are moral agents who know that we are free, acting in accordance with moral demand even against sensuous inclination (1993: 85). These, she claims, are aspects of moral experience that are not reducible to the categorical imperative procedure, and that can therefore play a role in determining moral salience.

However, Kant clearly had nothing like RMS in mind when he articulated the notion that the moral law appears to us as a fact of reason. The context in which he does so is as a premise in an argument leading directly to FUL, the formula of universal law (5:30), one of the formulations of the categorical imperative. Indeed, in this passage FUL is given the title "Fundamental Law of Pure Practical Reason" without any explicit transition between his introduction of the 'moral law' and FUL as the 'fundamental law', raising the possibility that what Kant means to pick out with the term 'moral law' in this context just *is* the most basic

formulation of the categorical imperative.

Even if we take the moral law and the categorical imperative to be relevantly distinct, it seems clear that the function of the fact of reason argument is to set up the introduction of the categorical imperative rather than any moral considerations external and prior to it. This undercuts Herman's claim that the moral law can be used to ground RMS independently of the categorical imperative procedure. Kant says nothing in the moral law passage of KpV about the features of experience that we take to be morally relevant and thus factorable into a given maxim; instead, he describes our *inner phenomenology* as moral agents whenever we propose to act. Finally, Kant's discussion in the relevant passage of KpV of the moral law as fact of reason is very brief. The discussion prior to introducing FUL is limited principally to the one sentence Herman explicitly cites, and therefore offers meager textual grounds for situating a theory of RMS.

Resting RMS on the moral law as fact of reason therefore has significant interpretive problems, appearing to bolster skeptics' suspicions that Kantianism has no robust theoretical resources to make sense of the particularity and texture of moral experience,[75] while Herman's insistence that Kantianism *must* have such resources risks appearing dogmatic. For those who do not already subscribe to Kant's moral theory, Herman does not supply much reason to think Kant may offer richer textual resources to make sense of such issues than other non-Kantian attempts do.

But Herman's approach also runs into independent philosophical problems, ones which, as I show in Chapter 4, Kant was sensitive to, and ultimately appears to have revised his account of practical judgment in light of. Herman argues that our *a priori* moral commitment is not

[75] See "Persons, Character, Morality" and "Moral Luck" in Williams (1981); Winch (1965).

sufficient to determine conclusive moral results, since we first need to know which features of a given situation are morally relevant. While Herman accurately establishes the necessity of moral salience or something akin to it for Kantian moral theory, her appeal to additional rules—not 'moral rules', but "rules of relevance", rules for practical judgment rather than rules issuing directly from the categorical imperative procedure (1993: 75)—poses a threat of regress.

To see this, it's helpful to compare the practical case to the theoretical. As Hannah Ginsborg argues, the introduction of additional rules in theoretical judgment can never adequately explain how it is that we come to *form* cognitive rules (or concepts) in the first place. In a famous passage from LJ, Kant describes how we form the empirical concept 'tree' on the basis of the logical operations of comparison, reflection, and abstraction:

> I see, e.g., a spruce, a willow, and a linden. By first comparing these objects with one another I note that they are different from one another in regard to the trunk, the branches, the leaves, etc.; but next I reflect on that which they have in common among themselves, trunk, branches, and leaves themselves, and I abstract from the quantity, the figure, etc., of these; thus I acquire a concept of a tree. (§6, note 1; 9:94-5)

Ginsborg holds that this explanation is necessarily incomplete, since it presupposes that we already have concepts such as *leaf, branch,* or *trunk*; given that "we need to explain the acquisition of these concepts on the basis of further concepts… a regress threatens" (2015: 151). Moreover, she argues that it is unclear how it is we know to group certain features into concepts rather than others. If spruces, willows, and lindens "typically have in common that they lack edible fruit, that they afford a degree of shelter from the rain, that they are composed of woody material, and that insects live in them", why is it that the concept we arrive at is one that excludes fruit trees and includes wooden houses (152)?

As this example shows, concept formation cannot be accounted for solely in terms of

preexisting rules. Instead, grasping an empirical concept such as 'spruce' or 'willow' requires that we first come to recognize the grouping of certain features—leaves, bark—as if they *should* belong together, and thus as forming *one* entity warranting cognitive unification. In other words, the subject has to take her associations to be subject to "a normative standard applicable to everyone" (Ginsborg 2015: 154, 162). Conceptual rules, rather than being grasped antecedently to experience, are acquired along with the disposition to associate certain features of representations with one another in such a way that conceptual synthesis thereby becomes possible, which in turn requires that subjects adopt a normative attitude towards their own mental activity (163). In short, concept acquisition depends on the *reflective* dimension of judgment.

Analogously in the case of practical reason, Herman takes RMS to respond to the problem of why certain features are grouped into one moral principle rather than another. In more recent work, Herman has focused on cases of 'moral improvisation', where new moral duties can be contingently brought into being, such as the adoption of anonymous review in academic refereeing practices or the screening of orchestral auditions (2009: 33). Once brought into being, Herman argues that the improvised solution constitutes a moral obligation, no less binding for the fact that it was contingently introduced, such that numerous alternate solutions (point systems, quotas) might have constituted appropriate responses to the problem. Herman's point is generalizable for the improvisation of other thick moral concepts, such as 'sexual harassment' or 'homophobia'. The features such terms pick out may have previously been grouped into altogether different interpretations of moral reality (a harmless joke, fun in the workplace), factoring into action types that are no longer acceptable once the improvisation is in place.

If the parallel between theoretical and practical uses of judgment is apt, the problems inherent to understanding salience as rule-governed are as present in the practical sphere as in

the theoretical.[76] We acquire moral concepts (selfishness, generosity) along with theoretical concepts (spruce, willow), and new moral rules (screening auditions, anonymizing peer review) along with cognitive rules for judging reality.

As I understand it, the problem Herman (2007: 26-8) emphasizes of how to critique moral salience, to which I referred above (§3.1), is linked to the mistaken understanding of moral salience as inherently rule-governed: if agents are taken to be committed to *both* moral rules *and* rules of relevance to guide the application of moral rules, it is hard to account for how it is that rules can change, how agents can be aware that they have changed, or how agents can be actively involved in such changes.

Hannah Arendt

In contrast to O'Neill and Herman's respective strategies, each of which overlooks the role of reflective judgment, one ethical approach famously embraces it: that of Hannah Arendt. In particular, Arendt uses the model of aesthetic judgment to counter what she takes to be Kantianism's overly rigoristic reliance on the application of rules—as in the example of Adolf Eichmann, who took himself to be acting on Kantian principles even when organizing Nazi death camps. What is needed, she argues, is not moral principle, but the exercise of judgment, as in the case of judging a beautiful object for which, Kant claims, a definite concept is lacking: "The faculty of judgment, which one may call, with some justification, the most political of man's mental abilities… is the faculty to judge particulars without subsuming them under those general rules which can be taught and learned until they grow into habits that can be replaced by other habits and rules" (1971: 446; see also 1963: 136, 295). Arendt can therefore be taken to concur

[76] See my discussion of passage A133/B172 in Chapter 4.

with a key line of argumentation in Aristotelian and particularist accounts of moral perception: genuinely moral action requires the judgment of particulars that should not ultimately be subsumed under an externally imposed 'general rule' or habit.

Yet, once more, Arendt's view faces both interpretive and broader philosophical challenges. On the first, she simplifies aspects of Kant's account of judgment, missing a more sophisticated reading of Kant's position. She claims repeatedly in her lectures on Kant that "judgment of the particular—this is beautiful, this is ugly, this is right, this is wrong—has no place in Kant's moral philosophy" (1970: 15, see also 53-5, 61, 67). Arendt takes the resistant nature of particulars to generalization to be incompatible with Kant's own views.[77]

In rejecting Kant's moral framework, however, Arendt effaces the complexity of Kant's account of practical judgment. Relying on the apparently non-conceptual nature of aesthetic judgment fails to shed light on how *new* moral concepts can be issued, since moral reasoning, on an Arendtian picture, only takes the form of judging each particular in independence from all others. Instead, attending to the complexities of Kant's account of practical judgment shows that the options are not restricted to either applying rules that are already given or treating particulars solely as particulars. As I attempt to show, the relation between particular and universal can be understood as a two-way dependence rather than in terms of mere subsumption or application.

[77] Arendt does not acknowledge the fact that, for Kant, a given moral principle or maxim must be *autonomous*, which is to say self-given by the agent who proposes to act on it. It cannot be mere habit, or externally imposed by a culture or authority, and must issue from within, namely from the moral law, rather than from without. This criterion alone would have ruled out Eichmann's actions as genuinely moral. Nevertheless, I do think Arendt articulates an important ambiguity in Kant's characterization of heteronomy in the sense of being ruled by sensuous inclination, and heteronomy in the sense of inheriting externally imposed doctrines one does not rationally assess.

3.4 An alternate approach

Arendt's appropriation of Kant is particularly important for critical theory, where it has been used as a model for invoking the relevance of 'reflective judgment' in social critique, as I myself do in this dissertation (Ferrara 1998, 2008; Lara 2008; Zerilli 2016). While the governing terms of discussion (reflective judgment, social critique) are consistent between my approach and neo-Arendtian approaches, the latter limit their discussion of reflective judgment to *aesthetic* judgment—and more precisely, to a reading of Kantian aesthetic judgment as inherently non-conceptual and particularist. In doing so, they leave open the possibility of considering the broader role Kant accords to reflective judgment in both Introductions of the *Critique of Judgment* not just in judging particulars, but in forming new generals or universals (concepts, principles) on the basis of particulars:

> To reflect… is to compare and to hold together given representations either with others or with one's faculty of cognition, in relation to a concept thereby made possible. [*Reflektieren (Überlegen) aber ist: gegebene Vorstellungen entweder mit andern, oder mit seinem Erkenntnisvermögen, in Beziehung auf einen dadurch möglichen Begriff, zu vergleichen und zusammen zu halten.*] (20:211; First Introduction)
>
> The power of judgment in general is the faculty for thinking of the particular as contained under the universal [*das Besondere als enthalten unter dem Allgemeinen*]. If the universal (the rule, the principle, the law) [*die Regel, das Prinzip, der Gesetz*] is given, then the power of judgment, which subsumes a particular under it… is determining. [*bestimmend*] If, however, only the particular is given, for which the universal is to be found, then the power of judgment is merely reflecting [*bloß reflektierend*]. (5:179; Second Introduction)

Thus, while determining judgment, one major topic of the first *Critique*, subsumes the experiential manifold under already given cognitive principles (such as the categories), reflective judgment forms *new* 'universals' in encountering matter that cannot be subsumed under the existing repertoire of rules.

This role, as I turn to below, is particularly clear in Kant's discussion of *teleological*

judgment, the other main paradigm for reflective judgment alongside the aesthetic. While Arendt appeals to aesthetic judgment on the basis of its resistance to generalization, teleological judgment furnishes an account of how new content encountered in the course of experience can lead us to productively *revise* our preexisting moral concepts and principles. Teleological judgment of a particular alters the universal that we initially started with, or provides a new one altogether.

While the case of teleology makes this aspect of reflective judgment most explicit, aesthetic judgment is not solely particularist or nonconceptual, either. In the aesthetic case as well, Kant makes room for the recalcitrance of particulars to universals,[78] but inevitably insists that recalcitrant particulars[79] normatively call for the creation of new universals: "The satisfaction in the beautiful must depend upon reflection on an object that thereby *leads to some sort of concept* (it is indeterminate which)" (5:207; my emphasis). So, while it is right to say that aesthetic particulars cannot be *subsumed* under existing universals, it is incorrect to conclude that they have no bearing on universals at all. Instead, they play a constructive role in forming *new* universals: concepts whose distinctive shape depends on the discovery of *that* particular rather than any other. Kant says that the "new rule" disclosed in this fashion "could not have been deduced from any antecedent principles or examples" and "*for that very reason* is original" (5:317; my emphasis).

This process is one clearly governed by the normative demand of broadening the current

[78] See, e.g., Kant's account of free beauty, as contrasted with dependent or adherent beauty, which is free in that it does not 'depend' on an already given concept: "In order to find something good, I must always know what sort of thing the object is supposed to be, i.e., I must have a concept for it. I do not need that in order to find beauty in something. Flowers, free designs... do not depend on any determinate concept, and yet please" (5:207); or Kant's definition of beauty in §6: "The beautiful is that which, *without concepts*, is represented as the object of a universal satisfaction" (5:211, my emphasis).
[79] I am grateful to Amia Srinivasan for suggesting this term.

bounds of discursivity. In responding to unsubsumable particulars, subjects attempt to grasp all that is currently "unnameable [*das Unnennbare*]" when adhering to "the mere letter of language"—all that eludes one's determinate conceptual schemes (5:317). Moreover, since the "aesthetic ideas" beautiful objects prompt in the spectator are ones "no language fully attains or can make intelligible [*unverständlich*]", they give her the "impetus to think more… than can be comprehended in a concept" (5:314-5). This process, while constructive, is thus one which is never complete.[80]

The example of maxim construction in teleological judgments

For both aesthetic and teleological judgment, then, particulars never remain *merely* particular; in both cases, they give shape to new universals.[81] However, given that Kant's account of aesthetic judgment is frequently (mis)taken to preclude conceptualization or universalization, the connection to practical reasoning is more perspicuous in the case of teleological judgment. After all, Kant makes explicit that teleological judgment constructs not just new *concepts,* but *maxims*, both *a priori* and empirical—for example, the maxim to judge nature as a purposive system, or to judge a natural being as a purposively ordered organism rather than as an arbitrarily arranged composition of parts. If practical judgment furnishes specific, 'empirical' maxims in

[80] The role of aesthetic judgment in expanding conceptual schemes, particularly as a helpful analogy to moral judgment, is also a theme repeatedly emphasized by Iris Murdoch; see, e.g., Murdoch (1970: 325).
[81] This point—that particulars are discursively mediated and vice versa—converges with Kukla's account of moral perception as conceptually articulated: "We can see nothing if we have not developed a space of concepts that would let us make sense of what we see. We need to be able to see something *as* placed within an articulated space of sense in order to be able to see it at all…. The objects of perception cannot be mute sense data or raw particulars; they must have the right kind of structure to be taken up by reason and to serve as an opening for inference and a ground for warrant" (2002: 324, 335). Since particulars show up to us within a discursive space of reasons, particulars must themselves be able to *inform* that discursive space.

order to apply *a priori* moral principle to particular cases, the example of teleology makes clear how this operation can be considered one of reflective judgment, whose role, as Kant specifies it in both Introductions, is precisely to generate 'universals' (maxims) from 'particulars' (moral experience). In teleological judgment, unlike in practical judgment, the power of judgment's role in constructing maxims on the basis of experience is isolated from the guidance of the moral law.

What does the process of teleological judgment consist in? In order to initiate a course of teleological reasoning, Kant claims that we must presuppose a determinate principle.[82] Kant characterizes reflective judgment as "a mere faculty[83] for reflecting on a given representation, in accordance with a certain *principle*, for the sake of a concept that is thereby made possible" (20:211, my emphasis). While in practical judgment the normative principle is already given, in teleological judgment a principle must be set as a heuristic in order to guide the procedure of judging. This principle initially lacks justification and must therefore be merely presupposed, but is gradually vindicated in the course of empirical discovery.

The relation between practical and merely reflective judgments therefore parallels the relation Kant establishes between determining and merely reflective judgments. Kant claims that "when we *reflect*… we need a principle just as much as we do when we determine, where the underlying concept of the object prescribes the rule to judgment and so takes the place of the principle" (20:211). While in the case of determining judgments, we have a concept, a cognitive

[82] This point introduces a complication to Ginsborg's (2015) account of the subjective normativity of our cognitive activity considered above. Kant does not claim that what is warranted to initiate teleological reasoning is the adoption of a normative attitude in general, as Ginsborg claims. Instead, he claims that we need to posit 'maxims', or determinate principles—heuristics, rather than a general normativity of reasoning.

[83] This constitutes yet another instance in which Kant opposes 'merely' reflective judgment to cognitive judgments, which have both determining and reflective (or, as I have argued, *practical* and reflective) aspects.

rule, that can guide the activity of judgment, reflective judgments lack any conclusive conceptual determination and must therefore *first presuppose* a heuristic rule or principle.

Kant characterizes these heuristic principles as 'maxims'.[84] As examples, he cites common sayings, or "stock formulae", such as: "Nature takes the shortest route—she does nothing in vain—she makes no leaps in the manifold of forms—she is rich in species but sparing with genera, etc." (20:210; compare 5:182). But because such principles "do not say what happens… but rather how they ought to be judged", they are intrinsically normative, claiming "logical objective necessity", in a sense that would not be possible for solely empirical principles (5:182). As a result, the initial guiding principle that nature is purposive is transcendental: "Thus the purposiveness of nature for our cognitive faculties and for their use… is a transcendental principle of judgments" (5:182). Here, too, the heuristic guiding principle of teleological judgment contrasts with the guiding role of the moral law in practical judgment: while both principles are transcendental and normative, these teleological maxims only *claim* objectivity, while the moral law *is* objective. Unlike the moral law and the categorical imperative that follows from it, transcendental teleological maxims fall short of objectivity because they only govern our own cognitive procedure in judging nature; we are not justified, Kant thinks, to apply them to nature from the outset.

Alongside these transcendental maxims, Kant claims that there are also more specific maxims that arise in the course of empirical investigation, thus deriving from experience. For example, the thought of a 'natural purpose', or an organism, which comes up in §66, several steps into Kant's argument for the purposiveness of nature, follows from the maxim of internal

---

[84] See, e.g., 20:205, 20:219, 5:182, 5:184.

purposiveness: “This *principle is of course to be derived from experience*, that is, experience of the kind that is methodically undertaken and is called observation” (5:376). The maxim of the purposiveness internal to an organism is subsidiary to the transcendental maxims Kant refers to in both Introductions: the former is guided by the normative procedure of investigation (‘experience… that is methodically undertaken and is called observation’) established by the transcendental *a priori* principle of the purposiveness of nature. Thus the specificity of the principle of internal purposiveness is derived from experience, while also being grounded in a more basic *a priori* principle that imbues it with normative force (5:376).

In summary, Kant argues that teleological judgment must follow a given normative procedure: presupposing a transcendental *a priori* maxim as a heuristic, proceeding in the course of empirical investigation on the basis of this maxim, and deriving subsidiary empirical maxims on the basis of the particulars encountered in experience (see Fig. 1). The empirical maxims are imbued with normativity, and are therefore not merely descriptive, due to their grounding in the *a priori* maxim that was initially presupposed. Meanwhile, the *a priori* maxim (of natural purposiveness, or what nowadays we might think of in terms of *functional explanation* in science), initially just a heuristic, is vindicated by its confirmation in the course of experience. *Vindicated,* but also *given form* by it: the stipulation of the *a priori* maxim leads to the construction of new, empirical maxims that further specify the content of the initial principle. ‘Nature does nothing in vain’, because, for example, natural organisms show up to us as perfect, as internally purposive, with every part serving a necessary function relative to the whole. Evidence for the initial maxim, then, is provided by reflecting on a particular feature (in this case, a natural organism), which then gives rise to new, empirically derived maxims, and in turn, these maxims give new form to

the initial, guiding principle.[85]

Figure 1: Transcendental and empirical maxims produced by teleological judgment

Type	Transcendental maxim *(heuristic)*	Empirical maxim *(only after encountering relevant particular)*	Vindicated maxim
Command	*Natural purposiveness*: Judge nature as a purposive system; 'nature does nothing in vain'	*Internal purposiveness*: Judge organisms as natural purposes: as purposively ordered	*External purposiveness*: judge natural purposes as themselves purposively ordered one to each other; as related by means-ends chains
Content	General	Specific	Generality imbued with specificity

For example, Kant claims that once, and *only* once, we have: a) stipulated the heuristic principle of natural purposiveness, b) considered particular features of beings we come across in nature (such as a bird's wing, or the structure of its bones, 5:360), and c) formulated the empirical principle of internal purposiveness, are we licensed to conclude that nature is *externally* purposive, or organized as a system of natural links of means-ends chains, with each natural being serving as an end for another. Thus, we are able to reason that "the grass is necessary for the livestock, just as the latter is necessary to the human being as the means for his existence", together comprising a chain of natural links—the grass, the livestock, the human being—each of which serve some other link higher up (5:378). The heuristic principle of natural purposiveness, once justified by

[85] At least when it comes to the particular teleological principles Kant presents in KU; nothing excludes the possibility that heuristic principles might be proposed that are instead *disconfirmed* by investigation.

the discovery of internally purposive particulars in nature, now has a newly specified form: these particulars—individual natural organisms—should be judged as *themselves* purposive to one another, and thus as together comprising a purposive *system* of nature. The vindicated principle retains the level of generality of its initial formulation—we here endeavor to explain the workings of *all* of nature, not just one particular feature of it—but imbues it with specificity, since, at the same time, speaking of external purposiveness still requires us to retain a degree of reflection on particular features *within* nature.

In Kant's account of the procedure of teleological judgment, there is therefore a two-way dependence between universals and particulars. Attributing salience to certain particulars rather than others is guided by *a priori* principles or maxims, which are themselves reciprocally informed by the particulars one encounters. The empirical maxims that result from this process are at once *vindications* of the initial principle and determine it with specific content.

How does the example of teleological judgment relate to practical reasoning? Unlike 'merely reflective' teleological judgments, the reflective dimension in practical deliberation is guided by an objective principle, the moral law, rather than proceeding absent any conceptual determination. But practical judgments are like teleological judgments in that they, too, must be informed by experience in order to render the objective principle more determinate, namely by generating subjectively given maxims to be employed in action. These subordinate maxims of practical reason, however, remain defeasible; upon finding new particulars, they may require revision, or may be articulated in terms of newly available concepts. Moreover, as these 'empirical' maxims change, agents' sense of the guiding universal—the moral law—and what follows from it may also change.

In so doing, practical judgments exhibit the same two-way dependence between

particulars and universals as teleological judgments, one that can be attributed to the fact that practical judgments have a reflective dimension.

The difference between 'mere' reflective judgments and the reflective dimension of theoretical and practical judgment

When it comes to theoretical cognition (e.g., in making claims to knowledge), the interdependence of reflective and determining judgment has been increasingly acknowledged. Guyer (1990) raises the example of a universal, such as causation, that can only be applied to a particular "through an intermediate causal concept", such as that of a "particular kind of chemical or mechanical causation" (18). In this case, both reflective and determining judgment will be required to apply the concept of causation to the particular, since doing so will first require investigation into whether the attribution of the specific causal concept is warranted. Longuenesse (2001) suggests that *all* judgments are reflective, while only aesthetic and teleological judgments are *merely* reflective, since the application of the categories presupposes "progress from sensible representations to discursive thought: the formation of concepts through comparison/reflection/abstraction, which is just what reflective judgment is: finding the universal for the particular" (165). Ginsborg (2015), as I gestured to above, argues that determining judgments are parasitic on the reflexive normativity of reflective judgment, and thus retain a reflective dimension even in straightforward cases of subsumption. These scholars are joined by others who have argued for the complementarity of reflective and determining judgment in theoretical cognition, including Bell (1987), Floyd (1998), and Allison (2001: 20).

By contrast, when it comes to Kant's account of practical cognition, either the relation between reflective and practical judgment has been overlooked, or the two types of judgment

have been taken to be incompatible.[86] In particular, reflective judgment, as Kant characterizes it in KU, is generally interpreted as contrasting with practical reason rather than as constituting an extension of it. In the First Introduction, Kant distinguishes "the power of judgment" from the understanding, which yields "laws of nature", and reason, which yields "the law of freedom" (20:202). Instead, the role of the power of judgment is to "mediate the connection between the two faculties" (20:202).

Here, the comparison to theoretical cognition, where the complementarity of determining and reflective judgment has been increasingly acknowledged, can prove instructive. Indeed, Kant is careful to characterize the aesthetic and teleological judgments that comprised the focus of the third *Critique* as "*merely* [*bloß, nur*] reflective judgments."[87] Thus, both determining and reflective judgments have a reflective aspect, but only reflective judgments are *just* reflective: "The peculiar feature of aesthetic and teleological judgments is not that they are reflective judgments *(for every judgment on empirical objects as such is reflective);* it is rather that they are *merely* reflective judgments, judgments in which reflection can never arrive at conceptual *determination*" (Longuenesse 2001: 163). All judgments, on this interpretation, comprise a *reflective dimension*, insofar as they presuppose the functions of comparison, reflection, and abstraction Kant describes

[86] See, e.g., Herman (1993), Korsgaard (2009), O'Neill (2002, 2018: 132), Papish (2018), Kleingeld (2017). For favorable assessments of the relation of reflective to moral judgment, see Recki (2001), Makkreel (2002), and Klemme (2012), although none of these authors advances a systematic account of the connection. Korsgaard (2009) seems to invoke something like reflective judgment in her reliance on teleology; Merritt (2018) puts emphasis on the importance of Kant's ideal of reflection for virtue. But neither of these accounts makes their reliance on reflective judgment explicit, nor do they elucidate what the connection between Kantian reflective and practical judgment might be: Korsgaard relies heavily on Aristotle and Plato when she invokes the importance of teleological reasoning in our self-conceptions as practical agents (see, e.g., 2009: 37-41); Merritt holds that reflection and reflective judgment, despite the fact that Kant tends to use the two terms interchangeably in KU, are distinct notions for Kant, and that reflective judgment is therefore not relevant to her account of virtue (2018: 16f).

[87] See 20:220-221, 20:408-409, 20:223-224, 20:223-224. Cited in Longuenesse (2001): 163-164.

as necessary "subjective conditions under which we can arrive at concepts" in the Amphiboly chapter of KrV (A260/B316).[88] But only 'reflective' judgments are *purely* reflective in that they lack the possibility of ultimately subsuming their objects under concepts while nevertheless retaining these reflective functions. By differentiating 'merely' or 'purely' reflective judgments from judgments that combine reflective and determining (or reflective and practical) aspects, we can account for the *reflective dimension* of practical judgment—much like the reflective dimension of theoretical judgment.

On this reading, aesthetic and teleological judgments, as 'merely' or 'only' reflective, are paradigm cases for studying the general function of reflective judgment, as both lack a corresponding theoretical or moral dimension. It is for this reason, in my view, that Kant divides his study of reflective judgment, the *Critique of Judgment,* into two parts on teleology and aesthetics. But the broader role Kant accords to reflectivity in cognition explains why Kant defines reflective judgment in both Introductions not narrowly in terms of aesthetic or teleological judgment, but more broadly, as what is generally required to form new concepts or principles (20:211; 5:179).

Technical practical principles and the technical activity of the power of judgment

In particular, reflective judgment's role in forming new principles includes principles of *action.* In both Introductions, Kant spends considerable time not merely on explaining the nature and function of aesthetic or teleological judgments, but on characterizing reflective judgment as

[88] There, Kant writes that "before all objective judgments we compare the concepts" implicated in the judgment with one another, and thus that "*transcendental reflection*… contains the basis for the possibility of objective comparison of representations with one another" (A262/B319, A262/B317). See also Makkreel's comparison of the orientational function of 'reflection' in KrV with that of 'reflective judgment' in KU (2015: 61-66).

"technical", or involving our general capacity to act on purposes—what Kant also terms "skill" or "art" (20:200; 5:172). Technical "propositions of practice" actually bring about their desired objects in the world, and are thus aligned with practical causality: "Where rules are practical… they can also be *technically practical*" (5:175). But the power of judgment in its 'merely' reflective use (*bloß reflektierend*) is also 'technical', Kant says, in that it judges nature in *analogy* with human technical activity, or 'art':

> We shall in the future also use the expression 'technique' where objects of nature are sometimes *merely* judged as if their possibility were grounded in art, in which cases the judgments are neither theoretical nor practical… since they do not determine anything about the constitution of the object nor the way in which to produce it. (20:200-1; my emphasis).

The pure employment of the power of judgment ('mere' reflective judgment) enables us to isolate the *form* of technical judging activity, since no object is produced: "The reflecting power of judgment thus proceeds with given appearances, in order to bring them under empirical concepts of determinate natural things, not schematically, but *technically*, not as it were merely mechanically, like an instrument, but *artistically* [or *artificially*; *künstlich*]" (20:214).

'Mere' reflective judgments are therefore technical without the practice, without the activity; they generally arise in the subject's stance as spectator of natural and aesthetic objects, rather than as actor. Technical propositions, in contrast, encompass *both* reflective and practical dimensions, as Kant suggests by devoting several sections of both Introductions to analyzing them. Indeed, any practical principle that requires empirical knowledge or incorporates empirical content is "technical" (20:200; 5:172). Empirical practical principles can (when employed, for example, in moral action) be both practical in the sense of technical, which is to say, reflective, as well as practical in the sense of moral.

Kant emphasizes this point by again adding the qualifiers '*bloß*' and '*nur*', this time to the practical rather than the reflective, to define the scope of moral legislation: "Legislation through the concept of freedom takes place through reason, and is merely [*bloß*] practical. Only in the practical alone [*nur allein im Praktischen*] can reason be legislative…. Conversely, however, where rules are practical, reason is not on that account immediately legislative, since they can also be technically practical" (5:175). Moral reason is legislative when it comes to the *moral dimension* of practical deliberation, whereas reflective judgment is implicated when it comes to the technical, empirical material of practical deliberation—the *reflective dimension*.[89]

In articulating what the technical consists in, Kant lists some of these empirical practical principles as "statesmanship and political economy, rules of good housekeeping as well as those of etiquette, precepts for good health and diet, of the soul as well as of the body", all of which have to do with the application of the principle of freedom to "the nature of things" rather than practical philosophy in its own right (20:195-6). In terms of his position as articulated in 1789-1790 when both Introductions were written, Kant therefore can sound somewhat dismissive about the role of empirical reflection in a theory of practical deliberation. He claims that "the art of bringing about that which one wishes should exist", insofar as it involves application to things as they are, "in the case of a complete theory is always a mere consequence and not a self-subsistent part of any kind of instruction" (20:200). Indeed, while in GMS Kant situated technically practical rules as part of 'general practical philosophy' (4:412-20), in KU Kant tells us that technically practical rules are instead to be situated on the side of "theoretical" rather than

[89] When I refer to 'practical deliberation', then, I mean the holistic domain of intentionality and action choice, encompassing both pragmatic (means-ends) and moral aspects of reasoning; for more on the practical/moral distinction, see Chapter 2.

"practical philosophy."[90]

Kant's definition of the 'technical' is therefore sometimes understood in terms of means-ends or instrumental reasoning and thus as opposed to practical reasoning; Zammito (1992: 132-3) even associates it with Weber's instrumental rationality in contrast to value rationality. Yet it makes little sense to dismiss the Kantian role for the empirical in practical deliberation outright. 'Technical' principles of statesmanship and political economy are themselves morally neutral; but when morally motivated, they can help to further the achievement of moral goods such as individual self-determination or self-realization.

Indeed, these passages of KU suggest an important ambiguity between the technical domain in the sense of means-ends or instrumental reasoning, which Kant dismisses as philosophically or morally irrelevant, and the technical in the sense of the artistic or creative, which constitutes one of the central topics of the *Critique of Judgment*. Understood in the latter sense, Kant would then be aligning the empirical matter of practical deliberation with the 'technical', 'artistic' subjective normativity of reflective judgment.[91]

Another way, then, to understand Kant's rejection of the prospect of a 'general practical philosophy' is not that Kant dismisses a role for empirical principles from practical reasoning altogether, but that the discovery of reflective judgment afforded him the theoretical latitude to accommodate them along the reflective dimension of practical reasoning rather than as the province of moral philosophy strictly speaking. As we will see in Chapter 4, while Kant remains somewhat ambivalent—when not outright dismissive—towards this prospect in the third *Critique*, he seems to come to more explicitly embrace it in the texts written after 1790.

---

[90] 5:172; see discussion in Papish (2018).

[91] See my discussion of the 'technical', 'artistic' self-given procedure of reflective judgment in §4.3.

Nevertheless, already in the third *Critique*, distinguishing between reflective and moral dimensions of practical deliberation allows us to make better sense of why Kant tells us that the power of judgment "can occasionally be annexed to either of them," to theoretical or practical philosophy, "in case of need", constituting an "intermediary between" them (5:168). In so doing, the activity of the power of judgment renders conceivable how "the concept of freedom should make the end that is imposed by its laws real in the sensible world", thereby securing a "transition from the manner of thinking in accordance with the principles of the [theoretical] to that in accordance with the principles of the [practical]", bridging a "gap" or "chasm" between the two (5:176; 20:244; 5:195). Given Kant's clear preoccupation with the *mediating* function of reflective judgment between theoretical and practical cognition, we should be wary to dissociate reflective from practical judgment. Instead, the function of both modes of judgment is helpfully elucidated by considering of them in relation to each other.

Natural teleology in deriving FRE

I have so far examined the argument from Kant's moral philosophy to reflective judgment, in particular when it comes to 'technical practical' reasoning and maxim construction. But my claim can also be approached from the other direction: Kant also argues from teleological judgment to *moral* reason in deriving one of the formulations of the categorical imperative. Consequently, the relation between the two can be conceived as one of reciprocal directionality of influence.[92]

In GMS, Kant's derivation of the formula of the realm of ends, FRE,[93] directly results

[92] For a discussion of these passages which resonates with my reading, see Klemme (2012).
[93] The maxim is given several paragraphs later: "Act according to the maxims of a member universally legislating for a merely possible kingdom of ends" (4:439)

from a comparison to natural teleology, later accorded to reflective judgment in KU. Kant claims, "All maxims from one's own legislation ought to harmonize into a possible kingdom of ends as a kingdom of nature" (4:436). In a footnote, Kant reinforces the parallel to natural teleology:

> Teleology considers nature as a kingdom of ends, moral science a possible kingdom of ends as a kingdom of nature. There the kingdom of ends is a theoretical idea for explaining what exists. Here it is a practical idea for the sake of bringing about—in conformity with precisely this idea—what does not exist but can become actual by means of our behavior. (4:436n)

That is, formulating the maxim of the kingdom of ends requires an analogy to natural teleology. It requires, first, the function of reflective judgment, specifically *teleological* judgment, to think the concept of natural purposiveness (as a 'kingdom of nature'. Second, this concept serves as a point of comparison on which the moral maxim, that of the kingdom of ends, depends. Kant reiterates the same point immediately before introducing the formulation of this maxim as the "supreme principle of autonomy" (4:440): "A kingdom of ends is possible only according to the analogy with a kingdom of nature—but the former just according to maxims, i.e. self-imposed rules, the latter just according to laws of externally necessitated efficient causes" (4:438). Since the 'kingdom of nature' as Kant characterizes it in KU[94] requires the thought of nature as an *externally purposive* system, formulating the maxim of autonomy requires thinking of other rational beings as related to each other in comparison to such a system. That is, in the domain of nature, each organism is thought as serving a determinate purpose in a natural chain, as an 'externally necessitated efficient cause'—the grass nourishes the deer, the deer nourishes the wolf, and so on. By contrast, in the domain of practical reason, each rational being must be thought as interrelated with all others *not* to serve a determinate, preestablished purpose for someone else,[95]

[94] See Kant's references to various 'kingdoms of nature' (5:426-7).
[95] As Kant establishes in KU, the human being is the *only* natural entity whose purpose, as a link in the

but in the sense that their end is up to them to determine—but only if this end is one that could be determined, or accepted, by everyone else. If FRE states that I must be able to legislate an end, or purpose, that every other rational being can accept (precisely *because* our purposes, as rational beings, are ones up to us to determine), the analogy to natural external purposiveness brings out the contrast between the two. The ends of a rational being can, for the most part, only be constrained negatively rather than determined positively, since they can be ruled out if they fail to be universalizable, but cannot be imposed from without if they are to count as free.[96] However, Kant seems to think that the principle of the external purposiveness of nature is required to fully make sense of this contrast between rational and natural beings, and thus to vindicate FRE.[97]

3.5 On the role of maxim construction

As the analogy between the kingdom of ends and kingdom of nature brings out, one

---

chain of nature, seems unclear; indeed, it is this lack of necessity of the human subject that leads to the thought of the *moral* subject as the final purpose of nature: "Yet one does not see why it is necessary that human beings exist… thus one does not arrive at any categorical end, but all of this purposive relation rests on a condition that is always to be found further on, and which, as unconditioned, (the existence of a thing as a final end) lies entirely outside of the physical-teleological way of considering the world" (5:378).

[96] As O'Neill (1975) discusses at length, Kant does think that there are certain 'special' ends which are duties and thus positively determined, such as self-perfection and the happiness of others.

[97] Although a systematic account of this is outside of my scope, such an analysis of natural teleology in GMS could be extended further. For example, in Section I of GMS Kant motivates the notion that the "true function" of practical reason is to produce a good will by appealing yet again to natural purposiveness: "In the natural predispositions of an organized being, i.e. one arranged purposively for life, we assume as a principle that no organ will be found in it for any end that is not also the most fitting for it and the most suitable" (4:395). In MS Kant again invokes a comparison between a 'system of freedom' and a 'system of nature': "Hence, philosophy can understand by its practical part (as compared with its theoretical part) no technically practical doctrine but only a morally practical doctrine; and if the proficiency of choice in accordance with laws of freedom, in contrast to laws of nature, is also to be called art here, by this would have to be understood a kind of art that makes possible a *system of freedom* like a *system of nature*, truly a divine art were we in a position also to carry out fully, by means of it, what reason prescribes and to put the Idea of it into effect" (6:217-8; my italics). Many other such examples populate Kant's writings on moral philosophy.

might think that practical reasoning does not need the subjective normativity of reflective judgment: it is already governed by an objective principle, the moral law, which Kant presents as a fact of reason admitting of no proof or deduction (KpV 5:29). So what makes the contribution of reflective judgment relevant to practical reason?

In the discussion of the problem of relevant descriptions above, I emphasized that the moral law, even the categorical imperative, presupposes that particular practical maxims have already been formulated. This poses a problem insofar as the moral conclusions we derive from universalization appear to depend on the way in which any given maxim is described. Kantians have noted many cases of apparent false negatives and false positives of maxim universalization, such as "I will buy clockwork trains, but not sell them" (O'Neill 1975: 163), save "money by shopping in this year's after-Christmas sales for next year's Christmas presents" (Herman 1993: 138, citing Scanlon ms.), or make "a false promise on Tuesday, August 21, to a person named Hildreth Milton Flitcraft" (Wood 1999: 102). While the problem of false negatives and positives seems to more often arise with highly specific maxims, as Wood correctly notes, "agents actually do form intentions at a wide variety of levels of generality and specificity"; thus, the problem is to find a "way of describing our intention at *exactly the right* level(s) of generality" (1999: 103, 104-5). Much stands or falls for the prospects of Kant's moral philosophy on what exactly he meant by 'maxim'.

The problem with these attempts at maxim-tinkering is that in each case, they presuppose independent moral knowledge: we can only know which is a case of a 'false positive' and which a 'false negative' if we *already know*, or think we know, what the right result *should* be, against which the apparently false result of the universalization test is judged. As Herman puts the point, "How would we determine the correct level of description except as the one that

produces the desired moral result?" (1993: 142).

In my view, it is misguided to attempt to solve this problem by defining maxims at exactly the right level of generality. Instead, the role of maxim formulation should be separated from the process of universalization. In fact, when we look at Kant's definition of individual maxims, we see that Kant distinctly separates them from the moral law. A maxim of practical reason, he tells us, is "the subjective principle of willing", while the "objective principle (i.e. the one that would also subjectively serve all rational beings as the practical principle if reason had complete control over the desiderative faculty) is the practical *law*" (4:400). Kant elaborates on the subjective-objective distinction in a later passage:

> A *maxim* is the subjective principle of acting, and must be distinguished from the *objective* principle, namely the practical law. The former contains the practical rule determined by reason conformably with the conditions of the subject (often his ignorance or also his inclinations), and is therefore the principle in accordance with which the subject *acts*; but the law is the objective principle valid for every rational being, and the principle in accordance with which he *ought to act*, i.e. an imperative. (4:421n)

Thus, in moral deliberation, subjective, individually given maxims are subordinate to an objective metamaxim by which all are bound, namely the moral law.[98]

Kleingeld and Willaschek (2019)[99] have recently explained this distinction by arguing for two levels of legislation: the categorical imperative as formal metaprinciple universally legislates for all agents, over all other agents, while substantive moral principles, or maxims, are self-legislated. Maxims, then, are generated not by the categorical imperative itself, but by agents'

[98] This distinction has also been taken up by the literature on hypothetical imperatives (as distinct from categorical imperatives), or the imperative to take the means to our ends; Mark Schroeder, for instance, classifies the hypothetical imperative as a "reflection of the auto" in *autonomous*, where the categorical imperative is a reflection of the "nomous" (2014: 225). My view of the 'auto' is somewhat more expansive than what is expressed by the hypothetical imperative, as I will elaborate further in the next chapter.
[99] See also Reath (2006) and Kain (2004).

subjective choices about what to do. While maxims *are* subject to endorsement or rejection by the categorical imperative test, this procedure can only occur once maxims *have already been formulated* by a given agent, and this on the basis of consideration of particular cases.

For example, Kant's discussion of the imperfect duties of moral self-cultivation in MS does not merely involve the individual judgment calls necessary to apply predetermined maxims. Instead, he argues that practical judgment *generates maxims*:

> Ethics, because of the latitude it allows in its imperfect duties, inevitably 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). (6:411)

That is, Kant claims here that the self-given procedure of judgment is one that *produces maxims,* and this in response to dilemmas that arise in lived moral experience and result from the inherent 'latitude' of practical deliberation. Practical judgment generates *new* maxims on the basis of experience in the course of its self-defined activity of judging.

Two kinds of maxims: a provisional distinction

The analogy to teleological judgment suggests an intrinsic relation between more general and more specific maxims. Yet this is a point that gets to the heart of a longstanding debate in the literature on the nature of Kantian maxims: whether maxims should be construed as more specific or more general, as 'particulars' or universals. Some commentators have argued that Kantian maxims should be taken to refer to specific intentions isolated to a single action (O'Neill 1975, Kitcher 2003, Schwartz 2006, Korsgaard 2009); others that maxims constitute general 'rules of life' (*Lebensregeln*) that must be acquired and cultivated (Bittner 1974, Höffe 1977);

others still that they should be understood as principles of greater or lesser determinacy that can be hierarchically ordered (Allison 1990: 93, Timmermann 2000). Much of this controversy is generated not just by philosophical debates over the nature of practical reason, but also by apparent inconsistencies in Kant's own various characterizations of maxims.[100]

As Allison and Timmermann's proposal of a potential hierarchical ordering of maxims suggests, we need not take Kant's divergent characterizations of maxims to necessarily be inconsistent with one another. As Allison writes, "One might think of maxims… as arranged hierarchically, with the more general embedded in the more specific, like genera in species" (1990: 93). The general and the specific, then, would both have a place in a unified account. Timmermann concurs, proposing we consider maxims in terms of principles of both first-order and higher-order volition, the latter of which may admit of cultivation and acquisition (2000: 43).

Both Allison and Timmermann attribute their hierarchical understanding of maxims to Kant's account of maxims in RGV, where Kant refers to "the ground of all specific maxims that are morally evil, which is itself a maxim", suggesting an intrinsic relation between a higher-order maxim and the lower-order maxims it 'grounds' (6:20). Or again: "Since the adoption [of a moral maxim] is free, its ground (e.g. why I have adopted an evil maxim and not a good one instead) must not be sought in any incentive of nature, but always again in a maxim" (6:21n).

The reflective dimension of practical deliberation can help to explain why Kant came to embrace a more complex view of maxims in the 1790s, one which incorporates both a conception of maxims as specific intentions presupposed in individual actions and a conception of maxims in

[100] See Gressis (2010a, 2010b) for an account of the interpretive problems at stake in these different accounts.

terms of general normative principles. We can call the first kind of maxim an 'intention maxim', or *i-maxim,* and the second a 'principle maxim', or *p-maxim.*

Rather than holding an inconsistent view of maxims that wavers between these definitions, I am suggesting that Kant came to hold a view on which, when morally motivated, the former—the more specific—can be systematically ordered by the latter—the more general. This process of systematization would then be ultimately encompassed, at the broadest level of description, under the 'objective' metamaxim of the categorical imperative.[101] A specific maxim of action, an i-maxim, counts as morally permissible when it is governed by the categorical imperative as a metamaxim (a p-maxim at the most general level of description), and may issue from p-maxims, general normative principles, also so governed. Respect for persons, for example, can shape the construction of a p-maxim, such as 'I ought to help people in need', which guides the formation of specific i-maxims, such as 'I am pulling over to help the person hurt on the side of the road', when faced with the appropriate circumstances. This helps to explain why the categorical imperative as well as normative principles such as 'I ought not lie' are *prescriptive,* while Kant often characterizes 'maxims' as *descriptive*.[102]

Thus, Kantian maxims admit of different scope and generality, matching the phenomenology of agents' self-described intentions: "We want to be able to assess *all* aspects and

[101] In this respect, Kant's account of the hierarchical ordering of maxims bears a close resemblance to Elizabeth Anscombe's theory of intention as action description. On Anscombe's view, the most specific description of an action, such as 'moving his arm', best captures the specificity of the action, while the most general, in her example 'poisoning the inhabitants', best characterizes the global nature of the intention: the intention given in the last term of a series of increasingly general action descriptions "so to speak swallows up all the preceding intentions with which earlier members of the series were done" (46). One could also progress up the chain of action descriptions by posing the question 'why?' at each stage: 'I am moving my arm.'—Why?—'Because I am operating the pump.'—Why?—'Because I am replenishing the water supply.'—Why?—'Because I am poisoning the inhabitants.'
[102] See Gressis (2010a, 2010b) for an account of the interpretive problems at stake in these different accounts.

phases of what we do.… It is one of the merits of Kant's solution… that it does not preclude us from assessing morally either small but intended components of our actions such as firing a gun or large intended sequences of actions such as 'committing murder' or 'betraying the cause'" (O'Neill 1975: 109). But, at the limit, we can distinguish between 'maxims' in the sense of intentions and 'maxims' in the sense of principles while positing an intrinsic relation between them. Indeed, the two-way dependence between particulars and universals holds both at the level of i-maxims and at the level of p-maxims. Given the wide range of descriptive specificity and generality by which agents make sense of what they are doing and why, the i-maxim/p-maxim distinction is advanced here as a merely *provisional* one.

I-maxims

The role of experience is perhaps clearest in the case of i-maxims, which, as we just saw, require responsiveness to a given circumstance. I can only intend to pull over to help someone suffering on the side of the road if I am first able to notice that there *is* someone so situated in my vicinity. And this capacity for responsiveness is itself an indispensable aspect of moral character, as moral salience theorists have emphasized: if I fail to notice the suffering individual on the roadside, this is indeed a moral failing, yet one that has historically been difficult for Kantians to capture. Consider the following situation:

> Marisol is at dinner with her friend, Lamar, who has just been laid off. Marisol spends the whole evening talking excitedly about new career opportunities without noticing Lamar's discomfort.

Marisol's subjective endorsement of her maxim of action ('sharing my recent good luck with my friend') may be morally permissible as far as she can determine it to be through the

categorical imperative procedure, but she has still missed important moral information that, we might want to say, *should* have been included in the circumstances factored into her maxim. For example, Marisol might have noticed that Lamar finds himself in a much more unfortunate situation, such that dwelling at length on her own career news rather than his own feelings of precarity may make him feel slighted. Attentiveness to the situation should have led her to revise her i-maxim as 'bragging' rather than 'sharing my news with my friend.'

As this example shows, the role of the empirical is indispensable to practical deliberation. It is not just technically practical maxims of 'art' and 'skill' that incorporate empirical content, but *all* i-maxims, as some influential formulas of maxims as including a description of the agent or of the circumstances suggest. O'Neill (1975) gives an account of maxims influenced by Anscombe's (1957) characterization of intention. On her view, maxims must have a similar intentional structure, of the form "To - - - if … in order to —", where '- - -' stands for an act description, '…' for an agent description, and '—' for a purpose description (O'Neill 1975: 99-102). Kitcher (2003) glosses O'Neill's view as "in C, I to do A for P", where 'C' stands for a description of the circumstances, 'A' for the action description, and 'P' for the purpose description (219). Kant's own examples of maxims, which vary widely in specificity or generality, often include explicit consideration of the agent's circumstances, such as "from self-love to shorten my life whenever its continuance threatens more evil than it promises pleasure" or "whenever I believe myself short of money, I will borrow money and promise to pay it back though I know this will never be done."[103] We can therefore think of i-maxims as contingent on a reflective dimension of practical reasoning: what Kant terms the 'subjective' aspect of moral deliberation. If the formulation of a

[103] GMS 4:422, both cited in O'Neill (1975): 101.

given maxim is, in part, dependent on one's responsiveness to context, this satisfies one of the guiding motivations for embracing particularism, which holds that moral conclusions are highly reliant on contextual features of a given situation.

On this view, an i-maxim is a particular action description rather than a general normative principle. Nevertheless, i-maxims encompass a normative aspect in that they are generally *reasoned*; they retain in principle a connection to what agents take to be *good* for them to do. As Anscombe claims about the nature of intention, while agents can describe their intentions at varying levels of generality, or emphasizing *this* feature rather than *that* one, the many possible descriptions must each characterize, and be internally connected by, what the agent *wants,* where 'wanting' must pick out the agent's seeing "what he wants under the aspect of some good" (1957: 75). While my i-maxim is, on one level, a mere description of what I am *already* doing, I should, on another level, take this description to be normative *for me*. In the previous chapter, I pointed out that Korsgaard affirms a version of this claim when she said that maxims must be 'intelligible', or responsive to reasons.

Yet, as I also suggested there, there can equally be instances in which I am *unable* to affirm the normativity of what it is I am doing, where some feature of my action or action-context is *unintelligible*, and here I am faced with a problem. Conflicts of this sort can result because i-maxims do not just depend on context, but on our conceptual repertoire and the domain of 'live options' conceived to be available to us—on 'practical intelligibility'. While these features are socially and historically shaped, as 'concepts' or 'action descriptions' they constitute Kantian universals rather than particulars. Our phenomenological sense of what is possible to do—and of what is *right* to do—depends in part on how we describe it. Affluent individuals moving into a cheaper neighborhood can understand themselves to be merely moving to an area

where they can finally afford space for a home office or a baby room. But they can also reinterpret what they are doing as combatting white flight, contributing to urban renewal, gentrifying the area, or displacing the original residents. Whether they understand their behavior to be right or wrong depends on the nature of the description they propose, which in turn depends on the discursive resources they can marshal in doing so. At the same time, new experiences that bring out gaps in our discursive schemes can also lead to the formulation of new action descriptions. The term 'gentrification' was only coined once sweeping processes of displacement in London neighborhoods like Chelsea, Notting Hill, and Paddington were already underway, bringing a previously neutral, overlooked practice newly under moral evaluation (Glass 1964).

Experience shapes the terms of practical deliberation, but moral terms also shape moral experience: a two-way dependence between particulars and universals. Consequently, at any level of specificity or generality, possibilities for description can be outstripped by particulars encountered in experience, and this is, as I showed in Chapter 2, where the limits of practical intelligibility in part manifest themselves.

P-maxims

More controversially, the same holds, I want to now suggest, for p-maxims: at the level of moral principle. This claim cuts against the grain of the dominant understanding of Kantian moral theory as strictly *a priori.* But several recent discussions in Kant scholarship have demonstrated an implicit understanding that empirical reflection can directly inform the scope and nature of the moral principles we hold.

First, it has long been argued that Kant, while originating a political and moral theory of crucial and continuing importance for the development of Western thought, was also one of the

originators of modern 'scientific' racism (Eze 1997, Bernasconi 2001, Mills 2005), and advanced numerous claims regarding the natural inferiority of women (Kleingeld 1993, Schröder 1997). While Kleingeld (1993) argues that Kant's bigotry poses a contradiction for his universalist and egalitarian position, Mills (2005) instead claims that it is more charitable to attribute to him the view that personhood comes in degrees, reducing "the degree of cognitive dissonance involved in his writings" (185). Consequently, we need not treat these apparent glaring oversights in Kant's views as cases of self-deception. Instead, Mills suggests, Kant may have been operating with a different notion of personhood from ours altogether, one on which nonwhites were not full moral persons, but 'subpersons': "What seem to be racist inconsistencies and anomalies in the writings of the classic political philosophers of the modern period would… now turn out to be simple and straightforward implications of racially restricted personhood" (171).

If this hypothesis is correct, many of the moral principles Kant advances have already implicitly undergone revision in ways that are then projected back onto Kant's initial theory. For example, Mills glosses the categorical imperative ("all persons should be treated with respect") under the racially restricted conception of personhood operative in the Enlightenment as: "'Person' is a technical term, a term of art, signifying beings of a certain level of intelligence and capacity for moral maturity, and on this planet, whiteness is a necessary prerequisite for being a person in the full sense" (177). 'All persons should be treated with respect', then, would have had a scope restricted from the current meaning, since nonwhites were systematically excluded from candidacy for the 'level of intelligence' and 'capacity for moral maturity' required for full moral personhood. It is a mistake for us to take Kant to hold our contemporary conception of moral personhood and to only have been self-deceived, or to otherwise have contradicted himself. Instead, it is far more likely that the scope of moral personhood has itself broadened since Kant's

first articulation of it.

The apparently empirical *a posteriori* status of Kant's racist conclusions need not entail that they are not central to his position.[104] Instead, even if we take Kant's racism to be the result of mistaken empirical beliefs, "there is something very strange about dismissing the issue of who gets counted in the moral community as merely a matter of incidental detail" (Mills 2005: 182). A difference in Kant's apparent understanding of the empirical facts entailed that the scope and definition of moral personhood, as he understood it, significantly diverged from our contemporary understandings of it. And the same holds for other central Kantian moral principles, such as cosmopolitanism, egalitarianism, autonomy, and respect.

By extension, then, a change in our empirical understanding—encountering a new particular feature, say—can entail a revision in the status and definition of our moral principles, even if implicit. The wording of the principle of respect did not change, but the tacit definition of 'personhood' underwent revision. Moreover, our moral reflection on the empirical history of colonialism and systematic racial discrimination may lead us to enact further revisions: for example, a race-sensitive rather than color-blind conception of universalization (Mills 2017: 17).

To take other Kantian perspectives, Korsgaard's (2018) recent argument for the ethical treatment of nonhuman animals, despite Kant's explicit rejection of our owing any direct duties to nonrational beings, takes into account the evidence of animal suffering to argue that we *do* in fact owe a certain degree of respect to animals. Other interpreters, from Cohen (1902) to Wood

[104] On this, Mills takes issue both with those who claim that the empirical *a posteriori* status of Kant's racism does not 'infect' his central theory, since it falls short of the synthetic *a priori* status that primarily concerns Kant (Wood 1995: 136-138; Louden 2000: 105, 177; Hill and Boxill 2001), and with those who hold it to be 'transcendentally grounded', and on these grounds indispensable to Kant's position (Eze 1997).

(1999) and Love (2017), have argued that capitalism violates Kantian principles such as autonomy and self-determination on the grounds that, e.g., "the needs of capital come to dictate the needs of society" (Love 2017: 589).

Each of these examples do not merely advance deductive arguments against Kant's own professed positions; instead, they demonstrate a *reflective* approach to general principles. They proceed from consideration of a particular, *prima facie* unjust feature of empirical (social) reality: colonialism, Jim Crow, animal cruelty, the machinations of capital. The basis for claiming injustice, however, is already suggested by the moral principles under interrogation. Even if currently incomplete or insufficient, these principles still play a schematic role in *guiding* the process of interrogation, picking out the morally relevant empirical features that can directly inform this procedure of moral reflection.

We can be led to suspect that there is something wrong with urban displacement by witnessing the suffering of those displaced or the reinforcement of class disparities: violations, say, of the principles of humanity or dignity. Once we discursively package this empirical process as 'gentrification', however—a negative moral evaluation—we come to understand the principles themselves in a new way; their semantic shape has been altered.

A model for this is already provided, in my view, by Kant's account of teleological judgment: the *a priori* principle of natural teleology, too, is initially insufficient or incomplete—here, for claiming objectivity for scientific inquiry—but it helps to pick out salient particulars, such as the internal functions of a given organism, which themselves give new scope to the initial principle.

Consequently, these examples do not stop short at noting unjust empirical features. By arguing for inclusion of animals into the moral community, for socialism, or for a 'color-

conscious' universalism, they then propose *revisions* to traditional principles—just as the empirical features picked out as salient by the principle of natural teleology themselves directly inform this initial principle. In deliberation on both moral and teleological issues, *a priori* principle can be readily informed by new particulars. In the moral case, our application of the categorical imperative (our understanding of its scope, say) can be modified on the basis of reflection on social experience. If we concede that the full articulation of our values may not always be transparent to us, as Kant's own racially restricted conception of personhood may not have been fully transparent to him, reflection on how current unjust social arrangements might be tacitly reflected in our values is itself morally pressing.

3.6 Responding to the particularist

I started this chapter by referring to some of the most influential critiques of Kantian morality, as elaborated in Aristotelian virtue ethics and particularist approaches. Charles Larmore articulates a typical line of objection, criticizing Kant on the grounds that a rule-based moral system is inadequate to capture the full nature of practical deliberation: "Is it always true, however, that the rules have enough content to settle by themselves whether something falls under their concept? Is it always true that judgment has no other task than simply to see that such rules indeed suffice to identify what things the concept may be predicated of? This is so neither in the moral domain nor elsewhere as well" (1981: 278).

In what I have said in this chapter, I have begun to articulate a response to these charges. First, understanding a maxim as a rule to be applied inflexibly to divergent circumstances trades on the two senses of maxim, i-maxim and p-maxim, I delineated above. In the sense of 'i-maxim', a maxim constitutes an *intention* that serves as a precondition for action as such, and which

therefore necessarily takes the particular circumstances of action into account. At this level of specificity, a given maxim can only be instantiated for one particular action at one particular place, moment in time, and agent, and therefore isn't a 'rule' in the sense that detractors like Larmore often understand it to be.

But p-maxims, maxims as general principles to be applied across varying circumstances, are also insufficiently understood as rules to be inflexibly applied. As shown above, moral principles—even Kantian moral principles!—are not just subject to application, but to interrogation and revision. This is a process that requires reflective judgment, including reflection on empirical particulars, and can itself be a site for virtue, a point to which I return in Chapter 5 and the Conclusion.

Consequently, distinguishing between i-maxims and p-maxims allows us to show that, at either limit, practical deliberation involves a two-way dependence between the particularity of experience and the generality of practical or moral description. Moreover, this distinction enables us to more closely examine the interrelation between i-maxims and p-maxims:

1. In a fashion that parallels how, in Kant's account of teleological judgment, the formulation of specific empirical maxims of natural teleology vindicate the general *a priori* maxim that posits natural teleology in the first place, i-maxims can 'vindicate' p-maxims by imbuing them with greater determinacy, showing how they can apply to a wider range of cases. In coming to practice vegetarianism, I affirm the guiding principle of valuing sentient life; in unexpectedly coming to love the sort of person I never thought I would, I vindicate my initial commitment to love (where, in the interim, there was perhaps just a hopelessness, a doubt that such love really existed, anyway).
2. Yet expanding one's repertoire of i-maxims can also *alter* p-maxims, subjecting them to

redescription if they fail to capture the complexity of a given situation. Whether it be vegetarianism or love, adopting my new practice—an i-maxim—alters the scope of my initial commitment. Where I first valued only humanity or narrow rationality, I broaden my moral attention to suffering or biological life; where I first had in view only a restricted picture of who I might come to love, I now find that image changed, newly oriented around *this* individual.

3. And, of course, there are the numerous instances in which i-maxims and p-maxims can come into conflict, where we don't in fact do what we take ourselves to be committed to do. While few of us in contemporary society would willingly task ourselves with killing an animal in the flesh, many of us go on eating animals we had no role in killing all the same; while we may feel the urgency *in principle* of warding off large-scale climate disaster, we often go on driving to work every day (indeed, many of our cities, particularly in the United States, are set up such that doing so is the only practicable option); while we may feel that everyone has the right to safety and shelter, the structure of urban life is one that promotes only callousness to those soliciting our aid on the street. In such examples, the live tension between i- and p-maxims puts pressure on one of the two to give way—though there is no inevitable dissolution of this tension.

In each of these cases, practical judgment is charged not just with applying rules, but with (re)formulating or (re)constructing maxims of intention and moral principle. It must therefore negotiate the two-way dependence between particularity and universality, a task which requires a *reflective* dimension of practical deliberation alongside the dimension of moral legislation. Application, then, is never merely iterative or rote; there is no guarantee that in the course of application the guiding universal will be left unchanged.

In this respect, I take the reflective dimension to better account for the role particularists accord to moral sensitivity. Such strategies often take a construal of virtue in terms of a "perceptual capacity" to be sufficient.[105] But what such views fail to acknowledge is that practical reason does not entail *perception alone*; it also involves an ensuing responsiveness to the ways in which new situations might *transform* one's preexisting moral schemes and practical repertoire.

Practical judgment should therefore not be understood as exclusively rule-governed. This is a conclusion that takes us back to Ginsborg's argument for the reflexive normativity of reflective judgment in forming new empirical concepts, initially posed as a response to Herman's rule-based account of moral salience. What we have since outlined are the beginnings of a *positive* account of what practical judgment must consist in once no longer understood to be exhausted by rules. We have seen that practical judgment must be accorded a reflective dimension by which new universals (maxims, moral concepts) are created and revised on the basis of particular features of experience. As we have begun to see, the function of the reflective dimension will require, first, practice in the course of experience, and second, as Ginsborg notes, a kind of normativity distinct from the notion of moral legislation alone: that is, a kind of virtue.

In Chapter 5, I will return to the provisional distinction between i-maxims and p-maxims I have here advanced, reassessing them in light of Kant's conception of *moral ideas*. But first, I will devote attention in the following chapter to the significant evidence that Kant's position on the empirical-*a priori* and theoretical-practical distinctions changed markedly after 1790. A close examination of the shift undergone in Kant's conception of 'autonomy' in his final decade

[105] "Virtue, in general, is [the] ability to recognize requirements that situations impose on one's behavior. It is a single complex sensitivity of this sort that we are aiming to instil when we inculcate a moral outlook" (McDowell 1998: 51, 53).

demonstrates the importance attributed to the distinctive normativity of reflective judgment, *heautonomy*, in both theoretical and practical reasoning, such that it comes to be encompassed within an expanded notion of 'autonomy'.

Chapter 4: Kant on Autonomy as Self-Making

4.1 The 'application presupposition' in Kantian conceptions of autonomy

Most Kant scholars have taken the Kantian principle of autonomy given in GMS (1785) to be indispensable to Kant's mature moral philosophy (principally MS, 1797):

> The principle of autonomy is thus: not to choose in any other way than that the maxims of one's choice are also comprised as universal law in the same willing. (GMS 4:440)

On the standard view, the principle of autonomy is to be applied to MS (what I will call the 'application presupposition'), despite the fact that 12 years passed between the publication of the two texts, during which Kant's views underwent crucial changes, including the publication of KU.

Recent scholarship has challenged the application presupposition by suggesting that the principle of autonomy "virtually disappears" from Kant's later moral philosophy, a dramatic shift occasioning "no debate" in the literature (Kleingeld 2017: 61). While I agree that Kant's account of autonomy undergoes a change from GMS to MS (and that this development has so far been underappreciated), in this chapter I will argue that Kant *does* continue to invoke autonomy in his mature views—but in the *Opus Postumum* (OP) rather than MS. Autonomy, as it is now presented in OP, is no longer identical to the standpoint in GMS of the ideal moral legislator freely conforming to the categorical imperative; instead, autonomy is now conceived as a unified principle of 'self-making' that straddles both the theoretical and practical spheres. Although it ultimately aims at morality, autonomy as self-making therefore does not take an exclusively moral

form. Moreover, it explicitly pertains to the empirically situated human subject, characterizing her active process of self-constitution as a rational being. Thus, while the principle of autonomy does 'virtually disappear' from MS as we had known it, its *re*appearance in OP has been neglected.

While Kleingeld argues that, despite the change in status of autonomy, Kant's account of morality stays largely the same, I attribute the shift to a larger alteration in the conception of Kant's philosophical system. In fact, Kant explicitly notes the influence of OP in the text of the *Doctrine of Virtue* (TL), in a passage (to which I return below) in which he acknowledges the need for a parallel 'transition' in both texts between pure principles (whether metaphysical or practical) and empirical cases.[106] In OP, this transition is secured by Kant's new doctrine of 'self-making', which he terms a form of autonomy; in MS, this transition is secured by practical judgment or casuistry, which applies pure moral principles to contingent circumstances. I argue that the conception of autonomy we get from OP—despite some important differences—is still relevant to the account of practical judgment Kant stresses throughout TL. As OP demonstrates, the new need for a schematism between pure moral maxim and contingencies in application leads Kant not just to *supplement* his initial conception of autonomy with accounts of moral training or moral character, as Kantians have often interpreted Kant's mature position,[107] but instead to expand the notion of autonomy to incorporate what Kant refers to as 'heautonomy' in KU, or the self-referential normativity of judgment: the *reflective* dimension of practical deliberation.

[106] Thorndike (2018) is, to my knowledge, the only commentator who has acknowledged Kant's own stated affinity between the two works.
[107] See, e.g., Herman (2007): 133, Hill (1991): 44.

Given that existing accounts of OP focus almost exclusively on Kant's post-critical theoretical philosophy,[108] the notion that this work could also help to elucidate Kant's post-critical practical philosophy has, to my knowledge, remained largely unaddressed.[109] Yet devoting attention to the latter is of direct consequence to current debates on autonomy in political and moral philosophy, where the Kantian conception of autonomy has often been set in opposition to contemporary accounts. Joseph Raz, for example, takes pains to distance his conception of personal autonomy as "self-creation" from the "only very indirectly related" Kantian conception, since for Kant "authorship reduced itself to a vanishing point as it allowed only one set of principles which people can rationally legislate and they are the same for all" (1986: 370). David Johnston concurs that the appropriate conception of autonomy for the purposes of political theory is based on an ideal of "self-authorship" or "self-fashioning", which he contrasts with Kantian moral autonomy (1994: 75).[110]

Christine Korsgaard has attempted to bridge the gap between these two conceptions by advancing a Kantian account of self-constitution. Her account makes space for individual variation and contingency by attributing it to differences in the *practical identities* individuals adhere to (such as 'mother' or 'student'). A conception of one's 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 1996: 101; see also 2009). However, Korsgaard does not claim to derive this account from Kant; she treats it as needed to supplement

[108] See, e.g., Friedman (1992), Förster (1989, 1993, 2000, 2012).
[109] For a recent exception, see Thorndike (2018), who—despite advancing an otherwise insightful reading of the connections between OP and Kant's practical philosophy—refrains from noting the new sense given to *autonomy* in OP.
[110] For a contrasting view that rejects such personal autonomy theories from a Kantian perspective, see Flikshuh (2013).

the Kantian account of moral autonomy with a source of the *contingent normativity* that structures everyday life (see §2.3).

The alterations undergone by Kant's conception of autonomy in his final decade may show a way of reframing the presumed opposition between 'Kantian autonomy' and the contingent normativity of self-constitution or self-fashioning. First, in OP Kant suggests that autonomy can serve as a more general normative criterion than the employment of the categorical imperative; it also characterizes the subject's 'self-making' of her own judgment in thinking experience in accordance with norm-guided rational and moral ideas. Second, this development in Kant's account of autonomy helps to bridge the gap between the criterion of universalization as such and particularities of practical reasoning, including individual variance in social station, as well as how one opts to structure one's own selfhood and course of life, which cannot be fully addressed by a universalization procedure. Moreover, it does so not through a commitment to the notion of *identity*, which fails to allow for the questioning of the identities we inherit and that can fail to fit, but by appeal to Kant's notion of *reflectivity,* which allows both for the creation of new maxims or principles and the critique of preexisting ones.

This chapter proceeds as follows: first, I present my reading of Kant's account of autonomy as self-making in OP, situating this account within his general aims in this work. Then, I present grounds for why autonomy as self-making can be seen to constitute a significant shift from Kant's better-known conception of autonomy as self-legislating the moral law. I draw attention to Kant's claim in TL that *both* a metaphysics of morals *and* a transition from metaphysics to natural science (the stated aim of OP) require a schematism. If this is the case, then one proposal I advance here for the role played by Kant's alternate conception of autonomy in OP is that it incorporates Kant's conception of the autonomy (or self-reflexive normativity) *of*

judgment—a notion christened 'heautonomy' in the *Critique of Judgment*.

4.2 Autonomy in the *Opus Postumum*

Self-making as a transition from metaphysics to physics

As Förster (1989) and Friedman (1992) have argued, Kant's central problem in OP[111] is that of securing a transition between the metaphysical foundations of natural science (already defended in Kant's 1786 text of this title) and physics. Kant conceived of this task as fulfilling a central function for his critical system as a whole: in letters from 1797 (the year of publication of MS) and 1798, Kant described the prospect of such a transition as the "final goal" that would pay "the unpaid bill of my uncompleted philosophy", filling "a gap that now stands open" (12:222, 12:257, 12:258).

In order to secure this transition, Kant argues that we can examine our own experience as human subjects: simultaneously organic, moving bodies subject to the same laws of physics as other natural objects, and thinking beings who *create* the form of experience in which such laws are cognizable. If "experience must be made" (22:322), and "we have insight into nothing except what we can make ourselves" (22:353), in order to make sense of the transition between metaphysics to physics, "first… we must make ourselves": that is, first we must generate our own self-consciousness, or experience of ourselves as both subjects and objects of thought. Through the "act of composition [or synthesis] (*synthetice*)" of the manifold of self-consciousness, "the subject makes itself, according to a principle, into an object as it appears to itself" (22:358). In doing so, it evidences the fact that physical laws are ones to which we are subject in order for

[111] Devoting more attention to this point is outside the scope of my paper; for more detailed accounts of Kant's aims in OP, see Förster (1987, 1989, 1990, 1993) and Friedman (1992).

cognition to be possible: "The subject in appearance, which collects the inner moving forces for possible experience (for the completeness of possible perceptions) in conformity with a formal law… affects itself according to a principle, hence appears to itself as compositive (by inner moving forces)" (22:359). On the other hand, however, these laws are ones that we ourselves create in the act of cognizing, since "one is oneself the originator of one's own power of thought" (22:79).

It has been established that this position constitutes a departure from Kant's critical theoretical philosophy.[112] As these commentators note, Kant's post-critical views go in a much more substantially idealist direction—one comparable to that of his German Idealist interlocutors, such as J.S. Beck, Fichte, and Reinhold.[113] Like these figures, Kant, too, develops a *Selbstsetzungslehre* (theory of self-positing or self-composition) in OP. Indeed, Förster argues that Kant's *Selbstsetzungslehre* is not influenced by such figures (particularly Fichte), but in fact precedes their own *Selbstsetzungslehren* (1989: 218). The OP, then, presents a 'common root' uniting the two heterogeneous faculties, one absent during the critical period: the activity of self-positing.[114]

As the OP progresses, the principle of self-positing (*selbstsetzen*) or self-affection (*sich*

[112] See, e.g., Förster (1990), Wallner (1984), di Giovanni (1985); Emundts (2004) instead argues that Kant's project in OP is an 'extension' rather than a transformation of the critical project, "since such a plan [namely, the transcendental foundations of empirical science] must be open to revision and alteration" [*Gleichwohl muß ein solcher Plan offen für Revisionen und Umgestaltungen sein*]" (Emundts 2004: 203). I am sympathetic to this proposal, which resonates with my own account of *moral* revisabillity. As far as I can tell, nothing substantively hinges on this point of disagreement among commentators for my purposes here.
[113] See also Kant's 1797 and 1798 letters to Tieftrunk (12:207 and 12:241), Garve (12:255), and Kiesewetter (12:258), in which Kant attacks these figures' popular new extensions of his critical project while jealously requesting news of how these new extensions are being received.
[114] See references to self-positing [*selbstsetzen*], or 'composition', throughout the entirety of OP, e.g. 22:326, 22:358, 22:384, 22:12.

afficiren) begins to take the form of self-*making* (*sich machen*). It is the notion of self-making, developed on the basis of the possibility of self-affection, on which Kant lays emphasis in order to establish continuity between the theoretical and moral domains. In the realm of theoretical cognition, Kant argues that the fact "that there is something outside me is my own product"—that is, that experience is something that must be produced by me, the cognitive subject (22:82). By extension, however, *I* am my own product, since I must also produce my experience *of myself*: "I am an object of myself and of my representations… I make myself" (22:82). As a result, both internal and external experience, experience both of myself and the world outside myself (and thus 'everything'), is a product of my own making: "We make everything ourselves" (22:82).

The notion of self-making is then extended to the moral domain: "The subject determines itself (1) by technical-practical reason, (2) by moral-practical reason, and is itself an object of both" (22:53, 213). Indeed, Kant posits a causal connection between one's recognition of one's own powers of self-affection and one's recognition of one's own status as a *moral* being:

> According to transcendental idealism, the subject constitutes itself *a priori* into an object—not as given in appearance… but as a being who is founder and originator of his own self, by the quality of personality: the '*I am*.' As a man, I am a sense-object in space and time and, at the same time, an object of the understanding to myself. *I am a person;* consequently, a moral *being who has rights*. (21:14, 221).

The notion of being one's own originator or maker grounds Kant's characterization of 'transcendental philosophy': "Transcendental philosophy is the act of consciousness whereby the subject becomes the originator [*Urheber*] of itself and, thereby, also of the whole object of technical-practical and moral-practical reason in one system" (21:78, 245). Being one's own maker establishes the basis for unifying theoretical cognition of natural mechanism (which Kant equates with the 'technical-practical' in 22:52) and practical cognition of human freedom under a

common principle. This 'progression' from one to the other, from natural determination to freedom, is nominally a consequence of the OP's stated goal: establishing a progression from metaphysical principles of natural science to physics. Over the course of the work, however, it seems to constitute an ulterior goal altogether: "One must progress from subjective principles of appearance to what is objective in experience. One must progress from technical-practical to moral-practical reason, and from the subject as a natural being to the subject as a person" (22:49, 210).

Autonomy and reflectivity in self-positing ideas and maxims of reason

This ability to generate ('originate', 'give rise to', *urheben*) one's own thought, and thus consciousness of one's own moral personhood, gives autonomy a new meaning: "Transcendental philosophy is autonomy, that is, a reason that determinately delineates its synthetic principles, scope, and limits, in a complete system" (21:59, 244). The concept of autonomy here, however, is broader than Kantian autonomy traditionally construed; here, autonomy characterizes Kant's conception of transcendental philosophy *in general* rather than a specifically moral principle. In fact, 'autonomy' now serves as the criterion governing the success of Kant's entire philosophical enterprise:[115] "Transcendental philosophy commences from the metaphysical foundations of natural science, and contains the *a priori* principles of the latter's transition to physics… without turning into heteronomy, it then progresses to physics" (21:59, 244).

[115] One might situate this view not as a departure from Kant's critical position, but as a development of it. O'Neill (1990) has written convincingly on Kant's use of political metaphor in KrV, thus situating the project of critique as such as a distinctly *political* task. Moreover, in crucial passages of GMS, KrV, and KpV, Kant presents practical reason, and therefore autonomy, as also being evidenced by the cognitive subject's activity of self-consciousness (A546/B574, 4:457). However, as this chapter aims to bring out, it is not until the post-critical period that this conception is developed into a broader account of autonomy than strictly *moral* autonomy.

This conception of autonomy cannot be dismissed as a mere turn of phrase; Kant reiterates it throughout the last fascicles of the manuscript,[116] particularly in describing the project of transcendental philosophy as such:

> Transcendental philosophy is the autonomy of ideas. (21:79)
>
> Transcendental philosophy is the subjective principle of ideas of objects of pure reason constituting themselves into a system, and of its autonomy according to the concepts: *ens summum, summa intelligentia, summum bonum*. (21:79, 246)
>
> This scale of ideas…. is *autonomia rationis purae.* (21:82, 248)
>
> The autonomy of ideas: to found experience as unity, *a priori*—not *from* experience, but *for* experience. (21:92, 253)
>
> *Ideas* are not mere concepts but laws of thought which the subject prescribes to itself. *Autonomy*. (21:93)

In these passages, autonomy is associated with the production of *ideas*. Such a connection may seem strange: they associate autonomy with *thinking* ideas rather than *acting* in accordance with them. But here it may help to reflect on what ideas are. In KU, Kant characterizes them as follows: "Ideas, in the broadest sense, are presentations referred to an object according to a *certain principle* (subjective or objective) but are such that they can still never become cognition of an object" (5:342). That is, ideas are inherently norm-guided, insofar as "they are produced according to principles" (5:342). In both KrV and KU, Kant refers repeatedly to the 'maxims' that guide ideas of reason. In KrV, Kant writes:

> I call *maxims* of reason all subjective principles that are obtained not from the character of the object, but from reason's interest concerning a certain possible perfection of the cognition of this object…. The systematic unity of the manifold knowledge of the understanding, as prescribed by reason, is a *logical* principle. Its function is to assist the understanding by means of ideas in those cases in which the understanding cannot by itself establish rules.

[116] As Förster (1993) notes, the exact ordering of the manuscript is unknown: the pages, as ordered on Kant's desk when he died, were mixed together before they could be transcribed, and the manuscript suffered decades of neglect. Nevertheless, the precise ordering doesn't much matter for my purposes.

(A648/B676)[117]

Already in KrV, then, had Kant attributed a role to maxims in *theoretical* cognition in allowing cognizers to proceed in accordance with ideas that would otherwise not be open to cognitive representation. For example, Kant names several of the principles, or maxims, of natural teleology on which he focuses in KtU: "That the manifold respects in which individual things differ do not exclude identity of species, that the various species must be regarded merely as different determinations of a few genera, and these, in turn, of still higher genera… is a logical principle… without which there could be no employment of reason" (A651-2/B680-1). Reason requires subjective maxims because, in the absence of determinate rules, it must give itself its own normative course to follow in making sense of experience.

In KU, however, as we saw in the previous chapter, Kant reassigns the employment of such idea-guiding maxims from reason, in its *regulative* use, to judgment, in its *reflective* use. He writes, for example, that the "idea" of a natural purpose "is not a principle of reason for the understanding, but for the power of judgment… where, indeed, the judgment cannot be determining, but merely reflecting" (5:405). The ideas involved in thinking the systematicity of nature are no longer classified as regulative principles of *reason,* but are principles accorded to *judgment,* insofar as the configuration of our cognitive faculties requires them in order to organize experience into ordered knowledge.

Heautonomy

KU also introduces a new conception of autonomy specifically *for judgment*. In both

[117] I am indebted to Anja Jauernig for directing my attention to the Appendix to the Dialectic.

introductions, Kant claims that the power of judgment "exhibits *autonomy*" insofar as it is "legislative with regard to the conditions of its reflection *a priori*" (20:225; compare 5:186). Kant specifies that the maxims that guide judgment are not "psychological" in origin, but normative: "They do not say what happens… but how they *ought* to be judged" (5:182). The normativity at issue makes judgment only *subjectively* autonomous, insofar as it legislates only to itself (rather than to all rational agents, through laws of freedom,[118] or to experience as such, through laws of nature). Thus judgment is not autonomous strictly speaking, but what Kant terms *heautonomous* (where the 'he-' prefix refers to the Greek reflexive pronoun, emphasizing its self-directed nature).[119]

That is, the power of judgment is subject to a self-given normative metaprinciple: in order to proceed in accordance with a given idea, it must methodologically dictate to itself the course of its own procedure, but this procedure can itself only be specified and ascertained in the course of the exercise of judgment itself. Judgment is thereby subject to an internal normative standard: its normativity is both subjective and made in the course of judgment. Calling judgment 'heautonomous', then, is another way of saying that the procedure of judgment is *self-made.*

The upshot of these shifts in Kant's views is that, on Kant's mature formulation, thinking ideas *requires* heautonomy, or the self-making of judgment, rather than reason under its regulative guise alone.[120] Consequently, Kant's characterization of the subjective construction of a system of

---

[118] For a convincing reappraisal of the heautonomy/autonomy distinction, see Kleingeld (2017).

[119] See Floyd (1998): "*Heautonomy* derives from the Greek definite article *he* being attached to the Greek pronoun for 'self', or 'itself': *auto*. The resulting term, *heauto*, means just what *auto* does except that it may only appear grammatically in a sentence reflexively (as in, e.g., 'I wash myself', 'he praised himself', etc.)… Philologically speaking, Kant is trying to emphasize a certain *necessarily* reflective character of the faculty of judgment" (205).

[120] In KU Kant claims that in teleological judgment, "we attempt to ascribe [causality] to nature by analogy with a *subjective principle*, namely that of *art*, i.e., causality in accordance with *ideas*", relating

ideas as a form of 'autonomy', in the sense of self-making, is consistent with his views post-1790. The form of autonomy Kant has in mind, I posit, encompasses heautonomy. Indeed, moral autonomy *already presupposes* heautonomy: moral actualization *also* requires the subjective, self-given generation of maxims. Thus, if Kant's characterization of reflective judgment in KU refers to cases that are *merely* [*bloß, nur*] reflective (rather than reflective *and* determining as in KrV,[121] or reflective *and* practical, as, I will argue, in GMS and KpV), 'autonomy', as the more general concept, will also comprise 'heautonomy'. Kant's frequent references in OP, not only to the making of (theoretical) ideas, but to 'moral-practical reason' and the generation of moral personhood, suggest that he had both conceptions—autonomy and the heautonomy it presupposes—in mind.

The 'self-making' of heautonomy constitutes a through line in the argument of OP, insofar as the conception of autonomy it describes is characterized in terms of an extensive process of self-making with both theoretical and practical dimensions. In theoretical terms, we 'make ourselves' by cognizing ourselves as objects: "Transcendental philosophy is the capacity of the self-determining subject to constitute itself as given in intuition.... *To make oneself*, as it were [*gleichsam sich selbst machen*]" (21:93, 254). In practical terms, we 'make ourselves' moral agents: "It is not even in the divine power to make a morally good man (to make him morally good): He must do it himself" (21:83). In both cases, the common principle being 'made' is personhood, both moral personality and the personhood of the rational cognizer:

> Transcendental philosophy is the doctrine of the complex of ideas, which contain the whole of synthetic a priori knowledge from concepts in a system both of theoretical-speculative and moral-practical reason, under a principle through which the thinking subject constitutes itself

heautonomy at once to 'art' or technical principles and to ideas (5:390).

[121] See discussion in Chapter 3 (§3.4).

in idealism, not as thing but as person, and is itself the originator [*Urheber*] of this system of ideas. (21:91, 252)

Kant's repeated insistence in these passages on the fact that the idealist system comprises *both* 'theoretical-speculative' *and* 'moral-practical' reason, and that the 'constitution' of the subject requires *both* constituting oneself as a person (a moral notion) *and* 'originating' a system of ideas, provides additional evidence that the conception of autonomy at issue must encompass both theoretical *and* practical uses.

While this conception of autonomy is in tension with Kant's characterization of autonomy in GMS and KpV as limited only to the legislation of the *moral law,* Kant's characterization of heautonomy shows the contours of an account that cuts across the theoretical-practical distinction.[122] Kant characterizes the heautonomous principles of judgment in the Second Introduction of KU as "themselves fit neither for theoretical nor for practical use", thus "mak[ing] possible a transition" from one domain to the other (5:176). As we saw in the previous chapter, in the First Introduction, Kant specifies that the use of judgment is 'technical', or practical in the contemporary sense (rather than practical as moral): "The reflecting power of judgment thus proceeds… not schematically, but *technically,* not as it were merely mechanically, like an instrument, but *artfully* [*künstlich*]"[123] (20:213-4). Reflective judgment is practical in the sense that an artist is: proceeding creatively in action and thought with the principles given to

[122] This possibility is already latent in Kant's characterization of ideas of reason in KrV: "Perhaps the transcendental ideas of reason make possible a transition from the concepts of nature to the practical concepts and in this way provide for the moral ideas themselves support and coherence with reason's speculative cognitions" (A329/B386).

[123] This term is often also translated as 'artistic', although 'artificial' is its more prevalent contemporary meaning. Another possible translation is 'manmade', which would emphasize the connection to the self-making of judgment. Kant's main point here is that reflective judgment establishes its own principles rather than having them be externally dictated.

itself.[124]

Moreover, Kant's repeated references throughout OP to 'organic matter', 'organisms', and 'organic bodies' (there are 140 instances of 'organic' alone) bolster the suggestion that reflective (specifically teleological) judgment becomes increasingly central to the expansive conception of autonomy he now endorses. Among these references is also an explicit reference to the 'autonomy' presupposed by our concept of an organism, suggesting that Kant really had broadened his conception of autonomy to include even straightforward instances of heautonomy: "Note. Of the autonomy of the concept of the organization of matter, without which we ourselves would have no organs" (22:86, 191). What Kant has in mind with his conception of 'self-making' also involves this consciousness of ourselves as purposive, corporeal beings: "Even the organism is contained in the consciousness of oneself. The subject makes its own form in accordance with *a priori* purposes" (22:78, 186). We 'make ourselves' not just by constructing systems of ideas (including the idea of one's self as a human being), but as embodied, purposive organisms. The purposive configuration of our bodily being is, in part, up to us (for example, through the development of physical skills and capacities).[125] Both of these notions, however—proceeding in accordance with the maxims necessitated by ideas, as well as conceiving of ourselves as purposive organisms, as I explain below—require the use of reflective judgment.

[124] In this respect, the need for moral schematism accords nicely with what Matherne (2014) emphasizes as the role of agency in empirical schematism, which Kant characterizes as a 'hidden art'.
[125] The connection between self-making and purposiveness suggests a Kantian reference point for a position most commonly ascribed to Aristotle; see Korsgaard (2009) on the Aristotelian elements of her account of self-constitution.

4.3 Self-making and heautonomy in the *Doctrine of Virtue*

Against the application presupposition, Kleingeld rightfully observes that the notion of self-legislation becomes increasingly less relevant for Kant's mature philosophy, since this notion also lacks a central place in the OP. What poses a problem for her account, however, is the fact that Kant nevertheless *retains* the concept of autonomy in OP, even without the analogy to legislation she takes to be definitional. Indeed, autonomy in OP is no longer governed by the metaphor of self-legislation (giving a *law* to oneself), but that of self-making (*making* oneself). Thus, autonomy does not drop out of Kant's mature philosophical position, but it does undergo an important shift.[126] As if to emphasize the distinction between the critical and post-critical conceptions of autonomy, Kant populates the latter notion with a corresponding new constellation of terms: self-constitution [*sich constituieren*], self-creation [*sich schöpfern*], self-origination or self-authorship [*sich urheben*], self-construction [*sich zimmern*]. While the notion of self-legislation is *juridical*, the notion of self-making is more explicitly *generative*. The subject as self-maker is no longer conceived on the model of an ideal moral judge deliberating over an already given law, but instead on that of creating a system of (ultimately self-directed) ideas. Autonomy here has to do with 'founding' oneself rather than legislating oneself, with self-production rather than self-constraint: "According to transcendental idealism, the subject constitutes itself… as a being who is founder [*Begründer*] and originator [*Urheber*] of his own self" (21:14).

[126] While Kleingeld makes no reference to OP, scholarship specifically on OP has also missed this point. In this literature, it is generally accepted that the OP presents a new account of theoretical reason; however, the conception of practical reason Kant advances therein has either been neglected, or read as conforming unproblematically with Kant's critical moral philosophy (see, e.g., Förster 1993). Yet such positing of an equivalence between the two elides the distinctions between the two accounts.

But is this conception of autonomy relevant to Kant's practical philosophy? On the virtually universal application presupposition, the answer would be a resounding 'no'. On this presupposition, the definition of autonomy presented in GMS is simply 'applied' to Kant's later moral writings, even if this conception no longer appears in these writings. Thus, as this consensus among scholars would have it, the conception of moral autonomy in GMS is much more relevant to texts such as MS than the (not specifically moral) conception presented in OP. After all, the title of the former work makes clear that it is to lay the *groundwork* for Kant's metaphysics of morals, whereas many scholars have dismissed OP as the product of Kant's senility.

Against this presumption, there is textual evidence *within MS itself* that Kant took the OP to provide a *model* for a metaphysics of morals. In TL, Kant explicitly draws a parallel between the aim of this work and OP, suggesting that both texts, composed in overlapping time periods in the mid-1790s, should be treated as complementary:

> Just as a passage from the metaphysics of nature to physics is needed—a transition having its own special rules—something similar is rightly required from the metaphysics of morals: a transition which, by applying the pure principles of duty to cases of experience, would *schematize* these principles, as it were, and present them as ready for moral-practical use. (6:468)[127]

[127] Some might take the 'metaphysics of nature' in this passage to refer to KrV rather than MAN (see Friedman 1992: 168); on this reading, the 'passage' from the metaphysics of nature to physics would be MAN rather than the OP. However, in the A-edition of KrV, Kant differentiates the project of this text from a 'metaphysics of nature': "Such a system of pure (speculative) reason I myself hope to provide under the title of *Metaphysics of Nature*. That system, though not half as voluminous as this critique, is to be incomparably richer in content. But first the critique had to establish the sources and conditions of the possibility of that system, and needed to clear and level a ground that was entirely overgrown" (Axxi). In MAN itself, Kant seems to take the subject matter of this work to be the metaphysics of nature (4:470). Yet in KU, Kant reiterates that the metaphysics of nature is a separate project yet to be completed, one which would complement the metaphysics of morals (5:170. Kant therefore seemingly reinforces both the need for a transition and the inadequacy of MAN to fill this role on its own (5:170). It seems clear, then, that the 'passage' from metaphysics of nature to physics cannot refer to MAN (which provides the passage from KrV to physics), but must refer to the transition project of OP. See discussion of the shifting

That is, the aim of OP is to secure a transition from the *a priori* to the empirical in the domain of theoretical philosophy, while MS requires a similar transition in the moral domain.

In noting this important point of commonality between the two works, I do not want to be taken to conflate their differences. Indeed, OP and MS could be seen to operate on two distinct intellectual registers: OP, at least at the outset, seems to be situated transcendentally, proceeding from its 'first principle' of self-positing; MS, if read as proceeding from the transcendental 'foundation' secured by GMS, could be taken to be merely supplementary or additive, and thus not transcendental in its own right. However, the incongruity between them appears less straightforward on closer inspection: OP aims to secure a transition from the transcendental to the empirical, and therefore cannot remain solely transcendental, while MS incorporates both *a priori* metaphysical first principles and a 'practical anthropology'—the latter of which Kant newly acknowledges as a necessary component of his moral system.[128] Thus, while there are important methodological differences between the two works, both aim at achieving *cohesion* between metaphysical principles and the empirical conditions to which they apply. Arguably, this need for both domains to *cohere* is a more pressing concern at this stage than in Kant's earlier works and, as I suggest below, may have been the direct result of the conclusions reached in KU.

The connection Kant draws between the respective functions of OP and MS suggests that the reappearance of autonomy is not attributable merely to the revision of his views on *political* legislation (as Kleingeld 2017 alleges), nor to Kant's senility in OP, but to a more

---

referent of 'metaphysics of nature' in Pluhar (1996: 39-40, note 149), Guyer and Wood (1998: 66), and Friedman (1992: 243, 260). I thank Sophie Cote for pressing me on this point.

[128] I return to the role of anthropology below.

pervasive problem posed by Kant's late practical philosophy.[129] Namely, the notion of self-legislation alone is insufficient to account for the connection between pure principles and contingent circumstances, as instantiated, for example, in the *formulation* of maxims of action, which instead requires an independent contribution by the moral agent. Thus, the need for a transition is felt not only from metaphysics to natural science, but also from pure moral principles to situated practical deliberation.

In OP, it is autonomy, conceived in terms of self-making, that secures the transition from pure principles to empirical experience. In MS, it is judgment, through casuistical applications of pure moral principles to contingent cases, that helps to secure this transition in the moral sphere (see Fig. 2):

Figure 2: The 'transition project' in the metaphysics of natural science and the metaphysics of morals

	Transition from -> to:	Transition from -> to:
Pure	Metaphysical foundations	Universal moral maxims
	Opus Postumum (autonomy as self-making)	**Metaphysics of Morals** (practical judgment)
Empirical	Natural science	Concrete cases

As we saw, however, Kant's characterization of autonomy in OP, as the construction of a system

[129] I don't want to be taken to suggest that this is the *only* problem Kant's new conception of autonomy is intended to solve. The more pervasive problem, I take it, is that of finding a principle to unify the spheres of theoretical and practical cognition. Kant felt this to be a problem well before OP: see Neuhouser (1990: ch. 1) on Kant's search for a deduction of the categorical imperative to establish a structural affinity to the transcendental deduction of the categories. However, it seems to have become more pressing in his final decade; see Kant's correspondence with Maimon (11:285) and Weiss (12:185-6), who press Kant on his 'universally valid principle' and 'actual ground' of his critical philosophy, respectively.

of ideas, already presupposes the autonomy of *judgment,* insofar as the maxims required for the regulative employment of ideas are reflexively given to judgment, constituting its heautonomy or self-referential normativity. If autonomy as self-making in OP also requires the role of judgment, can this connection shed any light on *practical* judgment, the point of similarity Kant emphasizes between MS and OP?

From the denial of a practical schematism to a 'reflective faith'

Kant continues the passage in TL in which he draws this explicit parallel between the 'schematizing' functions of OP and MS by suggesting that, in the moral case, a transition is required in order to answer questions such as:

> How should one behave, for example, toward human beings who are in a state of moral purity or depravity? Toward the cultivated or the crude? Toward the learned or the unschooled, and toward the learned in so far as they use their science as members of polite society or outside society, as specialists in their field (scholars)? Toward those whose learning is pragmatic or those in whom it proceeds more from spirit and taste? How should people be treated in accordance with their differences in rank, age, sex, health, prosperity, and poverty, and so forth? These questions do not yield so many different *kinds* of ethical *obligation* (for there is only *one*, that of virtue as such) but only so many different ways of *applying* it. (6:468-9)

As Kant argues throughout TL, ethics involves 'imperfect duties' that necessitate a certain 'latitude' in how to fulfill them.[130] Thus, as we saw in the previous chapter, ethics "unavoidably leads to questions that call upon judgment to decide how a maxim is to be applied in particular cases" (6:411). That is, to answer questions such as the ones Kant raises in this passage—how to

[130] Though, as my argument in Chapters 2 and 3 suggests, perfect duties are also not reducible to mere *applications* of principle to experience, but are also subject to 'moral intelligibility', or to how they are conceived or described. In the case Anscombe (1956) raises of Truman dropping the atomic bomb on Japan, the apparent moral dilemma arises in part due to collective uncertainty about how wide a scope to attribute to 'killing', as well as the moral conflict at stake in potentially saving millions of lives (though Anscombe would vehemently disagree with both sources of uncertainty). I am indebted to Don Garrett for pressing me for this example.

treat different *kinds* of individuals, in accordance with their social roles, capacities, and contingent characteristics—we can't rely on explicit *rules* directing us what to do in each case, but must use our individual[131] capacity of judgment. This is because the role of practical judgment, if conceived as mechanically applying rules, would give rise to a regress, one akin to the regress Kant notes in his discussion of theoretical judgment in the Analytic of Principles of KrV: the application of each rule would need to be guided by the introduction of a new rule, and so on ad infinitum (A133/B172).[132] The role of judgment therefore does not admit of systematization.[133] Kant concedes that here "ethics falls into a casuistry", which can only "fragmentarily", not systematically, be incorporated into ethics (6:411). Thus, casuistry cannot be conceived as a "science", but only as a "practice", specifically, the practice of moral judgment (6:411). Because matters of judgment, such as how to treat individuals in light of their varied conditions and social roles, therefore resist complete systematization, Kant claims that casuistical applications of ethical obligation, though necessary for a "complete presentation of the [moral] system", can only be "appended" to it fragmentarily (6:469).

This view of practical judgment explicitly contrasts with the position Kant advances in KpV (1788), where he rejects the role of a schematism for practical reason, affirming only the possibility of introducing further rules. In the Typic of Pure Practical Judgment, Kant writes that since judgment applies universal principles to particular cases, and since all particular cases can only be empirical, "it seems absurd to want to find in the sensible world a case… to which there

[131] Or collective—some of these questions are subject to collective deliberation, and indeed, this is my own interest in the other chapters of this dissertation. In either case, the determination of what to do remains perpetually open to negotiation rather than being laid down as a rule once and for all.
[132] Kant also refers to the 'regress' of judgment in TP (8:275) and ApV (7:199).
[133] For this reason, as Kant claims in ApV, "the power of judgment cannot be instructed, but only exercised" (7:199).

could be applied the supersensible idea of the morally good" (5:68). Thus, practical judgment is subject to "special difficulties" that theoretical judgment does not face, since it would seem impossible to apply laws of freedom to events of the sensible world (5:68). Consequently, Kant here denies that practical judgment, unlike theoretical judgment, could constitute a 'schema': "No intuition can be put under the law of freedom… and hence no schema on behalf of its application *in concreto*" (5:69).[134] In doing so, Kant *rejects the very point of connection he draws between OP and MS above*—namely, that *both* theoretical *and* practical reason require a 'schematism' between universal principles and conditions of application, and with it an *irreducible* role for practical judgment.[135] Because, at this earlier stage, Kant holds that practical reason does not admit of a schematism, he admits only the possibility that practical judgment take the form of a *law* (5:69). Thus, he explicitly rules out the characterization of judgment given in KrV and MS (thus both theoretical *and* practical) as a 'practice' which, due to the threat of a regress, precisely *cannot* be formalized as a law.

[134] Merritt (2018) helpfully distinguishes the "capacity to appreciate moral requirement *in concreto*" from the "capacity to apply rules, represented *in abstracto*, to a given situation", where the former consists in "a readiness to see facts about one's situation as themselves requiring certain responses of attention and action" (126). If Merritt is right, Kant would here be denying the role of this subjective moral assessment, but later come to embrace it.

[135] To recall, the complete passage is the following: "Just as a passage from the metaphysics of nature to physics is needed—a transition having its own special rules—something similar is rightly required from the metaphysics of morals: a transition which, by applying the pure principles of duty to cases of experience, would *schematize* these principles, as it were, and present them as ready for moral-practical use" (6:468). In MAN, Kant characterizes the task of a metaphysics of nature as application *in concreto*—which he glosses in terms of the *exemplification*—of transcendental principles: "A separated metaphysics of corporeal nature… furnishes examples (instances *in concreto*) in which to realize the concepts and propositions of the latter (properly speaking, transcendental philosophy), that is, to give a mere form of thought sense and meaning" (4:478). If a metaphysics of nature schematizes—produces examples *in concreto* of—transcendental principles, a metaphysics of morals must also schematize, or produce examples of, moral principles. Yet, by the time of MS, Kant seems to no longer identify the role of MS with this schematizing function. Instead, he calls for a transition *between* metaphysics (whether of morals or of science) and real-world cases, as can be gleaned from the former passage (6:468), thereby aligning the tasks of OP and MS.

How should we make sense of this significant change in Kant's conception of practical judgment? I suggest that the discovery of reflective judgment in KU was more broadly transformative for Kant's mature practical philosophy than has generally been recognized. As Makkreel (2002) and Munzel (1999) have noted, the role of 'reflection' is explicitly mentioned in the *Religion* (RGV, 1794) as what is required to effect a transition to a moral community.[136] Kant refers there to a moral faith which is "*reflective* [*reflectirenden*]" (6:52), insofar as it allows reason to "extend itself to extravagant ideas," which it "cannot incorporate… into its maxims of thought and action" (6:52). Moreover, this 'reflective' mode of faith is required if "something more" from the "inscrutable field of the supernatural" can be brought to the understanding, since "reason even counts on this something being made available to its good will even if uncognized" (6:52).[137] That is, Kant attributes a specifically *moral* role here to *ideas*, which are present to a good will even if they cannot be cognized, or integrated into maxims or (moral) concepts. Ideas—concepts beyond our cognitive grasp, that elide formulation in specific maxims of action—nevertheless have an important role to play in *guiding* our formation of maxims (our 'good will'). Moreover, it is only through *reflective judgment* (a 'reflective faith') that such ideas can make themselves available to us in this morally constructive fashion.

[136] The specific passage they cite is: "The basis for the transition to the new order of things… once grasped after mature reflection [*Überlegung*], will be carried to effect" (6:122). Since Kant here uses 'Überlegung' rather than 'Reflexion' (the term employed in KU to refer to reflective judgment), which is closer to 'consideration' than 'reflection' (in Kant's sense), this passage provides weaker textual support than the passage I cite here.

[137] The full passage is the following: "Reason, conscious of its impotence to satisfy its moral needs, extends itself to extravagant ideas which might make up for this lack, though it is not suited to this enlarged domain. Reason does not contest the possibility or actuality of the objects of these ideas; it just cannot incorporate them into its maxims of thought and action. And if in the inscrutable field of the supernatural there is something more than it can bring to its understanding, which may however even be necessary to make up for its moral impotence, reason even counts on this something being made available to its good will even if uncognized, with a faith which (with respect to the possibility of this something) we might call *reflective*" (6:52)

If Kant affirms the role here of reflectivity in practical reason, this suggests that heautonomy plays an analogous role in moral deliberation to theoretical reason. While Kant does not, to my knowledge, explicitly refer to reflective judgment in MS or elsewhere in RGV, Kant's discovery of reflective judgment would clarify key changes in Kant's account of practical judgment from GMS to MS: the independently necessary role of practical judgment, the necessity that judgment be *practiced* rather than governed by rules, and judgment's necessarily 'fragmentary' status in his philosophical system. In each of these respects, Kant's account of practical judgment comes to much more closely resemble his account of judgment in KrV, suggesting that he comes to reject the disanalogy initially established in KpV between theoretical and practical judgment. With it, he would affirm the possibility of a practical schematism on which the structural affinity between OP and MS depends.

If so, many of Kant's remarks in MS can be taken to refer to the *reflective* dimension of practical reason, in addition to the *moral* dimension as it has been commonly understood. For example, casuistry, insofar as it involves the subsumption of particular cases under universal principles towards the end of forming a complete ethical system, may be taken to constitute a form of reflective judgment. As Kant presents the 'transition' requisite for a metaphysics of morals in this passage, the formulation of specific principles as to how to treat individuals who are 'cultivated or crude', 'morally pure or depraved', or of different social standpoints must not be undertaken *a priori,* but can only result from having compared and considered different particular cases. Thus, the form of judgment necessary to secure the relevant schema accords with Kant's definition of reflective judgment in the First Introduction to KU,[138] Casuistry, if understood as

[138] Where he claims, as we saw in the previous chapter, that "to reflect… is to compare and to hold together given representations either with others or with one's faculty of cognition, in relation to a

the formulation of ethical principles which specify how to proceed in determinate cases, would have to be conceived as a form of reflective judgment, as well as practical judgment.

This outcome can be taken to accord with Kant's shift in emphasis in his mature moral philosophy. By the 1790s, Kant comes to emphasize the cultivation of virtue as what enables human moral actualization over the metaphysical status of morality as such. Thus, as Guyer (2013) has argued, autonomy must be conceived not just as the rational capacity accessible to all moral agents, Kant's focus in GMS and in the 1780s more generally; instead, by the 1790s, Kant stresses that autonomy must also be empirically realized in the lives of given agents—and on this empirical level, one's 'susceptibility' to being moved by duty requires cultivation.

Perhaps, then, Kant's late moral philosophy was transformed in ways the application presupposition has obscured from view. For example, if, as Herman notes, "there is the surprising fact that the CI [categorical imperative] procedure is not used in either part of the *Metaphysics of Morals*", this suggests the possibility that Kant's considered conception of moral philosophy *came to deemphasize* the role of the categorical imperative procedure given pride of place in GMS (1993: 133). This is not to say that the categorical imperative becomes irrelevant; Kant continues to refer to it (though less frequently) in both texts. However, Kant's account of practical deliberation appears increasingly to *broaden beyond* mere applications of rules, including the categorical imperative. The picture we get of practical deliberation in OP and MS seems to put more emphasis on the individual's role in *schematizing* moral ideas (including virtue), requiring their individual practice of judgment.

Yet, if this is the new emphasis of Kant's late moral philosophy, as Zuckert claims, such a

concept thereby made possible" (20:211).

schema would have an altogether different role than an empirical schema. Where an empirical schema functions to secure a fit, to enable application, between cognitive rules and the sensible manifold (between universals and particulars), insofar as ideas *lack* any determinate empirical object, "here the schema must always defer such fit [between universal and particular], obstruct any sense of seamless application" (2017: 101).[139] Emphasizing the role of ideas in moral reasoning therefore comes bound up with epistemic humility, with the recognition that there is necessarily contingency here, since there *is* no direct correspondence between idea and experience. Schematizing moral ideas involves the *employment* of judgment, rather than the *application* of principle, because it calls on one's agency in judging, on one's own subjective contribution.

Schematizing the anthropological and the moral

These considerations suggest a more immediate point of connection is established between Kantian anthropology and Kantian morality. Initially merely a domain of theoretical investigation of the empirical human being (akin to psychology), Kant comes, throughout the 1780s, to characterize it as '*practical* anthropology', and thus as a requisite domain of moral philosophy.[140] In ApH, Kant claims that anthropology, now in its practical or 'pragmatic' sense, concerns "what *he* [the human being] as a free-acting being *makes of himself*, or can and should

[139] Here Zuckert refers to *ideas* as 'schemata of regulative principles' where I refer to schemata *of ideas*; it should therefore be underscored that we are referring to different objects. Ideas, in my view, cannot be understood *as themselves* schemata, since Kant refers, in the several passages I have drawn attention to throughout this chapter, to the need to schematize ideas; in KU Kant again refers to the problem of schematizing ideas, concluding that the only possible schema of an idea is instead a 'symbol', an analogy, rather than the direct representation an empirical schema would afford (5:351-2). See Chapter 5 for my own account of ideas of reason and regulative principles as two sides of the same coin (as it were): where the latter is prescriptive, the former is descriptive, or so I claim, an account which can be extended to moral ideas and moral principles.

[140] See VA Mrongovius (1784-1785) 25:1367 for Kant's first reference to 'practical anthropology'. See also discussion in Wood and Louden (2012), as well as Wood (1991, 1999).

make of himself" (7:119, my emphasis).[141] That is, anthropology, if understood practically, centrally concerns a kind of 'self-making'.[142]

Moreover, anthropology in this sense is continuous with Kant's account of virtue, since it is not just descriptive ('what the human being makes of himself'), but *normative* (what the human being 'can and should make of himself').[143] The aim of anthropology is in part to depict possibilities for the human subject's course of moral actualization. It is both empirical—describing how human beings *do* actualize or 'make' themselves as subjects—and practical—advocating for their *moral* self-actualization or 'self-making'.

In MS, Kant draws another structural parallel between this text and OP. If OP is to provide the "step which connects both banks" (21:403, 15) between empirical natural science and a metaphysics of nature, MS is tasked with integrating empirical science of human nature (anthropology) into a metaphysics of morals:

> Just as there must be principles in a metaphysics of nature for applying those highest universal principles of a nature in general to objects of experience, a metaphysics of morals cannot dispense with principles of application, and we shall often have to take as our object the particular *nature* of human beings, which is cognized only by experience, in order to *show* in it what can be inferred from universal moral principles. (6:217)

If a metaphysics of morals will often involve taking as its object 'the particular nature of human

[141] As distinct from anthropology in its 'physiological' sense, as Kant had first characterized it in his earliest lectures, which "concerns the investigation of what *nature* makes of the human being" (7:119).
[142] For an analysis of ApH's account of 'self-making', see Wood (1999, 2003).
[143] The normative dimension of Kant's considered view of anthropology has often been missed; for a recent example, see Theunissen (2016)'s characterization of anthropology as purely descriptive and consequently "not strictly speaking a branch of moral philosophy at all" (109). As I argue, there is textual evidence against both claims, and thus against Theunissen's conclusion—that a metaphysics of morals cannot contain any empirical considerations—as well. Theunissen's interpretation relies heavily on GMS and KpV, not taking into consideration the changes in Kant's views between the publication of these texts and MS proper I have noted here. For discussion of the moral implications of Kantian anthropology (under its 'pragmatic' guise), see Cohen (2009).

beings', this refers to the domain Kant marks out for *anthropology.* At the same time, Kant increasingly comes to conceive of the domain of anthropology as having a distinctly *practical* component (25:1367, 27:1398). Thus, Kant draws together the two domains such that each comes to include aspects of the other. A metaphysics of morals has the task of schematizing universal *a priori* principles and conditions of application, which in practice will also require schematizing our self-understanding as at once anthropological, corporeal, socially situated beings and moral agents.[144] As Kant emphasizes here, the empirical realization of autonomy depends not just on cultivating one's 'susceptibility' to moral duty (Guyer 2013), but also reflection on how universal *a priori* moral principle is to be actualized in contingent circumstances which, often, are not immediately up to us.

4.4 Reflective judgment and practical reason

Given what I have said so far, the role of reflection in moral deliberation may look like it had no roots in Kant's initial views. Yet, in certain respects, reflective judgment had long been intrinsic to Kant's account of practical reason, putting further pressure on the application presupposition. In the remainder of this chapter, I will take up an example of the role played by natural teleology in Kant's conception of self-perfection as a moral notion as it changes from GMS to MS. Kant's appraisal of self-perfection shows that the discovery of reflective judgment allowed Kant to broaden his conception of practical reason to accommodate the *reflective*

[144] See KU 5:435 for Kant's claim that in order to connect our conception of nature to the human being as a moral subject, we require the thought of the 'culture' of the empirical human being, including anthropological considerations such as human skill, civil society, and the effects on humanity of war, luxury, and education (5:431-3). Zammito (1992: 267) and Allison (2001: 205) both position the task of schematizing the empirically conditioned human being and the autonomous moral agent as the principal 'gulf' KU is attempting to bridge.

dimension of independently formulating subjective maxims of action.

Natural teleology in Kant's shift on self-perfection

In MS, the connection between natural teleology and practical reason is even more robust. As Kant emphasizes in TL, the conception of self-perfection appropriate to moral philosophy is not *transcendental*, but *teleological.*[145] Understood 'teleologically'—through reflective judgment—perfection refers to "qualitative perfection", or "the harmony of a thing's properties with an *end*" (6:386). Only the teleological conception of perfection admits of more than one kind, since the transcendental conception "can only be one" (6:386). However, moral self-perfection can only be conceived as the former—on the teleological conception. In order to construe oneself as the potential object of self-improvement, one must conceive of oneself as the end of a teleological progression: at once an empirical, anthropological human subject and the bearer of the ideal of moral perfection as one's ultimate end, which is the product in turn of one's ever-present capacity to act in accordance with moral duty.

The distinction Kant draws here between transcendental and teleological conceptions of perfection suggests that Kant's mature elaboration of the cultivation of virtue through self-perfection is contingent on his account of reflective judgment in KU; in an important respect, Kant was unable to formulate the former prior to the latter. Indeed, in GMS Kant only

[145] The complete passage is the following: "The word *perfection* is open to a good deal of misinterpretation. Perfection is sometimes understood as a concept belonging to transcendental philosophy, the concept of the *totality* of the manifold which, taken together, constitutes a thing. Then again, as a concept belonging to *teleology*, it is taken to mean the harmony of a thing's properties with an *end*. Perfection in the first sense could be called quantitative (material) perfection, and in the second, qualitative (formal) perfection. The quantitative perfection of a thing can be only one (for the totality of what belongs to a thing is one). But one thing can have several qualitative perfections, and it is really qualitative perfection that is under discussion here." (6:386)

acknowledges the possibility of the “ontological concept” of perfection, which Kant rejects as heteronomous (4:443). In KU, Kant defines this ontological concept as “transcendental perfection”, contrasting it, as he does once more in MS, with teleological perfection (20:228). While the former can only be singular, admitting of the teleological conception of perfection allows us to “speak of *a* perfection (of which there can be many in a thing under the same concept of it)”, in which case the type of perfection at issue is “grounded in the concept of something, as an end” (20:228). Thus, in GMS Kant appears not to have yet admitted the possibility that self-perfection may take a different form than the transcendental, and thus that teleological judgment may be needed to complete his picture of virtue.

In KpV, Kant newly differentiates between “theoretical” (or “transcendental” and “metaphysical” concepts of perfection) and “practical” perfection, which accords with the teleological definition of perfection as “the fitness or adequacy of a thing for all sorts of ends”; considered “as a characteristic of the human being… [it] is nothing other than talent and… skill” (5:41). While Kant once again rejects the moral role of perfection (including in its ‘practical’ or teleological sense), he does so insofar as it is taken to constitute a practical determining ground of the will. That is, taking perfection to be the sole principle of morality, as Christian Wolff’s (1733) conception does, fails to confer a procedure on which subjects can determine for themselves which ends of action are appropriate; instead, it requires that ends “must first be given to us”, and thus constitutes heteronomy (5:41).[146]

As I read the progression from GMS to KpV to MS, in KpV Kant first introduces a

[146] On this point, see Guyer (2004) on Kant’s differences with Wolff when it comes to self-perfection. As Guyer notes, Kant initially rejects perfectionism; however, the version he comes to endorse puts emphasis on perfecting one’s powers of *choice* rather than perfecting one’s *condition* (312).

distinction between transcendental and teleological modes of perfection, associating the latter with practical self-perfection. But, in KpV, he still lacks the distinction between cultivating skill and cultivating discipline, or between the general aptitude for pursuing various ends and the ability to choose which ends are appropriate (that is, moral) ones to act on. The skill/discipline distinction is what Kant will elaborate in §83 of KU. The characterization of self-perfection Kant advances in MS is contingent on the "negative" notion of discipline, or the "training" of one's will, rather than skill (5:431). In MS Kant refers to self-perfection as the "cultivation of one's *will* (moral cast of mind), so as to satisfy all the requirements of duty", as well as the cultivation of "those concepts that have to do with duty" (6:387).

Consequently, in MS Kant advances a conception of self-perfection that newly retains the connection to moral duty he had first posited as lacking. But arriving at the appropriate formulation of this notion required the account of teleological judgment, and with it reflection on self-cultivation (as discipline rather than skill), Kant first elaborated in KU.[147] In KU, purposiveness is presented as *teleological* perfection, the thought of a perfection of a thing in relation to a possible end or purpose (5:374).

In MS, the appropriate (teleological) conception of self-perfection results only from a comparison between the internally purposive character of natural entities and our own (potentially moral) purposive action as rational agents. We share similar cellular processes and modes of regeneration as other natural organisms; we, too, are internally purposive organisms. Yet we are also *rational* beings, with the ability to *perfect ourselves* in the direction of our own choosing.

[147] On Kant's rejection of perfectionism, see Rawls (2000: 228); on Kant's eventual embrace of a conception of self-perfection in terms of choice of ends, see Guyer (2011).

Thus, while in both KU and MS Kant stresses that teleological perfection allows for the possibility of many possible kinds of perfection under the same concept, this point really only begins to attain force when it comes to our self-conception as perfectible and potentially moral beings. Kant claims repeatedly in MS that every individual must follow her own self-given procedure in perfecting herself, and that this involves the individual derivation of one's own maxims (concepts of duty): "It is a contradiction for me to make another's *perfection* my end.... For the *perfection* of another man, as a person, consists just in this: that he *himself* is able to set his end in accordance with his own concepts of duty" (6:386). One particular agent's chosen procedure of self-perfection may not be valid for another. Thus, while moral self-perfection ultimately aims at the ostensibly "*narrow* obligation" of doing one's "duty *from duty*", Kant insists that the law does not prescribe any specific "action" that would fulfill this duty; the moral agent must set her own individual course in determining which actions would best fulfill this end (6:392).[148]

The course of reflective judgment therefore extends *beyond* that delineated in the *Critique of Teleological Judgment:* the same intermediate conclusions vindicated there in the course of the procedure of teleological judgment—external and internal purposiveness—then serve as premises in the derivation of new moral maxims in GMS and MS, namely FRE (see discussion in §3.4) and self-perfection. Where FRE, as the maxim of autonomy, is vindicated by its dissimilarity to *external* purposiveness, the appropriate conception of self-perfection is vindicated by its dissimilarity to *internal* purposiveness. Autonomy requires us to consider the *context* of our moral action: namely, the abstracted community of other rational beings, who must be conceived in

[148] Although Kant does have certain broad suggestions; see, e.g., Kant's discussion of 'powers of the soul' at 6:445 or the cultivation of the feelings of beneficence, gratitude, and sympathy at 6:452-8.

contrast with the *external* purposiveness observable in nature. Self-perfection, meanwhile, requires us to reflect reflexively on our *own* nature as one intrinsically dissimilar from *internal* purposiveness.[149] In both cases, the purposiveness at issue is one that we can choose (or self-legislate) rather than one that is externally imposed; the analogy to nature enables us to reflect on our *dis*similarity with other natural organisms. In both examples, reflective judgment is required in order for the requisite *a priori* maxims to be properly conceived and derived.

As I argued in the previous chapter, however, reflective judgment is also involved in the formulation of subjective, subsidiary (or 'empirical') maxims. Indeed, in both FRE and self-perfection, the point of dissimilarity to natural organisms hinges on what I characterize as the heautonomy of practical judgment in human agents: the role played by our free choice to formulate and revise our maxims of action (that is, to choose the ends on which to act). In the case of self-perfection, for example, Kant suggests that reflective judgment is required not just to conceive of the appropriate definition of self-perfection in the abstract, but also to undertake a specific *course* of self-perfection as a given moral agent. For example, Kant articulates the command of moral self-perfection as "'be perfect.' But a human being's striving after this end always remains only a progress from *one* perfection to another" (6:446).[150] That is, the course of

---

[149] Both cases of comparison are *disanalogies* of a sort. As I said in §3.4, it is really in *contrast* to external purposiveness that FRE is framed. Moreover, while Kant defines 'teleological perfection' similarly in KU and MS, the notion of one entity's having 'many possible perfections' only seems fully applicable to the case of a rational agent who can choose a particular direction of self-perfection. In both cases, the analogy to nature serves to put into relief our unique capacity to freely determine our ends. However, a productive contrast or disanalogy serving to better elucidate or define a concept, principle, or rule is an instance of reflective judgment; see Kant's discussion of 'difference' as a defining function of reflection in the Amphiboly (A260/B316).

[150] Here, I draw from a point advanced by Makkreel: "Conscience sets the stage for Kant's duties of virtue, which are never determinate, but reflective. One of the duties of virtue is to perfect oneself, but Kant recognizes that 'it is a human being's duty to strive for this perfection, but not to reach it (in this life)'" (2002: 217).

realizing the end of a given mode of self-perfection results in the adoption of a *new* end of self-perfection; the formulation of new ethical ends (or rules) is achieved only in *practice,* in the course of undertaking a self-given procedure. Relatedly, undertaking one's own self-perfection must, in a certain respect, be not just autonomous, but also *heautonomous*: one's own procedure of self-improvement must accord with a normativity one claims for oneself, since this procedure is valid only for the particular agent who chooses to adopt it (even as it is *also* governed by universal general principles). Thus, in the case of natural or pragmatic self-perfection, Kant notes,

> Which of these natural perfections should take *precedence*, and in what proportion one against the other it may be a man's duty to himself to make these natural perfections his end, are matters left for him to choose in accordance with his own rational reflection about what sort of life he would like to lead and whether he has the powers necessary for it (e.g., whether it should be a trade, commerce, or a learned profession). (6:445)

The cultivation of one's own identity and way of life as a human being must, to a certain extent, be left up to the discretion of the individual in question, even as, in order to be virtuous, this telos must at the same time be guided by the maxim of moral self-perfection. And this point seems to be generalizable to moral self-perfection as such: "With regard to perfection as a moral end, it is true that in its end (objectively) there is only one virtue (as moral strength of one's maxims); but in fact (subjectively) there is a multitude of virtues, made up of several different qualities, and it would probably be impossible not to find in it some lack of virtue, if one wanted to look for it" (6:447). If 'it would probably be impossible not to find' in the multitude of virtues 'some lack of virtue', this indicates that there is no definitive answer to the question of how self-perfection can be universally attained. It must remain discretionary, up to the judgment of the individual moral agent—which, as I showed in Chapter 3, is the province not simply of applying rules or subsuming cases, but of one's independent formulation of particular maxims.

4.5 Personal autonomy, moral autonomy, and self-critique

This chapter has argued that the application presupposition, on which the moral principle of autonomy is to be exported from GMS to MS, inadequately accounts for the developments in Kant's views post-1790—particularly the influence of the discovery of reflective judgment, with its own independent principle of autonomy. Thus, while recent literature has been right to draw attention to the changing status of autonomy post-GMS (Kleingeld 2017, 2018), the conclusion that the principle of autonomy drops out of Kant's mature thought altogether fails to account for the role it plays in Kant's final work, the OP. As I have demonstrated, the notion of autonomy remains central to OP and consequently, given Kant's vision for the crucial role this work plays in his philosophical system, to his considered position. However, it resurfaces in a new form: no longer defined in terms of self-legislation, but now a mediating concept that, through its activity of self-making, effectively generates the Kantian critical apparatus. One way to read Kant's insistence in MS on the necessity for individual judgment in applying practical principles to experience is that it implicitly presupposes a form of autonomy as self-making. As I have shown, Kant explicitly situates the role of practical, casuistical judgment as a specifically moral transition that would complement the transition represented by OP, and vice versa. Through Kant's elaboration of the imperfect duties of self-perfection, which can only be conceived teleologically, practical judgment emphasizes the necessity of reflective judgment for moral deliberation in general. Moreover, I have demonstrated that the comparison to natural teleology, which was already invoked in formulating FRE in GMS, introduces new nuance to Kant's initial wholesale rejection of perfectionism following his discovery of a new form of judgment for thinking the teleology of nature in KU.

The reflective dimension of practical judgment in these cases is subject to its own

principle of autonomy, heautonomy, which governs the self-shaping of judgment. Indeed, I have attributed the fact that Kant defends a broader account of autonomy in OP than he does in the critical moral texts to Kant's initial articulation of the principle of heautonomy in KU. Thus, autonomy, on Kant's mature conception, incorporates *both* moral autonomy, or the practical dimension of moral deliberation that legislates a given maxim universally to all (and *ipso facto* to oneself), *and* heautonomy, or the reflective dimension, by which one produces new maxims whose scope extends only to oneself in practice.[151] We might represent these dimensions of practical judgment, as I have represented them in this and the previous chapter, in the form of a diagram (Fig. 3):

Figure 3: The two dimensions of practical judgment

Dimension of judgment	*Direction of reasoning*	*Scope of validity*	*Normativity*
Reflective	Particular -> Universal	Subjective	Heautonomy
Practical	Universal -> Particular	Objective	Autonomy

The consequences of this finding are twofold: first, autonomy remains indispensable to Kant's mature philosophical position. Second, the direction in which Kant develops his account of autonomy in this period puts into question the framing of the distinction between Kantian autonomy and contemporary accounts of autonomy referenced at the beginning of this chapter, or moral autonomy and personal autonomy respectively. If my interpretation has been convincing, Kant's account of autonomy as self-making is arguably more sophisticated than that of many contemporary theorists of (personal) autonomy, insofar as it provides us with a unified account of

[151] On the distinction between universal legislation and self-given maxims, see Kleingeld and Willaschek (2019) and Kleingeld (2018).

both morality and the creation and correction of principles of action (maxims) in general.

More centrally for my purposes in this dissertation, the reframing of this distinction suggests the possibility, currently underexplored, of *critiquing* our current constitution as subjects and practical agents. In order to answer to the demands of a contemporary account of autonomy, one which incorporates personal or political autonomy rather than limiting itself to moral autonomy alone, we need not take Korsgaard's route of appealing to 'practical identity' to *supplement* Kant's view, particularly given that this strategy prematurely forecloses important occasions for self-critique. After all, our identities are not wholly up to us and can thus often prove constraining, foreclosing productive possible avenues of action. Yet construing 'identity' so broadly as to encompass whatever it is a given agent happens to value, as Korsgaard (1996, 2009) does, precludes the possibility of *critique* of preexisting values. By contrast, Kant's account of the self-making of judgment is, on my interpretation, integrated into his account of practical maxim formation. Moreover, it integrates the possibility of critically evaluating contingent principles of action on the basis of experience, one that can be conceived as a kind of *virtue*.

Chapter 5: Kant on Moral Ideas and Virtue

5.1 The right and the good: virtue ethics and Kantian ethics

Kant is paradigmatically understood to be committed to the centrality of moral principles, or maxims (what I termed 'p-maxims' in Chapter 3), to moral reasoning. In this respect, Kant's moral philosophy (the widely supposed paradigm of deontology) is often positioned in opposition to virtue ethics, which takes basic moral facts to be constituted not by moral principles, but by perception of moral particulars (McDowell 1996, Nussbaum 1986) or by moral concepts, particularly the 'thick' concepts instantiated by the virtues themselves (Hursthouse 1999, Annas 2016). What is basic to the difference between the the virtue-ethical and deontological camps is whether moral reasoning principally takes the form of moral rules (prohibitions, commands, and 'oughts', instantiated in the form of maxims and the categorical imperative as metamaxim) or descriptive content (e.g., the salient features of one's circumstances, or the characteristic features of a given virtuous agent).

In Chapter 3, I did endorse an aspect of this view when I introduced the distinction between i-maxims and p-maxims. The notion of an i-maxim already undercut the interpretation of Kant's ethics as rule-based, since an i-maxim is not a moral rule: it is an intention, an action as we make sense of it or describe it to ourselves.[152] But p-maxims fit more straightforwardly into

[152] I-maxims are still normative in that they are reasoned or retain a connection to the agent's sense of the good. But they are less rule-governed than other interpretations have it. For instance, Korsgaard (2009) refers to "universal principles" as both "principles that can consistently be followed in any kind of case" (that is, they are *law-like*: they command consistency) and "specific to the situation at hand" (180, 73). For reasons I explain further in concluding this chapter, I don't take the appeal to consistency to be particularly helpful, particularly when it comes to the need to adapt to the ways in which the space of

the traditional conception of Kantian principles as moral rules.

Given the resources advanced in Chapter 4 regarding heautonomy and autonomy, we can now further complicate this initial distinction: like i-maxims, p-maxims are not reducible to mere rules, either. As I indicated there, Kant's moral philosophy increasingly broadens to acknowledge features of moral reasoning beyond the application of the categorical imperative to encompass the role of *moral ideas* in practical reasoning: moral concepts ('thick' concepts, as virtue theorists have rightly acknowledged) such as virtue, friendship, justice, or humanity. In this chapter, I return to Kant's conception of moral ideas to attempt to better clarify what they are, and what Kant takes their importance in moral deliberation to be.

The upshot will be that the prescriptive/descriptive (or rule/concept) distinction along which the Kantian and Aristotelian camps are often planted oversimplifies Kant's account, particularly in his post-critical works, of the necessary interplay between moral principles and moral ideas. Moral maxims are attempted realizations in prescriptive form of moral ideas, which—as indeterminate representations lacking concrete, sensible instantiations—inevitably elude our cognitive grasp. Correlatively, moral ideas can be thought of as attempted descriptions of ideals which help to guide maxim formulation. P-maxims, then, can be thought of as the prescriptive form taken by moral ideas, much like the 'v-rules' Hursthouse (1999) delineates (e.g., the virtue of fairness aligns with the prescription—the v-rule, or p-maxim—'be fair'). The prescriptive, then, goes hand in hand with the descriptive, the normative with the epistemic.

On the Kantian conception of virtue p-maxims fit into, our current characterizations of moral ideas must be understood as *working* or *attempted* descriptions, rather than as propositions

---

reasons can change.

claiming the status of epistemic certainty. Unlike certain readings of Aristotle's conception of virtue, for Kant virtue therefore consists in the "struggle", "striving", or "endless progress" toward perfection—toward the regulative, or asymptotic, realization of a moral idea—rather than in the state of moral perfection itself (5:83, 5:84, 5:122).[153] And, while interpreters typically parse the 'struggle' in which virtue consists in terms of the strength of the will against the passions or the steeling of oneself against sensuous inclination (Kohl 2017, Merritt 2018), I will suggest that they have missed the importance of the dimension of moral revision: of progress—not merely in the war of reason against affect—but of the progress "*of one's maxims*" (5:33), of our working formulation of moral prescription and our readiness to subject it to *re*description, to self-critique.[154]

This interpretation of Kant's moral philosophy requires us to consider it in closer analogy to Kant's theoretical philosophy than commentators have traditionally done. I present Kant's conception of practical reasoning not as the application of preexisting rules, but as our continual, never-ending striving to realize moral ideals in our working moral terms, both descriptions and

[153] Russell (2019) makes a similar point: On "McDowell's conception of the virtuous agent… all claims that run counter to morality are 'silenced.' Against such a reading, I again would argue that for Kant the claims of inclination and the dear self can never be wholly silenced; rather human beings can only strive for such perfection, while at the same time bearing in mind that ours is a mind that will forever require spur and [bridle]. This struggle and striving is virtue, it is not deficient in virtue, but it is not perfection" (70).

[154] This is not to say that the definition of virtue as strength of the will against inclination lacks textual basis, but just that its importance has been emphasized at the expense of the epistemic dimension of moral revision or moral intelligibility. That said, I think Kant has in view not just the obstacle sensuous inclination poses to the moral will, but also the obstacle of socio-historical contingency (what Kant calls 'habit' or 'imitation' [*Gewohnung*, *Nachahmung*], or our 'horizon' [*Horizont*] of thought, which Kant specifies as historically and socially conditioned; see 9:76, 9:40-45), or the habitual, unreflective reinstantiating of received, socially inculcated patterns of behavior in our individual actions). It is, I think, for this reason that Kant continually stresses the imperative or maxim of enlightenment, of thinking for oneself rather than passively inheriting the commands of the powers-that-be. Merritt (2018) does allude to the problem of 'prejudice', but this understates the issue.

prescriptions. Kant characterizes the critique of theoretical reason as the curbing of speculative reason's interests to claim knowledge of what lies beyond its bounds by means of a recasting of rational concepts not as constitutive of objects of knowledge, but as regulative (or, in other terms, as prescriptive, as what Friedman (1991) terms a 'task') for the course of empirical inquiry. The picture of moral reasoning that emerges once the 'regulative' role of moral ideas is made explicit comes to embody some of these characteristics. While practical reasoning lacks the same intrinsic dangers of speculative reasoning, it, too, can be characterized, on my view, in terms of the continual attempt to realize concepts that elude its cognitive grasp through the alteration of working descriptions of those notions.

Despite this proximity between Kant's conception of ideas of reason in both their theoretical and practical guises, the attention paid to Kant's conception of regulative ideas of reason in the first *Critique* (Friedman 1991, Rush 2000, Massimi 2017) has not extended to what Kant increasingly terms *moral ideas*. While interpretations have recently urged reconsidering elements of Kant's practical philosophy in light of his theory of regulative ideas, they have typically imported Kant's conception of regulative ideas from Kant's theoretical philosophy, in particular from KrV.[155]

Two fundamental points have therefore been missed. First, as Friedman (1992) puts it, "in the First Critique the problems later assigned to the faculty of reflective judgment are discussed under the rubric of the regulative use of reason" (244; see also Grier 2001: 289). As I have shown in Chapters 3 and 4, Kant's views change in crucial ways when it comes to his conception of the regulative employment of reason, in particular in regard to his introduction of the faculty of

[155] Kleingeld (2008) suggests we consider Kant's conception of history as a regulative idea of reason, while Kohl (2017) advances a reading of Kant's doctrine of radical evil as regulative idea.

reflective judgment. Second, in connection with this point, Kant's considered conception of moral ideas has a distinctly *practical* character, one lacking in his critical conception of regulative ideas of reason. As I argue in Section I, regulative ideas are distinguishable on the basis of their *employment*: the same idea of God or the world-whole admits of constitutive, regulative, or, ultimately (as Kant claims in the Canon of Pure Reason), practical employments, and its validity and function are to be evaluated on this basis. Moral ideas, by contrast, can be distinguished from the ideas used in scientific inquiry in virtue of their content as well as their employment. A moral idea is a concept that we attempt to realize or grasp in action rather than in thought, and whose semantic content is therefore aspirationally 'filled in' practically rather than theoretically. I suggest in §§5.2-5.3 that there is overlap between the two conceptions—an idea such as 'humanity' is implicated both in moral reasoning and in anthropological or scientific observation of human kinds and practices—but the shift in emphasis is instructive. In particular, it allows us to reconsider Kant's conception of virtue (§5.4) under the guise of a moral idea: while Kantian virtue is standardly interpreted in terms of strength against inclination, a view which exposes Kantianism to numerous criticisms of disregard for the role of sensibility and feeling in moral reasoning, Kantian virtue considered as moral idea can newly be seen to encompass the dimension of *moral revision*.

5.2 Ideas of reason in the *Critique of Pure Reason*

In the Appendix to the Dialectic and Transcendental Doctrine of Method sections of the first *Critique*, Kant devotes considerable attention to *ideas of reason*: pure concepts of the understanding that lack determinate objects. Indeed, Kant identifies ideas of reason as the source of both reason's interest in unifying cognitions into a system of knowledge and the fallacious uses

of reason that initially required a critique. What Kant has in view here are concepts such as God, the soul, and freedom (A798/B826). These are concepts that, when taken *constitutively*, to constitute real existing entities, lead us into cognitive error, misattributing our intellectual resources to litigating claims that can never be known. This, then, is the employment of reason in its 'speculative' use: as speculating about concerns outside our cognitive grasp. But ideas of reason also have a useful, even indispensable, function when understood *regulatively*: not as constituting objects in themselves, but as guiding reason in the task of systematizing cognition in the pursuit of knowledge. Ideas of reason can do this, Kant thinks, because reason has an *interest* in attaining knowledge of them, even if ultimately realizing this goal remains outside its reach. Ideas thereby 'regulate' empirical inquiry by galvanizing us to learn as much as we can in hopes that we might learn something more about these concepts in the process.

While Kant condemns, then, the *speculative* use of reason, he embraces its *regulative* use. However, in the Canon of Pure Reason, he also claims that our employment of ideas of reason, even when regulative, is directed at a still higher end (what Kant terms the 'ultimate end' of the pure employment of reason): a *practical* employment. Our interest in the ideas freedom, the soul, and God ultimately derive from their importance in the domain of moral action: from "the problem *what we ought to do*, if the will is free, if there is a God and a future world" (A800/B828).

As we can see, at this stage Kant marks out a practical role for regulative ideas, but one that pertains to their mode of employment rather than any distinctive content. The idea 'God' can be employed regulatively in theoretical reasoning, to further our progress in a given science. But the same idea can also be employed practically, as motivating us to reflect on how we should *act*. The real question, then, is not what an idea of reason is in its nature or essence, but what we should *do* if these ideas in fact hold.

In order to make sense of the ambivalent function of regulative ideas—illusory, yet necessary for the function of the understanding—commentators have frequently seized on this, drawing on the analogy to practical reason to specify the positive role of ideas for theoretical cognition. Grier (2001) claims: "Like maxims in the practical philosophy, the demand for systematic unity is a very general self-imposed rule, which, as such, reflects the particular interest of reason", playing a role "somewhat akin to that which the moral law plays in the practical philosophy" (285). Massimi (2017) suggests we take "ideas of reasons as imaginary standpoints—both epistemic and practical—for conferring unanimity and universality (to our scientific knowledge claims no less than to our moral actions, respectively)" (78).

In connection with this interpretive angle, it has often been supposed that a regulative idea *just is* a principle, or maxim, that orders the direction of empirical inquiry—or at least that the two are sufficiently close that they can be referred to in tandem, their differences not warranting explication. Friedman (1991) represents a common usage among commentators when he employs "regulative concepts and principles" interchangeably, claiming that ideas present inquiry with a "task" (73). Massimi (2017) stresses that the root of 'regulative' is the Latin *regula,* or rule, and argues for an interpretation of transcendental ideas as rules for the understanding (or 'idea-rules'), akin to practical rules, rather than as what she casts as the standard interpretation, 'idea-archetypes', or Platonic archetypes of objects that cannot be known (66).

I return to Massimi's idea-rule/idea-archetype distinction below, but it is first worth pausing to reflect on the nature of the connection between theoretical and practical reasoning as it is established in the Canon. What does Kant mean by saying that the 'ultimate end of the pure employment of our reason' rests in its practical employment? How does the regulative use of reason relate to the question of 'what we ought to do'?

The suggestion from Grier (2001) and Massimi (2017)—that ideas of reason provide reasoners with a *focus imaginarius*, an imaginary perspectival standpoint—is interesting on a number of counts, but strikes me as nevertheless incomplete. First, this proposal, insofar as the *focus imaginarius* is taken to confer unanimity on knowledge and action, establishes at most that ideas are important *both* when employed in science and in moral deliberation; it establishes, that is, only a *parity* between the two employments, theoretical and practical. Yet this is not what Kant claims in the Canon. Instead, Kant claims that the practical interest of regulative ideas is the *ultimate end* of the pure employment of reason. That is, he suggests that the regulative use of reason is in some way *subordinated under* practical reason—and this, not just when it comes to achieving consensus on scientific judgments, which, as many commentators have recently argued, might be thought to invoke a corollary to the categorical imperative for our use of understanding[156], but in considering questions of what we should *do.* What Kant means, then, is that the regulative use of reason is ultimately directed towards the question of action, and must in some way help us towards this end; it does not hold equal and distinct standing in either domain of theoretical and practical reasoning, respectively. In particular, Kant delineates a need for reassuring ourselves (for 'hope', A805-6/B833-4) that moral principles "*might* be met with in the *history* of mankind", that "the idea of a moral world has… objective reality" (A807-8/B835-6).

To anticipate the response I offer below, my view is that Kant comes to advance a more complex account of the role of ideas in moral reasoning by the time of the Doctrine of Virtue. On this later account, however, ideas are not the *same* as rules, principles, or maxims, although they do remain tightly bound up with their formulation and refinement (and vice versa, when it

[156] See O'Neill (1989), Chignell (2007), Cohen (2014), Kohl (2015), Merritt (2018) for current discussions on the role of epistemic normativity in Kant's theoretical philosophy.

comes to the 'sharpening' and 'development' of moral ideas, as we will see). Moreover, moral ideas are not exhausted by the "three official transcendental ideas of God, soul, and world"[157] on which Kant focuses in the Appendix to the Dialectic. My account already suggests two responses to these lines of discussion: first, that we shift from assimilating ideas of reason to prescriptive rules to instead consider the relation *between* ideas and rules (or maxims); second, that we reconsider the role of the *descriptive content* of any given idea.[158]

To set up these claims, I want to first draw attention to several elements in Kant's account in the first *Critique*. First, the question of the prescriptive function of ideas of reason: are ideas of reason simply *rules*, in this case rules for judgment?[159] This is the account commentators have

---

[157] Massimi (2017): 74; see also Geiger (2003) and Zuckert (2017).

[158] Massimi (2017) argues that "it does not matter whether we are dealing with the ideas of soul and God; or with the idea of 'pure water' and 'fundamental power'…. Instead, systematicity is delivered by ideas qua *foci imaginarii* that act as rules (*[Idea]-Rule*) for the correct empirical use of the understanding" (64). This seems to me to be too dismissive; for Kant, I think, the distinctive moral content of an idea such as 'justice' or 'humanity' became increasingly important. Thus Kant was led to develop a theory of the differences in content between aesthetic ideas and rational ideas, respectively, subdifferentiating the latter into theoretical and practical ideas, where each of these differences are carved up in accordance with the semantic content of the idea at issue as well as its employment.

[159] At one point, Kant characterizes ideas as analogous to schemata: "The psychological and theological ideas…. should not be assumed in themselves, but their reality should hold only as that of a schema of the regulative principle for the systematic unity of all cognitions of nature" (A665/B693, A674/B702). Grier (2001) comments: "Just as, in the Transcendental Deduction, [Kant] needed a synthetic step to connect the categories to empirical intuitions in general (or spatiotemporal intuitions), so too, Kant now needs something analogous to a synthetic connection between reason and the matter of *empirical cognition*. In the Critique, Kant attempts to forge such a connection… by the ideas as analoga of schemata, although later on (in the *Critique of Judgment*) he does so by judgment" (2001: 289). This would then suggest that ideas are the "conditions for the application of various principles and maxims of reason" (235). For this reason, Grier's view should not be assimilated to Massimi's: rather than presenting ideas as rules as Massimi does, Grier suggests they function as schemata for principles of reason. I still think this is not right; in the third *Critique*, Kant makes clear that it is not ideas that *are* analogous to schemata, but that *require* a function analogous to schemata—one provided, not by the ideas themselves, but by 'symbols' (5:351-2). And this is reinforced by other passages in the first *Critique* where Kant claims that ideas are not themselves (a kind of) schema, but require schemata for their execution (A645/B674, A835/B863). So this still leaves the question open of what ideas are, if not rules or schemata. See also my discussion in Chapter 4 of my opposition to Zuckert's reading along similar lines, as well as my consideration below of A643/B671, where Kant attributes 'all errors of subreption' to judgment.

frequently gravitated towards, whether explicitly, as in Massimi's (2017) endorsement of ideas as rules rather than ideas as archetypes, or implicitly, in discussing ideas as of a piece with principles or maxims (Rush 2000: 839, Friedman 2010: 73). But, although Kant is admittedly not always consistent in his characterization in the first *Critique*, I think he already gives us reason to suspect that the assimilation of ideas to rules *tout court* is too quick. In particular, I want to suggest that one aspect of Kant's account of the regulative employment of rational ideas remains relevant for Kant's account of specifically *moral* ideas in the third *Critique* and Doctrine of Virtue: the tight connection Kant establishes between ideas of reason and maxim formation.

Are regulative ideas principles of reason?

First, although the two conceptions of ideas and principles remain closely aligned, Kant generally retains a distinction between them. He claims, for example, that when employed regulatively in knowledge formation, an idea "*helps to find a principle* for the manifold, thereby guiding [the understanding] even in those cases that are not given" (A646/B674; my italics). That is, Kant does not claim that ideas *are* principles, but that they help to *find* principles. The principles formed in the process of systematizing experience under an idea of reason are termed 'maxims':

> I call *maxims* of reason all subjective principles that are obtained not from the character of the object, but from reason's interest concerning a certain possible perfection of the cognition of this object.... The systematic unity of the manifold knowledge of the understanding, as prescribed by reason, is a *logical* principle. Its function is to assist the understanding by means of ideas in those cases in which the understanding cannot by itself establish rules. (A648/B676)

The role of maxims in *theoretical* cognition, then, allows cognizers to proceed in accordance with

ideas that would otherwise not be available to cognitive representation. Here again, Kant claims that the 'function' of systematic unity as a 'principle' is to 'assist the understanding by means of ideas': he does not claim that the principle *is* an idea, but that it assists *by means* of ideas. A close connection between ideas and principles, to be sure, where each 'helps' reason to secure or grasp the other, but not a straightforward assimilation of one to the other.

Whether employments of ideas of reason are fallacious (constitutive) or effective (regulative) depends on the recognition of this intrinsic connection between ideas and maxims; if not, ideas are hypostasized into concepts of objects: "If merely regulative principles are considered as constitutive, then as objective principles they can be in conflict; but if one considers them merely as *maxims*, then it is not a true conflict, but it is merely a dif-ferent interest of reason that causes a divorce between ways of think-ing" (A666/B694). A passage like this one may lead us to think Kant understands ideas as principles, since he refers to 'regulative principles' rather than ideas, while describing the same conflicts of reason to which ideas give rise. Similarly, there are instances in which Kant refers to the same object, say, 'systematic unity', both as idea (A646/B674) and as maxim (A680/B708). As I elaborate below, however, this need not lead us to subsume the two conceptions: we can understand the relation between ideas and principles to consist in the descriptive and prescriptive *aspects,* respectively, of the same cognitive object.

There are also textual instances in which Kant does seem to assimilate the two conceptions: in KU, for example, Kant remarks, "Ideas, for which no appropriate objects can be given in experience… could therefore serve only as regulative principles in the pursuit of experience" (5:405). As I said above, Kant is not always consistent in his usage, but I take the preponderance of the evidence to fall on the side of the interpretation I advance in this chapter. Indeed, Kant also insists explicitly on the distinction between them, particularly when it comes to

morality: "The *estimation* of morality, in regard to its purity and consequences, is effected in accordance with *ideas*, the *observance* of its laws in accordance with *maxims*" (A812/B840). The *descriptive* or *epistemic* dimension of moral reasoning, Kant claims, is the purview of ideas, while the *prescriptive* or *law-governed* dimension is the purview of maxims.

There is a broader reason why I raise the issue of the relation between principles and ideas. As I understand it, the assimilation in the literature of regulative ideas to rules attempts to read Kant's practical philosophy into his theoretical philosophy, as suggested by the Canon of Pure Reason and Kant's attribution of a positive cognitive function to rational ideas: on this account, a regulative idea is a kind of maxim or principle for the understanding or for theoretical reasoning, analogous to the way in which we act on maxims in practical reasoning. Yet this interpretation cannot explain why Kant devotes increasing attention in his practical philosophy to *moral* ideas. If ideas *just are* maxims, albeit in helping the understanding form empirical judgments, why do we need ideas in practical reasoning at all? Why not just take practical reasoning to be exhausted by moral rules, and let ideas fall out altogether? To understand what might be *distinctive* about the role of ideas in morality, we need a better understanding of the relation between ideas and maxims—an issue more evident when it comes to the role of ideas in Kant's moral philosophy.

Is the content of regulative ideas relevant?

Commentators on regulative ideas often focus on the three "official" ideas of God, the soul, and the world (Massimi 2017: 73; see also Grier 2001, Zuckert 2017). It is these to which Kant devotes his attention in the second part of the Appendix, and it is these which are taken to relate to moral actualization; indeed, these are the ideas mentioned in the Canon). Yet Kant also

refers to other, 'unofficial' ideas in the first part of the Appendix, such as "pure earth, pure water, pure air" (A646/B674) or "fundamental power" (A649/B677), a fact which has puzzled commentators.[160]

Some have responded to this problem by denying the relevance of the content of any given idea altogether.[161] But this comes at a cost, primarily when it comes to Kant's increasing differentiation of various kinds of ideas—aesthetic, moral, and theoretical—on the basis of content, as we will see. Moreover, it's not clear what it means for the notion of 'pure water' or 'fundamental power' to serve as an imaginary standpoint for attaining "unanimity and universality" (Massimi 2017: 65). An idea such as God or the soul might help to foster a "shared conversational scoreboard", namely by fostering conversation about what these unrepresentable notions consist in; this is a role akin to the one Kant assigns to the function of aesthetic ideas in fostering communication (see below) (77). But Massimi does not explain how the same can be said for 'pure water' or 'pure air', concepts no longer in our repertoire and, as such, tending not to spark much debate.

I think there is a better explanation for why Kant groups seemingly disparate notions under the rubric of regulative ideas in the Appendix, one suggested by several other, less-cited

[160] Massimi (2017): 67; see also Geiger (2003), Zuckert (2017). Indeed, Friedman (1992) argues that Kant's method in MAN, a text which concerns itself with justifying the *a priori* role that moving forces, powers, and pure concepts of matter (e.g., pure water, pure air) play in the scientific method, is constitutive rather than reflective (or, presumably, regulative, since he acknowledges that "in the First Critique the problems later assigned to the faculty of reflective judgment are discussed under the rubric of the regulative use of reason", 244; see also 254). On this score, I find Emundts' (2004) proposal instructive: the metaphysical foundations of science, as they are increasingly specified (thus rendered increasingly empirical), must admit of revision, and therefore entail a reflective dimension of evaluation. That this is the case can be shown by the fact that we no longer employ terms such as 'ether' or 'pure water' in science.

[161] "It does not matter whether we are dealing with the ideas of soul and God; or with the idea of 'pure water' and 'fundamental power'" (Massimi 2017: 64).

examples of ideas Kant gives in the Ideal of Pure Reason section. The "moral concepts" of humanity, virtue, and human wisdom are really *ideas*: Kant refers to "our idea of perfect humanity", and claims that "virtue, and therewith human wisdom in its complete purity, are ideas" (A568-9/B596-7). That is, Kant already suggests that there may be many more ideas than the 'official' three, some of which are more morally significant than others. Consequently, while Kant's main emphasis on the *employment* of a given idea might be interpreted as suggesting that an idea's content is irrelevant, there are also indications of a broadening out of the notion of an 'idea' (not just 'soul', 'God', or 'world', but 'humanity', 'virtue', 'wisdom'), admitting of ideas with explicitly *moral* content.

This conclusion is bolstered by considering the extent to which many aspects of Kant's account of reflective judgment in the third *Critique* are anticipated in the Appendix to the Dialectic of KrV, albeit not yet under this rubric (Friedman 1992, Grier 2001). Kant claims that, where "the particular is certain but the universality of the rule for this consequent is still a problem", the universal is "assumed only *problematically* and is a mere idea" (A646/B674). In this process, the "idea is only a heuristic and not an ostensive concept; and it shows… how, under the guidance of that concept, we ought to *seek* after the constitution and connection of objects of experience in general" (A671/B699). While in the case of a certain universal "only *judgment* is required for subsuming", the case of a problematic universal instead calls for the "regulative" use of reason (A646/B674).

Thus, Kant already acknowledges cases in which particulars lack universals, and already accords a problematic (or hypothetical), heuristic role to the universal rules found for them. These rules are termed 'maxims', as in the third *Critique*, which guide the understanding in its methodological course of inquiry. Yet the role of judgment in all of this is not yet perspicuous. On

the one hand, Kant claims, "All errors of subreption are to be ascribed to a defect of the power of judgment, never to understanding or to reason" (A643/B671).[162] The fallacious employment of regulative ideas (here, the substitution of concepts constitutive of knowledge claims for ideas of reason, or 'subreption') is, in this passage, accorded to the power of judgment. Yet Kant does not yet emphasize the power of judgment's role: Kant tends to refer only to the 'regulative use of reason' or the 'regulative employment of reason', without acknowledging that it is the *power of judgment* that is responsible for the way in which reason is 'employed'.

5.3 Ideas of reason in the *Critique of Judgment*

In KU, Kant largely retains the general definition of ideas of reason: "Ideas, in the broadest sense, are presentations referred to an object according to a *certain principle* (subjective or objective) but are such that they can still never become cognition of an object" (5b:342). Yet, as I emphasized in the previous chapter, Kant newly accords the role of thinking ideas to judgment in its *reflective* use—thereby newly accentuating reflective judgment in his account of ideas of reason. He writes, for example, that the "idea" of a natural purpose "is not a principle of reason for the understanding, but for the power of judgment… where, indeed, the judgment cannot be determining, but merely reflecting" (5:405). Indeed, the ideas involved in thinking the systematicity of nature are no longer classified as regulative for (the use of) *reason,* but are more explicitly accorded to *judgment,* insofar as the configuration of our cognitive faculties requires them in order to organize experience into ordered knowledge. As such, they, too, are thought in terms of principles, or maxims, "without which there could be no employment of reason" (A651-

[162] See also discussion in footnote above on ideas as 'analoga of schemata.'

2/B680-1).

I established in the last section that there is reason to doubt any easy assimilation of regulative ideas to rules or principles. Yet, as I noted, this is not to say that Kant is consistent in his treatment, particularly between the first and third *Critiques*. Indeed, Kant seems to indicate more than one possible mode of relation between ideas and maxims of reason. In KrV, Kant often makes it sound as if ideas are *a priori* and, due to reason's interest in attaining cognition of them, help to produce maxims of reason to guide our pursuit of knowledge. Ideas, then, would *guide* the formation of maxims. This is how I interpret claims such as "systematic unity (as mere idea)… helps to find a principle for the manifold" (A646/B674). In KU, by contrast, Kant sometimes makes it sound as if ideas are themselves *guided by* maxims, insofar as they are "related to an object in accordance with a certain principle" (5:342). In this second sense, then, the maxim is given independently of the idea. Between the two texts, Kant also sometimes seems to reassign the role of one to the other. For example, while in KrV it is the "idea" which is "only a heuristic" and whose "guidance" indicates how "we ought to *seek* after the constitution and connection of objects of expe-rience in general", in KU—as we saw in Chapter 3—it is the principle or maxim of natural purposiveness which serves as a heuristic to guide inquiry (A671/B699).

These apparent discrepancies have contributed to the assimilation of ideas to principles in the literature. Yet there is another interpretive possibility: the relation between ideas and principles of reason, where each is variously characterized as reciprocally informing the other, can be understood as demonstrating the simultaneously normative and (aspirationally) descriptive character of ideas of reason. Regulative ideas, as concepts lacking determinate objects, must be treated *as* prescriptive—as both maxim-guided and maxim-guiding—for the activity of reason. The idea of nature as a system, for example, does not pick out any given particular, but our

interest in giving distinctness to this inchoate concept spurs us on in empirical inquiry, "thereby approximating the rule to universality" (A646/B674). The idea can be expressed both *descriptively* as a problematic or indeterminate concept which fails to correspond to a given object ('natural systematicity'), or *prescriptively* in the form of a maxim, as prescribing a procedure for judging nature ('nature *ought to* be judged as a system'). Reason requires subjective maxims in attempting to grasp ideas because, as I claimed in Chapter 4, in the absence of determinate rules, it must give itself its own normative course to follow.

When we encounter a given organism, for example, Kant thinks we find ourselves faced *both* with the idea of a natural purpose, which we project onto it even though it cannot be taken to objectively correspond to it, *and* with the maxim that dictates that we *ought* to judge the organism as internally purposive rather than as mechanical: "Certain products of nature, as far as their possibility is concerned, *must*, given the particular constitution of our understanding, *be considered by us* as intentional and generated as ends" (5:405). If we consider organic purposiveness as an idea, we take in view its subjectively descriptive content (to the extent, at least, to which it can be approximated in a given particular); if we consider it as a maxim, we take in view its prescriptive command to the activity of the power of judgment. I raised similar issues in Chapter 3 regarding reflective judgment's formation of 'empirical maxims' (including organic purposiveness): they are empirical because they are contingent on our *initial confrontation with an empirical particular* (with an object we can *describe,* or discursively represent; see Fig. 1). Prescription follows from description, but goes on to guide us in searching for *new* possibilities for description.

This example helps to show how Kant can hold that ideas of reason are both guided by maxims and themselves guide the formation of maxims. The idea of a natural purpose or

organism, for example, is posited in a course of inquiry guided by the maxim of natural purposiveness, and can itself be taken as a sort of specification or further determination of the initial, more general idea of the purposiveness of nature as a whole. In turn, reflection on a given organism gives rise to the new, more determinate maxim of internal purposiveness. Consequently, the relation between ideas and maxims of teleological reasoning demonstrates the simultaneously descriptive and normative character of the procedure of reflective judgment: presupposing a transcendental *a priori* maxim as a heuristic in accordance with an idea of reason, proceeding in the course of empirical investigation on the basis of this maxim, and deriving subsidiary maxims on the basis of the particulars encountered in experience—which are themselves judged in accordance with a better specified idea.

It is for this reason that Kant ceases to refer to the 'regulative' employment of ideas in the third *Critique*. Ideas are still central to Kant's account, as we have seen in this section (as well as in the preceding chapters). But, if their regulating function is now specified as applying to *reflective judgment* rather than to *reason*, ideas must take on a new role. Since the power of judgment gives itself its own normativity (heautonomy), ideas can no longer regulate the activity of reasoning all by themselves. Instead, they are to be thought as the *descriptive* aspect of the *principles* of reflective judgment (*a priori* or empirical), which do the regulating or *normative* work (as I showed in the discussion of autonomy as self-making). Both ideas and principles 'guide' reasoning (in terms of my own focus, practical reasoning), but, as I turn to below, in different ways.

The introduction of reflective judgment also enables Kant to make clearer distinctions in the content of ideas: since the employment of ideas is now the purview of a distinct theory—namely, the theory of reflective judgment—the question of how ideas are *used* can be separated

from the question of what ideas themselves *are*, of their *content.*

5.4 Moral ideas

How does all this bear on Kant's account of *moral* ideas? I claimed that while, in KrV, Kant largely did not differentiate moral ideas from theoretical ideas in terms of content, he did differentiate divergent *employments* of ideas of reason: speculative, regulative, and, ultimately, practical. This changes in KU: there, for example, Kant newly differentiates ideas of sensibility (*aesthetic ideas*) from ideas of reason,[163] the latter of which, in turn, are more explicitly divided into ideas involving theoretical and practical *content,* not merely their employment.[164]

Aesthetic ideas and thick concepts

If a rational idea is understood as a problematic universal or undeterminable concept, an aesthetic idea can be considered to be an unsubsumable manifold. An aesthetic idea is a sensible representation which resists a determinate description, for which "no expression designating a determinate concept can be found for it", allowing for the addition of "much that is unnameable" (5:317). A rational idea, by contrast, is a pure 'problematic' concept lacking a determinate sensible particular, and which thereby precludes application to experience.

[163] Although there is already an initial characterization in the first *Critique* of "ideals of sensibility", "creatures of imagination, of which no one can give an explanation or an intelligible concept… constituting more a wavering sketch, as it were, which mediates between various appearances, than a determinate image, such as what painters and physiognomists say they have in their heads, and is supposed to be an incommunicable silhouette of their products or even of their critical judgments" (A570/B578). Kant contrasts these 'ideals of sensibility', which lack rules or concepts, with the "ideal of reason", which "always rests on deter-minate concepts and must serve as a rule" (A570/B578).

[164] A note on terminology in what follows: since Kant includes ideas of both theoretical and practical reasoning under the term 'rational ideas' , for simplicity, I will term ideas employed in theoretical reasoning 'theoretical ideas' and ideas employed in practical reasoning 'moral ideas'. I will use 'regulative idea' to refer to Kant's critical conception of ideas of reason admitting of theoretical and practical employments, and I will continue to use 'rational idea' or 'idea' as an umbrella concept when none of the above specifications are necessary.

Because aesthetic ideas are representations of the imagination to which "no language fully attains or can make intelligible [*verständlich*]", they "animate the mind by opening up for it the prospect of an immeasurable field of related representations", giving "the imagination an impetus to think more, although in an undeveloped way, than can be comprehended in a concept" (5:315). In so doing, the imagination embarks on a line of thought that can never be fully brought to a determinate conclusion. That is, the beautiful artwork encourages the imagination in perpetuating its activity of thought to the point where the aesthetic ideas it generates "approximate a representation of concepts of reason (of intellectual ideas)" (5:314). One way to put the relation, then, might be to say that an aesthetic idea affords an approximation of a sensible representation of a rational idea, while a rational idea affords an approximation of a concept of an aesthetic idea.[165]

Perhaps this way of thinking of the relation between ideas of reason and ideas of imagination can help us to make sense of why, as Paul Guyer (1979: 362) notes, the examples cited as ideas of reason made sensible by aesthetic ideas are moral or morally significant notions, such as, positively, the "idea of reason of a cosmopolitan disposition", "the consciousness of virtue", "the kingdom of the blessed", "love", and negatively, "envy and all sorts of vices", "fame", or "the kingdom of hell" (5:316, 5:314). Kant focus his attention in the third *Critique* on a proliferation of ideas of reason, many of which are explicitly practical in nature, beyond God, the soul, and the world—including thick ethical concepts, like 'fame' or 'envy', that have both

[165] Chignell (1998) claims: "It is not the *content* of these thoughts [promoted by aesthetic ideas] that is of primary normative importance, but rather the *formal* manner in which these thoughts are strung together by the mind into a 'coherent whole' that has the phenomenological feel of both unity and inexhaustibility" (424). I think this is fine as far as it goes (namely, in terms of cashing out the distinction between aesthetic and rational ideas), but doesn't help much when it comes to moral and theoretical ideas.

normative and descriptive content. Aesthetic ideas, it seems, must at some stage of reflection *give way to* moral ideas. Kant states explicitly that a given beautiful artwork, and the aesthetic ideas it promotes (5:313), must be connected by the subject to 'moral ideas' in order to avoid becoming "loathsome": "If the beautiful arts are not combined, whether closely or at a distance, with moral ideas, which alone carry with them a self-sufficient satisfaction, then [charm or emotion] is their ultimate fate. They then serve only for diversion" (5:326).[166]

The 'development' of moral ideas

Since aesthetic and moral ideas can each at best serve as *approximations* of the other, our sense of any given idea amounts to a kind of working conception. Kant claims, in referring to the idea of universal history, that "nature only requires of us that we should approximate to this idea" requiring "many fruitless attempts", since "reason itself does not operate instinctively, but rather needs attempts, practice and instruction in order gradually to progress from one stage of insight to another" (8:18, 8:19).[167] Kleingeld (2008) emphasizes: "This openness implies that Kant never closes off the possibility that a different idea may actually do a better job of organizing the empirical materials than the one he proposes. He does not stipulate that he has found the only or even the best account of the course of history" (2008: 525).[168]

The same point applies to Kant's conception of moral ideas more generally. Kant claims repeatedly that moral ideas admit of 'development':

It is a fact that what is called sublime by us, having been prepared through culture [*Kultur*],

---

[166] I say more about the connection between aesthetic and moral ideas in Vaccarino Bremner (forthcoming).
[167] See also Kant's other repeated references to giving a particular characterization of this idea as an "attempt", as Kleingeld (2008: 525) notes (8:29, 8:30, 8:31).
[168] I would put the point in terms of a different *conception* of a given idea rather than a different idea altogether, but otherwise concur with the general sentiment.

> comes across as merely repellent to a person who is uncultured and lacking in the development of moral ideas [*Entwicklung sittlicher Ideen*]. (5:265)
>
> The propaedeutic that will truly establish our taste consists in developing our moral ideas [*Entwicklung sittlicher Ideen*] and the culture of the moral feeling. (5:356)

The 'development' at issue is not just that of individual moral character, but one that has also taken a historical course of progression, as Kant had already claimed in the conclusion of the Canon of Pure Reason in KrV:

> Thus we do indeed find, in the history of human reason, that before the moral concepts were sufficiently purified and determined…. our acquaintance with nature and even a considerable degree of the culture of reason in various other sciences could produce only crude and erratic concepts of the deity, and in part they left people with an amazing general indifference regarding this question. A greater treatment of *moral ideas*… sharpened reason for dealing with this object. (A817/B845)

As Kant claims here, moral concepts necessitated, over the course of history, being 'purified' and 'determined', such that moral ideas could be better accounted for. An analogous process, Kant contends, has occurred in the history of science, in a passage I cited in Chapter 1: "It is first possible for us to glimpse the idea in a clearer light… only after we have long collected relevant cognitions haphazardly like building materials and worked through them technically with only a hint from an idea lying hidden within us", initially appearing as a "mere confluence of aggregated concepts, garbled at first but com-plete in time" (A835/B863). In LJ, Kant goes so far as to claim that "most men *lack* the idea of humanity, the idea of a perfect republic, of a life in felicity, and the like. Many men have no idea of what they want, proceeding therefore by instinct or authority" (9:93).

What to make of Kant's talk of 'developing', 'determining', or 'sharpening' moral ideas, if not as references to the effects of aesthetic judgment? To make sense of this, Kant's claims about

moral ideas in MS are helpful.

Regulative ideas, to recall, elide a determinate cognitive grasp, insofar as they do not correspond to any object one might come across in experience. Yet, insofar as they are thought in accordance with principles or maxims, they also have a normative dimension.

Both of these points hold true, albeit in slightly different ways, for moral ideas. In TL, Kant claims that moral ideas are concepts of reason "to which no object given in experience can be adequate"—including, as we will see, in our own attempted enactment of the moral idea in action (6:371). In ApV, he writes, "Ideas are concepts of a perfection that we can always approach but never completely attain" (7:200). The normative dimension of moral ideas does not extend to the activity of theoretical reason in advancing the ideal of unifying experience into a system (as theoretical ideas do), but to practical reason, in establishing an ideal to which we strive in formulating maxims of action—or so I will argue.

Concepts counted as 'ideas' in TL include a 'system of freedom like a system of nature' (6:218), duty (6:219), freedom (6:225), God (6:439), virtue (6:447), humanity (6:404), and friendship (6:469). There are also numerous references to 'ideas' throughout the Doctrine of Right, including: justice (6:334), a civil constitution (6:372), the original (social) contract (6:315), sovereignty (6:371), the state (6:313), and acquisition (6:291). Since it might be argued that these ideas are not purely moral ideas but political (in other words, theoretical) ideas, I will restrict my discussion to the moral ideas referred to in the Doctrine of Virtue (TL) in what follows.

Moral ideas, like theoretical ideas, evade a determinate cognitive grasp, albeit not when it comes to finding particulars already found in experience to which these concepts correspond, but in *enacting* new particulars in action, which will, Kant thinks, only serve as imperfect approximations of the ideal which they aspire to realize. As Kant characterizes virtue as moral

idea (a passage I also cited in Chapter 4), "With regard to perfection as a moral end, it is true that in its Idea (objectively) there is only *one* virtue (as moral strength of one's maxims); but in fact (subjectively) there is a multitude of virtues, made up of several different qualities, and it would probably be impossible not to find in it some lack of virtue, if one wanted to look for it" (6:447). The moral idea of virtue guides our 'subjective', self-given constitution of virtue—our formulation of maxims of action and moral principle. Consequently, there are numerous ways in which the moral idea of virtue might be approximated in experience, although never to the degree of certainty required to claim that the idea has in fact been *realized*. As Kant claims, "The depths of the human heart are unfathomable. Who knows himself well enough to say, when he feels the incentive to fulfill his duty, whether it proceeds entirely from the representation of the law or whether there are not many other sensible impulses contributing to it that look to one's advantage…?" (6:447). Similarly, "Friendship is only an idea… and unattainable in practice, although striving for friendship (as a maximum of good disposition toward each other) is a duty set by reason" (6:469).

Moral ideas as working notions guiding maxim formulation

A moral idea such as 'friendship' or 'virtue', then, entails no prospect of ever being fully realized; as Kant says, it remains 'unattainable in practice'. Instead, it sets us a goal to which we can 'strive', and thereby attempt to approximate in the maxims we set ourselves. To anticipate my concluding argument, the upshot is that Kant's notion of moral ideas seems to come bound up with the prospect of revising self-given principles (the prospect, of course, elaborated in the rest of this dissertation). Practical deliberation, then, is not just about the application of rules, but also extends to a critical attention towards the 'rules' themselves.

It is for this reason that I insisted above on keeping ideas and principles (or maxims) of reason apart. If we think of moral reasoning as acting on rules[169] (and theoretical reasoning as conceptual, and therefore as rule-governed, as well), the normative dimension of ideas has to be thought as a kind of *metanormativity*, a mode of guidance in determining how to *form* the first-order rules of thought (concepts) and behavior (maxims).[170] However, we should resist thinking of the metanormativity of ideas as second-order rules, which might be taken to entail that they are fixed, unchanging, and merely in need of application. Instead, as I suggested above, they are better conceived as *working conceptions*, as valid insofar as they prove "useful [*brauchbar*]" for cognition and can "serve us as a guiding thread [*Leitfaden*]" (8:29).[171] Their very "semantic depth", as I have suggested elsewhere, can therefore change in the course of use (Platts 1988, Moody-Adams 1994).

Already in the first *Critique*, Kant had claimed that ideas of reason call for schemata, and thus, as we saw, the activity of judgment: "The idea [of knowledge as a system] requires for its realization a schema" (A645/B674). In Chapter 4, I argued that a parallel function for a *moral* schematism is established for a metaphysics of morals, which requires "a transition which, by applying the pure principles of duty to cases of experience, would *schematize* these principles, as it

[169] The reasons I advanced in Chapter 4 as to why this is too simplistic an account of moral reasoning can be extended to theoretical reasoning, although these concerns are outside the scope of this dissertation; Bell (1987) gives a convincing account of the need of reflective judgment to account for Kant's conception of theoretical reasoning as rule-governed spontaneity.

[170] As Kant claims in LJ, "ideas serve to lead the understanding through reason with regard to experience and using its *rules* to the greatest completeness [*Vollkommenheit*]" (9:92; my italics).

[171] The full citation is as follows, with reference to the idea of a universal history: "Yet if it may be assumed that nature does not work without a plan and purposeful end, even amidst the arbitrary play of human freedom, this idea might nevertheless prove useful [*brauchbar*]. And although we are too short-sighted to perceive the hidden mechanism of nature's scheme, this idea may yet serve as a leading thread [*Leitfaden*] to us in representing an otherwise planless *aggregate* of human actions as conforming, at least when considered as a whole, to a *system*" (8:29).

were, and present them as ready for moral-practical use" (6:468).

When it comes to the regulative role of moral ideas, the parallel continues to hold. Much as ideas of theoretical reason can never be determined by any given particular, Kant claims that ideas of practical reason are 'unattainable in practice'. While many Kantians would take this to be a point about the threat sensuous inclination poses to moral motivation,[172] I parse Kant's account as broader in scope. The account Kant posits of the role maxim (or concept) formation plays in our attempts to grasp theoretical or regulative ideas also pertains to moral ideas. Ideas of *both* theoretical and practical reason help to guide our formulation of maxims, in the former case maxims of empirical inquiry, in the latter maxims of action. Our working principles—our maxims—serve as defeasible approximations of an idea that cannot be fully grasped in thought.[173]

Moral ideas and moral archetypes

Chignell (2008) claims that moral ideas are really "ideals by which we can guide our actions in the world", having in mind, presumably, the topic of the Ideal of Pure Reason prior to the Appendix (420). I don't fully agree with this reading, but the point of disagreement is a productive one for my purposes. Already in the Ideal, Kant is careful to distinguish moral ideas, which present the problem of application *in concreto*, from moral ideals, which are to be applied

[172] See Kohl (2017: 659-60) and Merritt (2018: 152-5) for recent expressions of this view.

[173] Much as a theoretical idea helps to guide the discovery of new universals (empirical concepts) under which particulars can be subsumed, moral ideas, too, guide the production of subjectively valid principles or maxims—that is, characterizations of action which extend only to the individual moral agent in scope, by which she then shapes the world through her action. As Kant puts it in the third *Critique*, "Pure reason, as a practical faculty, i.e., as a faculty for determining the free use of our causality by means of ideas (pure concepts of reason), not only contains a regulative principle for our actions in the moral law, but at the same time also thereby provides a *subjectively* constitutive one, in the concept of an object that only reason can think and which is to be made actual by means of our actions in the world in accordance with that concept" (5:453).

in individuo: "What I entitle the *ideal* seems to be further removed from objective reality even than the idea. By the ideal I understand the idea, not merely *in concreto*, but *in individuo*, that is, as an individual thing, determinable or even determined by the idea alone" (A568/B596).[174] A moral ideal is not a moral idea (or concept, and therefore a *universal*), such as humanity or virtue, but an archetype—an individual given in intuition, a *particular*, who exemplifies moral properties:

> Virtue, and therewith human wisdom in its complete purity, are ideas. The wise man (of the Stoics) is, however, an ideal, that is, a man existing in thought only, but in complete conformity with the idea of wisdom. As the idea gives the *rule*, so the ideal in such a case serves as the *archetype* for the complete determination of the copy; and we have no other standard for our actions than the conduct of this divine man within us. (A569/B597)

Kant refers to the Platonic conception of ideas in the Ideal, but I think what he also has in mind here is the Aristotelian conception of virtue as the moral standard set by a particular individual, by example, rather than by a universal rule. A moral *idea*, however, as Kant presents the issue in the first *Critique* (a usage consistent with his usage in the third *Critique* and Doctrine of Virtue), is not a particular (the 'wise man'), but a universal ('human wisdom'). And, as we saw in Chapter 4, what Kant has in view with the problem of a 'moral schematism' in TL is application of moral principle *in concreto,* not *in individuo.*

The distinction is muddied a bit by a passage from the second *Critique* referring to moral *ideas* that seemingly instead describes moral *ideals*:

> If I understand by an *idea* a perfection to which nothing adequate can be given in experience, the moral ideas are not, on that account, something transcendent, that is, something of which

[174] On the *in abstracto/in concreto* distinction, see Callanan (ms). Callanan claims that 'in concreto' refers to autological representation, where the representation is a token of the type it expresses (such as 'black' written in black ink, or a geometrical diagram of a triangle). I am not sure how well this account works in the moral case (see Chapter 4 for discussion of moral application 'in concreto'), and Callanan does not address the distinction between *in concreto* and *in individuo*, but we might think of *in individuo* as akin to the application of a proper name, a particular which can never be generalized.

> we cannot even determine the concept sufficiently or of which it is uncertain whether there is any object corresponding to it at all, as is the case with the ideas of speculative reason; instead, the moral ideas, as archetypes of practical perfection, serve as the indispensable rule of moral conduct and also as the *standard of comparison*. (5:127n)

While in the first *Critique* Kant associated ideas with rules and ideals with archetypes and standards,[175] in the second *Critique* it is moral ideas which are 'archetypes' serving as the 'standard of comparison'—while nevertheless also giving the 'rule' of conduct.

As I have argued, Kant's account of moral ideas bears fruitful comparison with the earlier conception of regulative ideas, and are therefore better characterized as universals (working conceptions or 'attempts' at understanding notions ultimately unavailable to cognitive representation) than particulars. Yet this is not to say that the 'archetype' model fails to play any role in Kant's account.

This brings us back to Massimi's (2017) distinction between idea-rules and idea-archetypes. While Massimi argues that a regulative idea should be understood as a rule rather than an archetype, I suggested that Kant's account is more nuanced than this. Although a given moral idea is 'unattainable in practice', the *attempted* realization of it—its instantiation in experience—is what the idea normatively impels us to achieve. The idea, then, functions like an asymptotic limit, conferring on us the prescription to approach it, even if it can never ultimately be reached. On this interpretation, a given moral exemplar (or archetype) would provide us with a partial, necessarily incomplete empirical representation of a moral idea, which would help to *broaden* or *reshape* our working conception of the idea, allowing us to flesh out its description. A full consideration of Kant's account of moral ideas needs to account for both the role of maxims

[175] Cf. *supra*: "As the idea gives the *rule*, so the ideal in such a case serves as the *archetype* for the complete determination of the copy" (A569/B597).

and archetypes, for both the attempted *prescriptive* and *descriptive* forms taken by our endeavors to grasp moral ideas.

Bootstrapping

The dual-aspect view I have developed emphasizes the interrelation *between* moral idea and moral principle, one perhaps best understood as the 'bootstrapping' relation I alluded to in Chapter 1. As Barbara Herman characterizes bootstrapping arguments, "The attraction of bootstrapping is that you use a little bit of what you already have to get someplace you haven't been before, but need to go.... The bootstrapping move is prompted when further thought about what the basic subject is like reveals complexity... that outstrips the carrying capacity of the first explanation" (2007: 154).[176] Recall the examples of moral revision I referred to in Chapter 3: a restricted conception of, say, moral personhood (or, what Kant also refers to as the moral idea of 'humanity') in terms of race, sex, or species exhibits openness to revision on the basis of empirical reflection—reflection itself *guided by* our initial, working moral notions.

Kant's conception of moral ideas accommodates the critical attention to—and possibility of revision in—the formulation of moral principle exhibited in these examples. A moral idea, Kant tells us, such as 'justice' or 'equality', can only *partially* be determined by any given 'ought', by any given particular maxim of action. Moral work is required not to *apply* a moral idea (how

[176] Herman invokes the notion somewhat differently than I do; her reference point is to the bootstrapping *arguments* of Frankfurt's account of reasons and Korsgaard's conception of the will—arguments that begin from the more basic building blocks of, say, desire and belief. Herman, in contrast, claims that Kant does not invoke bootstrapping arguments at all, but holds a metaphysical conception of the will through and through. My reference, by contrast, is to the bootstrapping *process* of moral ideas, which occurs at a later stage. Thus the term 'bootstrapping' is not employed in any technical sense of a given mode of argument but to a commonsensical notion of a self-sustaining process that uses available resources to achieve greater sophistication or complexity.

does one 'apply' justice or virtue?), but to attempt to *approximate* it in our working definitions of these notions, thereby rendering these concepts more *distinct* to us—as, for example, in extending 'humanity' beyond a racially restricted conception, or 'duty' from an indirect duty to a direct duty to non-human animals. We 'bootstrap', then, between our subjective working sense of a given moral idea and the maxims we employ in action. As a given maxim is revised, this renders the idea it approximates more distinct or better developed—which may in turn render us able again to revisit our initial maxim, and so on.

The same, I think, applies to moral exemplars as empirical representations *in individuo* of moral ideas: they demonstrate to us *what it is like* to form maxims under the guidance of moral ideas and to successfully act on them. In either case, we bootstrap between *a priori* concepts to the empirical and back again. Neither the concept (our working definition of the moral idea) nor the empirical representation (the moral exemplar or ideal, when we observe it in others; or the maxim we act on when we attempt to realize the idea ourselves) is understood to be complete, determinate, or conclusive, but each can help to variously 'develop', 'determine', or 'guide' the other.

Interestingly, the same can be said for Kant's view of the role of moral ideas in aesthetic judgment. As we saw in Chapter 3, the connection of an aesthetic object to moral ideas results, for aesthetic judgment as well, in the formation of new concepts: the "universally communicable" pleasure beauty occasions in the viewer "requires a faculty for apprehending the rapidly passing play of the imagination and *unifying it into a concept* (which for that very reason is original and at the same time discloses a *new rule*, which could not have been deduced from any antecedent principles or examples)" (5:317, my emphasis). And "the satisfaction in the beautiful must depend upon reflection on an object that *thereby leads to some sort of concept* (it is indeterminate which)"

(5:207, my emphasis). While we might assume that these new concepts are solely aesthetic concepts of various sorts (picking out, perhaps, aspects of form, color, and method), Kant's insistence on the connection with moral ideas shows that he must also have *moral* concepts, in particular thick moral concepts, as Guyer (1979) notes, in mind.[177] A similar move is employed here, bootstrapping from the empirical to the *a priori* and back.

From this perspective, we can begin to see why it is not just the descriptive aspect of moral ideas which admit of alteration, but also their prescriptive aspect: p-maxims. In TP Kant claims that when theories are incomplete, they can be "perfected only by future experiments and experiences from which… [the theorist] can and ought to *abstract new rules* for himself to complete his theory" (8:275, my italics). In IaG, Kant invokes a similar account of *moral* rules, referring to the transformation over time of the "natural capacity for moral discrimination into *definite practical principles* [*bestimmte praktische Prinzipien*]"; similarly, in KrV, Kant refers to a future time "when *true principles have been developed*, and have become part of our way of thought" (8:21; A747/B775, my italics). Here again, Kant invokes a bootstrapping relation between the empirical and the prescriptive *a priori* aspects of moral reasoning.

This evidence serves to show that the process of working out the implications of and better specifying our moral notions is itself an integral part of moral reasoning. Since this process

[177] Bernard Williams (1985) initially characterized thick concepts as bearers of practical knowledge for a given historically and geographically situated social group: as both action-guiding and world-guided, thick concepts both track the truth for a given society, and give agents within it reasons for action. Williams' thesis was that ethical reflection, in breaking up such marriages of fact and value, destroyed thick concepts and replaced them with 'thin' ones ('right', 'good'; concepts which are bearers of value alone, with no empirical content)—in the process destroying the moral knowledge bound up in thick concepts. The considerations I have been raising, however, show that reflection—here understood as *reflective judgment*—can also create *new* thick concepts ('sexual harassment', 'the problem that has no name', 'gentrification') just as easily as it can destroy them.

amounts to critical revision of our guiding universals (namely, our moral ideas), it amounts to the employment of the reflective dimension of practical reasoning, resulting in the "progress of [our] faculty of judgment" (KpV5:154).

As such, the parallel 'transition' Kant sought between the metaphysical foundations of science and morality comes more sharply into focus. Friedman (1992) terms the Transition Project of OP "paradoxical", insofar as it takes the form of an "intersection between the constitutive and regulative domains—the demand for principles that are, at the same time, both constitutive and regulative", that proceed both top-down (constitutively), from the *a priori* to the empirical, and bottom-up (regulatively), from the empirical to the *a priori* (262). Emundts (2004), however, insightfully suggests the metaphysical foundations of science, as they are increasingly specified (thus rendered increasingly empirical), must admit of revision, and therefore must always entail a *reflective* dimension of evaluation (203). That this is the case can be shown by the fact that we no longer employ terms such as 'ether' or 'pure water' in science—such terms have been revised out of our scientific repertoire. Zuckert (2017) claims that this reflective capacity is precisely what is called for in schematizing regulative principles of reason. Since a fit between universal and particular can here never be secured, but only 'deferred', the ultimate task of the schema is to "prevent empirical complacency", to ward off any sense that our scientific concepts hold categorically: "One also needs to think that these concepts might be wrong, might require (even drastic) revision", just as mechanical conceptions of matter replaced Aristotelian kinds (101). If there is to be a parallel 'transition' in the respective theoretical and moral domains, the same reflective capacity, the same capacity to revise, that functions in altering scientific schemes is at work in our *normative* conceptual repertoire.

5.5 Virtue as meta-idea

The dual-aspect prescriptive and descriptive character of moral ideas, as I have characterized it, bears a close resemblance to discussions of virtue in the Aristotelian tradition as involving *both* a virtue or an archetype *in individuo* (a morally 'thick' concept), *and* a rule, or a 'v-rule', as Hursthouse (1999) puts it. Since each virtue can be enumerated in terms of a thick concept, such as honesty, charity, fidelity (and each vice as, say, irresponsibility, fecklessness, intolerance), they also each generate a prescription or a prohibition, a 'v-rule': namely, "do what is honest, charitable, generous", and so forth (Hursthouse 1999: 29; see also discussion in Annas 2016).

The account I have advanced of Kantian moral ideas is similar in certain respects, broaching morally relevant thick concepts, the subjective formulation of maxims, and the standards set by particular morally exemplary individuals or behaviors. Indeed, Kant's own account of virtue must be situated here, as a kind of *meta-idea*: as picking out our continual attempt to *grasp* moral ideas. The "holiness of will" as a "practical idea", for example, serves "as a model to which all finite rational beings can only approximate without end" (5:32). The same passage continues, "The utmost that finite practical reason can effect is to be certain of this *infinite progress of one's maxims and of their steady disposition to advance*, that is, *virtue* [...*ins Unendliche gehenden Progressus seiner Maximen und Unwandelbarkeit derselben zum beständigen Fortschreiten sicher zu sein, d.i. Tugend*]; and virtue itself, in turn, at least as a naturally acquired ability, can never be completed" (5:32-3; my emphasis). Virtue, then, takes the form of our subjective striving to realize moral ideas, which consists, in turn, in the progressive construction and revision of maxims: the 'indefinite progress of... maxims'.

Kant reiterates this characterization of virtue in TL: "Virtue is always *in progress*...

because, considered *objectively*, it is an ideal and unattainable, while yet constant approximation to it is a duty.… Virtue can never settle down in peace and quiet with its maxims adopted once and for all […*die Tugend mit ihren einmal für allemal genommenen Maximen niemals sich in Ruhe und Stillstand setzen kann*]" (6:409). Again, Kant claims that virtue involves continual 'progress' in the maxims we are in a position to construct. To say that virtue is 'in progress' is to suggest a moral role for the continual *perfecting* of our maxims, subjecting them to redescription as our sense of the morally relevant features at stake evolves. By contrast, to 'settle down in peace and quiet' with our 'maxims adopted once and for all' amounts to skipping over important moral work: that of critically examining our moral conceptual repertoire and scrutinizing where it admits of redescription.[178]

Virtue and moral revision

What I have described characterizes Kant as a moral theorist who gives pride of place to moral concepts as well as moral rules, thereby aligning him in certain key respects with the virtue ethical tradition and complicating the deontology-virtue distinction. This should not necessarily come as a surprise to Kantian moral philosophers, given the importance Kant himself gives to the notion of virtue in the appropriately named Doctrine of Virtue, as well as the role accorded to the broader conception of moral ideas I have attempted to elucidate here.

But, in certain respects, it is actually in Kant's *own* account of virtue that his most important differences with the virtue-ethical tradition may be glimpsed. Because, indeed, part of what the latter tradition misses in its account of the virtues is the asymptotic role moral ideas

[178] Here, Murdoch's insistence that descriptions are value-laden is helpful: "It is important to remember that language itself is a moral medium, almost all uses of language convey value…. If we attempted to describe this room our descriptions would naturally carry all sorts of values" (Murdoch 1977: 27).

play. Since any given moral idea is ultimately 'unattainable', *there is no* ultimate 'standard of comparison': there are only *partial instantiations* of moral ideas which help us, in bootstrapping fashion, to revise our working conceptions of those ideas. This aspect of Kant's view is, as I understand it, radically different from Aristotle's (on certain prominent interpretations of it), on which a given virtue (or thick moral concept; *moral idea*) is *only given to us by* the standard set by a given virtuous individual. What distinguishes the two conceptions is the place perpetually left open for *moral revision.*[179] For Kant, then, virtue *is itself* the striving towards realization of moral ideas both in action and thought, rather than the realization itself, since this last remains 'unattainable'. The process of attaining to virtue therefore remains perpetually incomplete and never done, requiring, most fundamentally, an attitude of *epistemic humility* on the part of the virtuous agent.

While this difference between the two conceptions may appear slight, its implications are ultimately crucial. After all, virtue is typically defined in terms of *character*, or a permanent state of an agent: as a "lasting feature of a person, a tendency for the person to be a certain way"; as a "*reliable* disposition" (Annas 2011). Kant's conception of virtue as moral idea undercuts the easy identification of virtue with stable character. For what is 'stable' in Kantian virtue is an agent's openness to *revise* her character: to revise her ready response to moral features; to revise her normative resources—e.g., her moral conceptual repertoire, her working conceptions of moral ideas—in doing so; to revise her own sense of what she is doing. What 'persists' or 'lasts' in a virtuous state of the Kantian agent is her propensity for self-critique. There is therefore a

[179] Russell (2019) puts the point, to my mind, correctly: "For Kant, while virtue is an ideal to strive for, virtue is not the perfection of practical reason but is rather the attitude in the struggle towards such perfection" (60), where perfection, Kant tells us, "is *not attainable* by any creature, but is yet the archetype which we should strive to approach and resemble in an uninterrupted but endless progress" (5:83).

fundamental *instability* at the quick of Kant's conception of virtue: namely, in the association of virtue and critique

Conclusion: Virtue as Critique

A metanormative stance on normative revisability

On one of the best-known accounts of Kantian 'self-making', constituting oneself as a moral agent requires, not the capacity to revise one's own character, but the activity of *integrating* one's character, assimilating roles into a consistent identity and life narrative: "The task of self-constitution involves finding some roles and fulfilling them with integrity and dedication. It also involves integrating those roles into a single identity, into a coherent life. People are more or less successful at constituting their identities as unified agents, and a good action is one that does this well" (Korsgaard 2009: 25).[180] The operative ideal here is not self-critique, but consistency and authenticity: normative stability rather than normative change.

Tiisala (2016) astutely remarks that parsing the norm of 'coherence' in terms of consistency is only one way in which it might be understood: "The unity of a self can be achieved just as well by radically *changing* one's commitments, by *abandoning* a set of commitments and with them *repudiating* the old self whose commitments they were" (155). In parsing coherence as self-consistency or self-permanence (where it may, just as easily, lend itself to self-alteration), we must recognize that we are stipulating a further substantive normative judgment. Once we begin to recognize how our moral repertoire may, in certain instances, prove inadequate, may fail to accommodate new particulars, we can begin to see that this further normative judgment may not

[180] See also: "Such identities are the sources of our reasons, but of course the idea is not just that we decide which ones we want and conform to them. We have many particular practical identities and so we also face the task of uniting them into a coherent whole" (2009: 21).

be warranted.

Where we encounter a resistant particular, we can retain self-consistency by attempting to assimilate the discordant element in terms of our preexisting normative commitments; if this fails, it can be rejected altogether (Tiisala 2016: 154). Korsgaard might be taken to advocate an approach along these lines when she claims that "an action chosen in a way that more successfully unifies and integrates its agent is more authentically, more fully, an action, than one that does not", identifying 'unity' with 'integrity' and 'authenticity' (2009: 25). Yet, as I have argued, in such instances, the ideal of authenticity or self-consistency may very well be at odds with the best way of accommodating the incompatible particular: rather than reject it as morally unintelligible, our preexisting normative commitments might require reexamination. There are instances, then, in which what may be called for is to assimilate moral content at *odds* with our prior conception of ourselves, rather than only embracing what is in continuity with our earlier self-conception.

This way of conceiving of virtue can be brought to bear on what I called 'reflective constructivism' in Chapter 1. There I claimed that the account of the moral revisability of moral universals I would develop by emphasizing the activity of reflective judgment in practical reasoning was compatible with constructivist views about value. Values are 'constructed' from the evaluative standpoint of valuers (including, but not limited to, human reasoners). This standpoint is one structured not just by basic preferences such as avoiding suffering or treating humanity as an end in itself, but, given that we live and are trained as reasoners in complexly ordered communities with intricate histories, also by the many complexities of our social world. Or in other words, the 'ideologies', as some might put it, the false empirical facts in terms of which we often understand our social reality (given the broader social forces whose interests have helped to structure it), also inevitably infuse our evaluative standpoint. In altering our judgments

of the empirical facts that pertain to our value system (to our judgments of rationality, humanity, or justice), so, too, can our values change. I argued in Chapter 1 that this task—namely, altering the empirical facts that bear on our value judgments—is one central to genealogical investigation in the continental tradition.

This way of understanding the nature of value raises questions about how to construe the metanormative attitude we should take to our normative commitments.[181] Values are, on this view, simultaneously *constructed* by us, but not always in ways fully *transparent* to us as individuals. What view should we take to this fact? Does it follow, for instance, that we should *make* our own normative commitments fully transparent to ourselves, if aspects of them remain obscured?

Against Korsgaard, Street argues that there *is* no universal constructivist norm—aside from, implicitly, coherence, the norm presupposed by the process of a normative judgment 'withstanding scrutiny'.[182] Against these two pictures of constructivism (one universalist, one anti-universalist), I have instead argued for a *revisable universalism*, one guided both by practical reason and by an openness to question the deliverances of judgment: two roles represented by the moral and reflective dimensions of practical deliberation, respectively.

On this picture, the relevant ideal for reflective judgment is not permanence, consistency, or authenticity, but *critique*, albeit understood in an idiosyncratic sense. Indeed, one way of understanding the sense of 'critique' I have in mind is in terms of the account of virtue I

---

[181] See also discussion in Allen (2016).
[182] "For one normative judgment to withstand scrutiny from the standpoint of other normative judgments… is for that judgment not to be mistaken as determined by the standards of correctness that are constitutively set by those other normative judgments in combination with the non-normative facts" (Street 2008: 230). For a normative judgment to withstand scrutiny, it must, then, *cohere* with all the other normative judgments.

presented in the previous chapter. There I claimed that Kant's elaboration of virtue as moral idea contrasted in crucial respects with certain prominent Aristotelian conceptions of virtue. Yet here, I want to stress that Kantian virtue as moral idea, as involving the "infinite progress in one's maxims" and the striving towards a "perfection that we can always approach but never completely attain" (5:32, 7:200), also contrasts with many self-professedly *Kantian* construals of virtue.

On the conventional view, Kant defines virtue as strength of the will against inclination. As a recent paper on Kant's conception of intellectual virtue asserts, "The claim that [Kantian] virtue is strength is relatively uncontroversial" (Thomason 2020: 387). Yet scholars have directed almost no systematic attention to Kant's definition of virtue in the same text as the 'progress' of maxim formation (5:32). Even when the passages regarding virtue as progress are cited, they are usually interpreted as a specification of the claim that virtue is, primarily, strength of will. Another recent account citing a passage from MS identifying virtue as progress (6:409) has it: "Kant argues that 'virtue is always in progress, but also always begins from scratch'... This claim has a two-fold basis. First, finite agents are constantly susceptible to temptations to deviate from moral laws. Second... since our attempts to ensure that we reliably act from good maxims cannot count on habituated volitional mechanisms, such attempts involve a struggle against temptation that arises anew in every particular case" (Kohl 2017: 659-60). Yet, here again, the same passage at 6:409 continues by identifying virtue with the *revision of maxims*, not merely with resisting inclination: "Virtue can never settle down in peace and quiet with its maxims adopted once and for all" (6:409).[183] Virtue, in these passages, is associated with a willingness to subject moral

[183] The full citation is the following, which does attribute causality to 'inclination' : "Virtue is always *in progress* and yet always starts *from the beginning*. It is always in progress because, considered *objectively*, it is an ideal and unattainable, while yet constant approximation to it is a duty. That it always starts from the beginning has a *subjective* basis in human nature, which is affected by inclinations because of which virtue

principles (i-maxims, p-maxims) to critique.

Because virtue is almost overwhelmingly understood in terms of consistency of character in both the Kantian and Aristotelian traditions, not much attention has been paid to the 'critical' aspect of virtue, so understood. As I pointed out in the previous chapter, positioning virtue *as critique* has the somewhat paradoxical result that what is here laudibly situated as a persisting, enduring trait of one's character is precisely what tends to *undercut* the current constitution of one's own character. The virtue of critique has a perpetual instability built into it in a way that the more traditional virtues arguably do not.

The end of history

Famous critics of the very notion of morality can nevertheless be situated as practitioners of virtue as critique. Marx and Engels, for instance, famously rejected morality as empty: moral theory and moral talk falsely represent the interests of the dominant class as the "common interest of all members of society", and thereby only have the "*form* of universality", taking its own, class-specific moral concepts as "the only rational, universally valid ones" (1846: 68, my emphasis). Marx and Engel's critique of morality is extended to Kant's own moral philosophy: "Kant was satisfied with 'good will' alone, even if it remained entirely without result, and he

---

can never settle down in peace and quiet with its maxims adopted once and for all but, if it is not rising, is unavoidably sinking. For moral maxims, unlike technical ones, cannot be based on habit (since this belongs to the natural constitution of the will's determination); on the contrary, if the practice of virtue were to become a habit the subject would suffer loss to that *freedom* in adopting his maxims which distinguishes an action done from duty" (6:409). Nevertheless, the kind of inclination Kant has in mind is further identified with 'habit', which is not necessarily a form of desire, or 'sensuous'; see, e.g., 16:403, 16:409, 24:16. And regardless of one's views regarding the prominence of inclination here, Kant still posits here that virtue must still involve a critical attention *to one's adopted maxims*, rather than just to the inclination to act in ways which frustrate them.

transferred the *realization* of this good will, the harmony between it and the needs and impulses of individuals, to *the world beyond*" (208). Moreover, they claim, a 'good will' is not an *a priori* moral notion at all; it is merely the expression of the less-developed stage of the forces of production of 18th century Prussia as compared to the already industrialized British and French empires.[184]

Yet Marx and Engels employ normative notions themselves in criticizing the social order which has produced these 'impotent' moral concepts. They champion a world in which universality can be *realized*, in which individuals can live a "many-sided life" embracing a "wide circle of varied activities and practical relations to the world", in which "thought has the same character of universality as every other manifestation of his life" (1846: 281). In the current social order, by contrast,[185] they claim that "it is evident that individuals undoubtedly make *one another,* physically and mentally, but do not make themselves"—they lack the capacity for self-determination, for freedom, for autonomy (59). They criticize the fact that the scope of universality is artificially restricted to the domain of abstract rights,[186] with no attention given to the many ways in which labor under capitalism, by its very nature, curtails individual freedom. The nature of Marx and Engels' critique of morality, then, is not so much a *rejection* of morality

[184] "While the French bourgeoisie, by means of the most colossal revolution that history has ever known, was achieving domination and conquering the Continent of Europe, while the already politically emancipated English bourgeoisie was revolutionising industry and subjugating India politically, and all the rest of the world commercially, the impotent German bourgeois did not get any further than 'good will'…. Kant's good will fully corresponds to the impotence, depression and wretchedness of the German bourgeois" (208).

[185] Namely one in which "each man has a particular, exclusive sphere of activity, which is forced upon him and from which he cannot escape. He is a hunter, a fisherman, a shepherd, or a critical critic, and must remain so if he does not want to lose his means of livelihood" (53).

[186] See also Marx's critique of human rights in "On the Jewish Question". The right to 'liberty', for instance, is "the right to do everything which does not harm others…. It is a question of the liberty of man regarded as an isolated monad, withdrawn into himself" (1844: 42).

as the *reflective alteration* of moral concepts (universality, autonomy, freedom) to accommodate particulars which cannot be subordinated to the available moral repertoire, given the latter's currently restricted extension to dominant class interests alone. They push on us the revision of our maxims, to strive towards their 'infinite progress', and thereby exemplify the critical dimension of Kantian virtue I have tried to bring out.

Why not think, then, with Marx and Engels, that morality is ultimately dispensable?[187] They advocate a social order in which individuals would be able to behave 'egoistically', acting on their own interests, rather than needing to bother about the moral repercussions of their behaviors, since, in a communist society, egoistic and moral ends would naturally align. If such a social order would *necessarily* come about with the historical development of the means of production, the only problem would be to dispel the false consciousness of the proletariat that left them deluded about their true interests. Once workers attained an accurate view of their situation, the activity required to overthrow capitalism would necessarily follow, since it would align with workers' own self-interest.

Of course, as of yet, history has not taken the deterministic shape Marx and Engels thought it would. And, absent their historical teleology, their premature rejection of 'bourgeois morality' also needs to be revisited. For the time being at least, we will need to retain an inventory of normative concepts to assess the social world. And this state of affairs makes the exercise of a *critical* attitude toward those concepts (as well as to the social reality to which they are applied) all the more important.

[187] "The communists do not preach morality at all…. They do not put to people the moral demand: love one another, do not be egoists, etc.; on the contrary, they are very well aware that egoism, just as much as selflessness, is in definite circumstances a necessary form of the self-assertion of individuals" (1846: 264).

Given these premises, we can begin to see what motivated another strand of the continental tradition to characterize critique in *ethical* terms, specifically as a virtue—and, reciprocally, to understand moral actualization in terms of critique (Foucault 1978: 35).[188] As Butler (2002) defines it, "Virtue is most often understood [as] a quality that conditions and characterizes a certain kind of action or practice. It belongs to an ethics which is not fulfilled merely by following objectively formulated rules or laws. And virtue is not only a *way* of complying with or conforming with preestablished norms. It is, more radically, a critical relation to those norms" (215). If we reject the faith in the arrival of a social order in which each individual's egoistic satisfaction of her own desires will harmonize with everyone else's, we must reassert the importance of *moral* deliberation—albeit one reconceived in terms of critique.

Moral regression

In this dissertation, I have admittedly often made it sound as if the exercise of the reflective dimension necessarily results in forms of moral progress. I have characterized maxims as admitting of systematicity or hierarchical organization, where increasingly determined and particularized maxims can be ordered under our most general moral principles, and I have spoken of moral 'revision' and of the 'extension' in scope, or depth, of our moral repertoire, as well as the 'bootstrapping' relation between moral concepts and principles.

[188] "A more general imperative underlies critique—more general than that of dispelling error. There is something in critique similar to a virtue. And, in a certain way, what I want to speak of is this critical attitude as a virtue in general" (Foucault 1978: 34-35). Foucault goes on to characterize this virtue in various ways: "desubjugation [*désassujetissement*]" (39), the "art of not being so governed, or not governed at such a price" (37), the "art of voluntary insubordination" or of "reflective indocility" (Foucault 1978: 39). See also references to *parrhesia* as a critical attitude and as a virtue (Foucault 1983: 24, 63, 68; 1984d: 29, 283; 1984b: 24, 322).

Such talk makes it sound as if moral alteration only goes one way. It doesn't, of course. The reasons why it doesn't are more complex than I can fully defend in the space of a conclusion. To put my position succinctly: we should not presuppose moral progress because reflection takes current social conditions as its object, which need not be gradually improving—and often are not in anything but a partial and fragmentary way. Indeed, many social facts have been constructed to some degree 'behind our backs', such that the content of social practices is not always fully transparent to us. We may never be able to fully ascertain the implications of altering social arrangements, in part because it is so difficult to fully get their current functioning into view; changing one aspect may itself pose a new problem that then, once again, requires reflection, such that the employment of the reflective dimension will remain perpetually necessary.

To elaborate a bit more on the two considerations raised in the above characterization of my view:

First, as I also noted in Chapter 1, the collective agency at work in the making of our social world is not equally distributed; certain material interests are accorded more expression than others, and the forces at work in shaping the social order (and, with it, the normative concepts brought to bear on it) are not symmetrical in their effects. We can think here of the 'normalization' of unjust states of affairs—where, instead of being rectified, they merely become ingrained as part of the fabric of reality. In the last century, the rise in deaths caused by automobile accidents (then newly invented) were 'normalized' in part through the persistent efforts of the automobile lobby;[189] a similar process is underway with regard to the normalization

[189] This led to the introduction of laws against jaywalking and the transformation of roads from public space into areas exclusively for cars; the onus was put on individuals to *avoid* cars, rather than on car manufacturers or urban planners to restrict car access (Norton 2007).

of the coronavirus epidemic at the time of this writing.

Second, the effects of certain 'revisions' can be unpredictable. The 19th century normative conclusion that the mentally ill required institutionalization created new problems, and with them, the occasion for a new critique, resulting, in the mid-20th century, in the release of patients from asylums back into the community—occasioning yet another critique in turn.[190]

These examples are each somewhat mundane, but they serve to demonstrate, first, that instances of moral *regression* do occur (as judged, of course, from within our evaluative perspective), and second, that the reflective dimension can also be actualized in identifying them. As Wendy Brown claims, this is an important function of genealogical critique.[191] Most prominently, perhaps, moral regression is a persistent theme in Adorno's writings, which he characterizes at turns as the erosion of the 'good universal' or the 'shrinking of morality' (1951: §11, §99).[192] While the apparently 'regressive' aspects of such critiques have often been taken as evidence of underlying normative confusion, if we construe them as operations of reflective judgment within practical reason, we can see how identifying instances of regression is necessary in order to be able to revise our normative repertoire in (what we take to be) *better* ways going forward. Thus, reflective judgment is always employed *in the service of* (what we understand to be) moral progress, even where it identifies the opposite taking place. Yet the results of its exercise are

[190] The sociological literature on 'moral panics' is instructive here: new causes for moral concern can be identified which are merely expressions of collective anxiety, or whose 'solutions' fail, or create problems worse than the originally identified one. The American temperance movement of the 1920s is one such example (Gusfield 1963).
[191] "Genealogy also conjures a reversal in the course of history—it challenges progressive accounts with intimation of regression, as it suggests that the present may be 'living at the expense of the future' rather than paving the way to that future" (2001: 98).
[192] "Neither Timur nor Genghis Khan nor the English colonial administration in India systematically burst the lungs of millions of people with gas[;] the eternity of horror… manifests itself in the fact that each of its new forms outdoes the old" (1951: §149).

limited in scope to merely *local* forms of progress, ones always vulnerable to reversal as events continue to unfold in the social world—and as the meaning of our behavior is sometimes reinterpreted by others (given a new 'legibility'). Employing reflective judgment should therefore involve a healthy dose of epistemic humility.

Systematicity and intelligibility

As I discussed in Chapter 5, Kant claims that ideas of reason serve to guide reason's demand for systematicity: in other words, they help to order and unify cognitions into a system of knowledge. While this role has been widely recognized in interpretations of Kant's view of theoretical reasoning, the way I construed moral ideas and moral principles gave similar emphasis to systematicity in *moral* reasoning—which can make it sound as if maxims are being fitted into their fated slots, like pieces of a puzzle clicking into place.

Ido Geiger (2003) suggests that this need not be the only way in which systematicity is understood. We can also understand the ideal of systematicity as the ideal of *intelligibility*: "The ideal of a systematic unity of knowledge represents the world of objects as fully intelligible", as objects that can be exhaustively comprehended in discursive terms (279). This ideal is consistent with the transcendental maxim of teleological judgment, namely the principle of purposiveness (Chapter 3): to aim for the ideal of the systematic unity of knowledge, we must presuppose that the world is *fit to be known*, that it is purposive for human cognition (A686-7/B714-5); "but to be purposive for human cognition is to be intelligible" (Geiger 2003: 279). Geiger points out that the model of systematic unity moves from the idea of a hierarchical taxonomy of species and genera in the Introduction to KU to the notion of *organic* unity in the body of the Critique of Teleological Judgment—a model, that is, that "significantly… does not posit a hierarchical

taxonomy of concepts" (293).[193] And this is even more evident in the case of aesthetic judgment, whose role Kant explicitly characterizes as making "intelligible" new concepts or rules (5:314, 5:207, 5:317).[194] The demand for systematicity increasingly comes to take the form of a demand for *intelligibility*.

In KU Kant describes a process by which we continuously (re)make our judgment in light of how best to *make sense* of the world. Kant characterizes the maxim of purposiveness as how "the power of judgment… *must think* of nature", representing "the unique way in which we *must proceed* in reflection on the objects of nature with the aim of a thoroughly interconnected experience" (5:184). This maxim, as we saw in Chapter 3, is nonetheless 'revised', and lent support, by the maxims of internal and external purposiveness, respectively. Kant emphasizes that the "unity of nature" thought through reflective judgment is "contingent", in that it cannot be known *a priori*, but can always be revised by new empirical features; nevertheless, it "must still necessarily be presupposed and assumed" (5:183).

The difference between systematicity and intelligibility is relevant to the question of how to treat obdurate particulars, known, in the domain of science, as counter-examples or anomalies. If our governing ideal is *systematicity*, a counter-example to our system risks simply being

[193] Geiger does not lend support to this claim, and I find it somewhat incomplete; as I discussed in Chapter 3, Kant continues to lend credence to the ideal of the hierarchical ordering of natural purposes or 'links in the chain' of nature, or what he terms 'external purposiveness' (§3.4). Nevertheless, it is true that Kant emphasizes, more than ever before perhaps, the provisionality of the claim to systematicity here, insisting that a 'chain of nature' can only be posited once internal purposiveness, or the functional explanation of natural organisms, has been epistemically vindicated (5:378).

[194] Though we should be careful here, since Kant associates the 'intelligible' both with the linguistic and with the noumenal (or 'intellectual'); a purely discursive intellect would also be a *divine* intellect. See A249, where he terms "noumena" "*Intelligibilia*", or refers to "*mundi sensibilis und intelligibilis*" (A256). In the KU passages cited here, however, Kant refers to what is '*unverständlich*' or '*unnennbar*' (incomprehensible or unnameable), which is much closer to the contemporary meaning of intelligibility, as well as how I employ the term in Chapter 2.

discarded, if admitting it would threaten the integrity of our system. But if our ideal is instead *intelligibility*, what comes to take precedence is the comprehension of the resistant particular over the preservation of the system.[195] The ideal of intelligibility presses us to *revise* our system rather than to only train our attention on features that can be successively *added* to it.

When I spoke of 'moral intelligibility' in Chapter 2, I had something like this distinction in mind. Making sense of unintelligible particulars may lead us to *discard* aspects of our moral system, or *revise* firmly held principles embedded deep within it. Some of these alterations may look like something of a regression, even where, ultimately, they result in collective improvement. To understand reflective judgment as aimed at *intelligibility* rather than systematicity, to affirm its operations as promoting normative *instability* rather than securing normative grounding, allows us to make space for such possibilities.

As the reasons articulated here begin to show, the Kantian reflective dimension of practical reason may be thought to provide a genuine alternative to the Hegelian model of immanent critique. For Kant, unlike Hegel, there is nothing necessary about the revisions reflective judgment brings to bear on its own principles. As Herman claims, "The Hegelian distinction between *Moralität* and *Sittlichkeit*… the first marking a set of timeless obligations and rights, the other the obligations that belong to historically specific forms of life… tends to miss the contingency that occurs at all levels of obligation" (2009: 47). Kant perpetually insists that the norms reflective judgment gives itself are in place only to the extent to which they prove *useful* for

[195] The debate over coherentism in epistemology dovetails nicely here: a more coherent or complete system may be less conducive to truth (if, e.g., all of its components are false) than one riddled with internal inconsistencies (if, e.g., some of them are true); see Klein and Warfield (1994, 1996). This problem also faces Street's constructivism, since it also affirms the norm of coherence (Street ms.), but it is all the more pressing if coherence is construed, more strongly, as consistency or permanence over time, as Korsgaard advocates.

cognition. They can never be thought to inhere in the object, but only in directing us in how best to *think* of the object in question.

Kant characterizes critique as a historical process, one specific to his own time: "Our age is, in especial degree, the age of criticism, and to criticism everything must submit" (KrV Axi).[196] Kant, too, leaves open just how historically restricted the need for critique is.[197] In rejecting the end of history, the metanormative attitude of *critique* takes on a heightened relevance. The exercise of critical attention to our working moral universals, to the very construal of universality as such, the cultivation of this reflective activity as a *virtue*—it all becomes all the weightier. If past historical experience is any guide, there are bound to be epistemic lacunae in our working conception of universality, ones which we may never be able to say conclusively are filled. In this regard, the reflective and the moral dimensions can be construed as holding each other in check. We can never be sure the 'leading-strings' are cast off once and for all; lessening their grip will

[196] See also the reference in the same passage to "the *matured judgment* of the age, which refuses to be any longer put off with illusory knowledge" (Axi). In the Discipline of Pure Reason Kant elaborates on the social conditions required for critique. Kant calls against keeping students "under tutelage" until "their faculty of judgment is mature", referring instead to the prospect of "enlightening" youthful reason "through critique" (A754-5/B782-3). The process Kant advocates involves allowing the student "the opportunity of testing for himself, one by one, by reference to the critical principles, how groundless are the assertions of those who have launched these attacks" (A755/B783). In addition, critique is characterized a *collective* endeavor, requiring conditions for *publicity*: "Reason depends on this freedom [of critique] for its very existence. For reason has no dictatorial authority; its verdict is always simply the agreement of free citizens, of whom each one must be permitted to express, without let or hindrance, his objections or even his veto" (A738/B766).
[197] This is brought out, in particular, in Kant's claims regarding enlightenment: "It will require a long, perhaps incalculable series of generations, each passing on its enlightenment to the next, before the germs implanted by nature in our species can be developed to that degree which corresponds to nature's original intention" (8:19); "by a continued process of enlightenment, a beginning is made towards establishing a way of thinking which can with time transform the primitive natural capacity for moral discrimination into definite practical principles" (8:21); "although folly and caprice creep in at times, enlightenment gradually arises" (8:28); "a germ of enlightenment always survived, developing further with each revolution, and prepared the way for a subsequent higher level of improvement" (8:30). It's unclear, given these pronouncements, whether this process will ever reach an endpoint.

always constitute an ongoing task.

References

Adorno, Theodor W. 1951. *Minima Moralia: Reflections from Damaged Life.* Translated by E.F.N. Jephcott. London: Verso.

Adorno, Theodor W. 1963. *Problems of Moral Philosophy.* Edited by Thomas Schröder. Translated by Rodney Livingstone. Stanford: Stanford University Press.

Adorno, Theodor W. and Max Horkheimer. 1944 [2002]. *Dialectic of Enlightenment: Philosophical Fragments.* Edited by Gunzelin Schmid Noerr. Translated by Edmund Jephcott. Stanford: Stanford University Press.

Allais, Lucy. 2016. "Kant's Racism." *Philosophical Papers* 45.1-2: 1-36. DOI: 10.1080/05568641.2016.1199170.

Allen, Amy. 2008. *The Politics of Our Selves: Power, Autonomy, and Gender in Contemporary Critical Theory.* New York: Columbia University Press.

Allen, Amy. 2016. *The End of Progress: Decolonizing the Normative Foundations of Critical Theory.* New York: Columbia University Press.

Allison, Henry E. 1990. *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.

Allison, Henry E. 2001. *Kant's Theory of Taste: A Reading of the Critique of Aesthetic Judgment*. Cambridge: Cambridge University Press.

Ameriks, Karl. 2012. "Vindicating Autonomy." In *Kant on Moral Autonomy*, edited by Oliver Sensen, 53–70. Cambridge: Cambridge University Press.

Annas, Julia. 2011. *Intelligent Virtue*. Oxford: Oxford University Press.

Annas, Julia. 2016. "Learning Virtue Rules: Developing Thick Concepts." In *Developing the Virtues: Integrating Perspectives*. Edited by Julia Annas, Darcia Narvaez, and Nancy E. Snow. New York: Oxford University Press, 2016. DOI: 10.1093/acprof:oso/9780190271466.003.0011.

Anscombe, G.E.M. 1956 [1981]. *Mr. Truman's Degree.* Privately published pamphlet. Reprinted in *The Collected Philosophical Papers of G.E.M. Anscombe*, volume III: *Ethics, Religion and Politics*. Oxford: Blackwell, 62-71.

Anscombe, G.E.M. 1957 [2000]. *Intention*. Second edition. Cambridge, Mass: Harvard University Press.

Anscombe, G.E.M. 1958 [1981]. "Modern Moral Philosophy." Reprinted in *The Collected*

Philosophical Papers of G.E.M. Anscombe, volume III: *Ethics, Religion and Politics*. Blackwell, Oxford: 26-42.

Anscombe, G.E.M. 1963 [1981]. "Two Kinds of Error in Action." Reprinted in *The Collected Philosophical Papers of G.E.M. Anscombe*, volume III: *Ethics, Religion and Politics*. Blackwell, Oxford: 3-9.

Arendt, Hannah. 1963. *Eichmann in Jerusalem: A Report on the Banality of Evil.* New York: Penguin Classics.

Arendt, Hannah. 1966. "Some Questions of Moral Philosophy." *Responsibility and Judgment*. New York: Schocken Books, 49-146.

Arendt, Hannah. 1970. *Lectures on Kant's Political Philosophy.* Chicago: University of Chicago Press.

Arendt, Hannah. 1971. "Thinking and Moral Considerations: A Lecture." *Social Research*, 38:3.

Aristotle. 2001. *Nicomachean Ethics*. In *The Basic Works of Aristotle*. Edited by Richard McKeon. Translated by W.D. Ross. New York: Modern Library.

Arp, Kristana. 2001. *The Bonds of Freedom: Simone de Beauvoir's Existential Ethics*. Chicago: Open Court.

Austin, J.L. 1962 [1975]. *How to Do Things with Words: The William James Lectures Delivered at Harvard University in 1955*. Oxford: Clarendon.

Bagnoli, Carla. 2011a. "The Exploration of Moral Life." In *Iris Murdoch, Philosopher*, edited by Justin Broackes. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199289905.003.0008.

Bagnoli, Carla. 2011b. "Emotions and the Categorical Authority of Moral Reason." *Morality and the Emotions.* Edited by Carla Bagnoli. Oxford: Oxford University Press, 62-81.

Balibar, Étienne. 1994. "Racism as Universalism." *Masses, Classes, and Ideas: Studies on Politics and Philosophy Before and After Marx.* Translated by James Swenson. New York: Routledge.

Bauer, Nancy. 2001. *Simone de Beauvoir, Philosophy, and Feminism*. New York: Columbia University Press.

Beauvoir, Simone de. 1947. *Pour une morale de l'ambiguïté*. Paris: Gallimard.

Beauvoir, Simone de. 1949a. *Le deuxième sexe.* Two volumes. Paris: Gallimard.

Beistegui, Miguel de. 2018. *The Government of Desire: A Genealogy of the Liberal Subject.* Chicago: The University of Chicago Press.

Bell, David. 1987. “The Art of Judgement.” *Mind* 96.382: 221-244.

Bergoffen, Debra. 2008. “Getting the Beauvoir We Deserve.” In *Beauvoir and Sartre: The Riddle of Influence*. Edited by Christine Daigle and Jacob Golomb, 13–29. Bloomington: Indiana University Press.

Bernasconi, Robert. 2001. “Who Invented the Concept of Race? Kant’s Role in the Enlightenment Construction of Race.” In *Race.* Edited by Robert Bernasconi. Malden, MA: Blackwell.

Bittner, Rüdiger. 1974. ‘Maximen.’ *Akten des 4. Internationalen Kant-Kongresses*. Edited by Gerhard Funke. Berlin: Walter de Gruyter, 485–98.

Brandom, Bob. 2007. The Structure of Desire and Recognition: Self-Consciousness and Self-Constitution. *Philosophy & Social Criticism* 33: 127-150. DOI: 10.1177/0191453707071389.

Brown, Wendy. 2001. *Politics out of History.* Princeton: Princeton University Press.

Burgess, Alexis and David Plunkett. 2013a. “Conceptual Ethics I.” *Philosophy Compass* 8.12: 1091–1101, DOI: 10.1111/phc3.12086

Burgess, Alexis and David Plunkett. 2013b. “Conceptual Ethics II.” *Philosophy Compass* 8.12: 1102–1110. DOI: 10.1111/phc3.12085

Butler, Judith. 1986. “Sex and Gender in Simone de Beauvoir’s *Second Sex.*” *Yale French Studies* 72, 35-49.

Butler, Judith. 1989. *Gender Trouble*. New York: Routledge.

Butler, Judith. 1999. “Preface (1999)”. In *Gender Trouble*. New York: Routledge.

Butler, Judith. 1996. “Universality in Culture.” In *For Love of Country?*. By Martha C. Nussbaum. Edited by Joshua Cohen. Boston: Beacon Press, 45-52.

Butler, Judith. 2002. “What Is Critique? An Essay on Foucault’s Virtue.” In *The Political*. Edited by David Ingram. Malden, Mass: Blackwell, 212–26.

Calhoun, Cheshire. 1989. “Responsibility and Reproach.” *Ethics* 99.2: 389-406.

Calhoun, Cheshire. 2016. *Moral Aims: Essays on the Importance of Getting It Right and Practicing Practicing Morality with Others.* Oxford: Oxford University Press.

Callanan, John. Manuscript. “Kant on Signs *in Concreto* in Geometry.” Accessed October 21, 2020.

Cappelen, Herman. 2018. *Fixing Language: An Essay on Conceptual Engineering.* Oxford: Oxford

University Press.

Card, Claudia. 1996. *The Unnatural Lottery: Character and Moral Luck.* Philadelphia: Temple University Press.

Celikates, Robin. 2006. "From Critical Social Theory to a Social Theory of Critique: On the Critique of Ideology after the Pragmatic Turn." *Constellations* 13.1, 21-40.

Celikates, Robin. 2009. *Kritik als soziale Praxis: Gesellschaftliche Selbstverständigung und kritische Theorie.* Frankfurt: Campus Verlag.

Chalmers, David J. 2018. "What is Conceptual Engineering and What Should it Be?" Paper given at NYU workshop on "Foundations of Conceptual Engineering," September 14, 2018. Accessed October 8, 2019. http://consc.net/papers/engineering.pdf

Chignell, Andrew. 2007. "Belief in Kant." *The Philosophical Review* 116.3: 323-360.

Chignell, Andrew. 2008. "Kant on the Normativity of Taste: The Role of Aesthetic Ideas." *Australasian Journal of Philosophy* 85.3: 415-433.

Clark, Maudemarie and David Dudrick. 2007. "Nietzsche and Moral Objectivity: The Development of Nietzsche's Metaethics." *Nietzsche on Ethics and Politics.* Edited by Maudemarie Clark. New York: Oxford University Press.

Cohen, Alix. 2009. *Kant and the human sciences: Biology, anthropology and history*. London: Palgrave Macmillan.

Cohen, Alix. 2014. "Kant on the Ethics of Belief." *Proceedings of the Aristotelian Society* 114: 317-34.

Cohen, Hermann. 1902. *System der Philosophie, Erster Teil: Logik der reinen Erkenntnis*. Berlin: Bruno Cassirer.

Conant, James. 1995. "Reply: Nietzsche, Kierkegaard, and Anscombe on Moral Unintelligibility." In *Religion and Morality.* Edited by D.Z. Phillips. New York: Palgrave MacMillan.

Congdon, Matthew. 2017. "What's Wrong with Epistemic Injustice?: Harm, Vice, Objectification, Misrecognition." *The Routledge Handbook of Epistemic Injustice*. Edited by Ian James Kidd, José Medina, and Galle Pohlhaus, Jr. London: Routledge, 243-253.

Craig, Edward. 1990. *Knowledge and the State of Nature: An Essay in Conceptual Synthesis.* Oxford: Oxford University Press.

Crary, Alice. 2018. "Wittgenstein Goes to Frankfurt (and Finds Something Useful to Say)." *Nordic Wittgenstein Review* 7.1: 7-41. DOI: 10.15845/nwr.v7i1.3493

Curzer, Howard. 2016. "Rules Lurking at the Heart of Aristotle's Virtue Ethics." *Apeiron* 49.1:

57-92.

Dancy, Jonathan. 1993. *Moral Reasons.* Oxford: Blackwell.

Dancy, Jonathan. 2004. *Ethics Without Principles.* Oxford: Oxford University Press.

Danto, Arthur Coleman. 1997. *After the End of Art: Contemporary Art and the Pale of History*. Princeton University Press.

Darwall, Stephen, Allan Gibbard, and Peter Railton. 1992. "Toward *Fin de siècle* Ethics: Some Trends." *The Philosophical Review* 101.1: 115-189.

Daston, Lorraine and Peter Galison. 2007. *Objectivity.* New York: Zone Books.

Davidson, Arnold I. 1997. "Structures and Strategies of Discourse: Remarks Towards a History of Foucault's Philosophy of Language." In *Foucault and His Interlocutors.* Edited by Arnold I. Davidson. Chicago: University of Chicago Press, 1-20.

Davidson, Donald. 1973. "On the Very Idea of Conceptual Scheme." *Proceedings and Addresses of the American Philosophical Association* 47: 5-20.

Descombes, Vincent. 2014. *The Institutions of Meaning: A Defense of Anthropological Holism*. Translated by Stephen Adam Schwartz. Cambridge: Harvard University Press.

di Giovanni, George. 1985. "Introduction: The Facts of Consciousness." In *Between Kant and Hegel: Texts in the Development of Post-Kantian Idealism*, edited by H. S. Harris and George di Giovanni, 2–50. Albany: State University of New York Press.

Diamond, Cora. 1988. "Losing Your Concepts." *Ethics* 98.2: 255-277.

Egan, David. Forthcoming. "Rule Following, Anxiety, and Authenticity." *Mind*. Accessed October 18, 2020. https://philarchive.org/rec/EGARFA-2

Emundts, Dina. 2004. *Kants Übergangskonzeption im Opus Postumum.* Berlin: de Gruyter.

Eze, Emmanuel Chukwudi. 1997. "The Color of Reason: The Idea of 'Race' in Kant's Anthropology." *Postcolonial African Philosophy: A Critical Reader*. Edited by Emmanuel Chukwudi Eze. Oxford: Blackwell, 103-131.

Ferrara, Alessandro. 1998. *Reflective Authenticity: Rethinking the Project of Modernity.* London: Routledge.

Ferrara, Alessandro. 2008. *The Force of the Example: Explorations in the Paradigm of Judgment.* New York: Columbia University Press.

Flikshuh, Katrin. 2013. "Personal Autonomy and Public Authority." In Sensen (2013), 169-190.

Floyd, Juliet. 1998. "Heautonomy: Kant on Reflective Judgment and Systematicity." In *Kants Ästhetik*. Edited by Herman Parret. Berlin: Walter de Gruyter.

Forst, Rainer. 1994 [2002]. *Contexts of Justice: Political Philosophy Beyond Liberalism and Communitarianism.* Translated by John M. Farrell. Berkeley: University of California Press.

Förster, Eckart. 1989. "Kant's *Selbstsetzungslehre*." In *Kant's Transcendental Deductions: The Three Critiques and the Opus Postumum*, edited by Eckart Förster. Stanford University Press.

Förster, Eckart. 1990. "Fichte, Beck and Schelling in Kant's Opus Postumum." In *Kant and His Influence*, edited by G. MacDonald Ross and Tony McWalter, 146–69. Thoemmes Continuum.

Förster, Eckart. 1993. "Kant's Third Critique and the Opus Postumum." *Graduate Faculty Philosophy Journal* 16.2: 345–58. DOI: 10.5840/gfpj199316221.

Förster, Eckart. 2000. *Kant's Final Synthesis: An Essay on the Opus Postumum*. Harvard University Press.

Förster, Eckart. 2012. *The Twenty-Five Years of Philosophy: A Systematic Reconstruction*. Translated by Brady Bowman. Cambridge, Mass: Harvard University Press.

Foucault, Michel. 1964 [2001]. "Nietzche, Freud, Marx." In *Dits et écrits*, volume I, no. 46. Paris: Gallimard, 592-607.

Foucault, Michel. 1971 [2001]. "Nietzsche, la généalogie, l'histoire." In *Dits et écrits*, volume I, no. 84. Paris: Gallimard, 1004-1024.

Foucault, Michel. 1973 [2013]. *La société punitive: Cours au Collège de France*, *1972-1973*. Paris: Seuil/Gallimard.

Foucault, Michel. 1975. *Surveiller et punir: Naissance de la prison.* Paris: Gallimard.

Foucault, Michel. 1976 [1997]. « *Il faut défendre la société* »: *Cours au Collège de France*, *1975-1976*. Paris: Seuil/Gallimard.

Foucault, Michel. 1978 [2015]. "Qu'est-ce que la critique?". In *Qu'est-ce que la critique? Suivie de La culture de soi.* Paris: Vrin.

Foucault, Michel. 1982 [2001]. "Vérité, pouvoir, et soi." In *Dits et écrits*, volume II, no. 362. Paris: Gallimard, 1596-1602.

Foucault, Michel. 1983 [2019]. "Discourse and Truth." In *"Discourse and Truth" and "Parresia".* Chicago: University of Chicago Press.

Foucault, Michel. 1984a [2001]. "What Is Enlightenment?." In *Dits et écrits*, volume II, no. 339.

Paris: Gallimard, 1381-1396.

Foucault, Michel. 1984b [2008]. *Le gouvernement de soi et des autres*: *Cours au Collège de France*, *1982-1983*. Paris: Seuil/Gallimard.

Foucault, Michel. 1984c. "Polemics, Politics, and Problematizations." In *Dits et écrits*, volume II, no. 342. Paris: Gallimard, 1410-1417.

Foucault, Michel. 1984d [2009]. *Le courage de la vérité*: *Cours au Collège de France*, *1984*. Paris: Seuil/Gallimard.

Fraser, Nancy. 1985. "What's Critical about Critical Theory? The Case of Habermas and Gender." *New German Critique* 35: 97-131.

Fricker, Miranda. 2000. "Confidence and Irony." *Morality, Reflection, and Ideology.* Edited by Edward Harcourt. Oxford: Clarendon Press.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Friedan, Betty. 1963. *The Feminine Mystique*. New York: W.W. Norton.

Friedman, Michael. 1991. "Regulative and Constitutive." *Southern Journal of Philosophy* 30: 73-102.

Friedman, Michael. 1992. *Kant and the Exact Sciences*. Cambridge: Harvard University Press.

Gadamer, Hans-Georg. 1960 [2013]. *Truth and Method.* Translated by Joel Weinsheimer and Donald G. Marshall. London: Bloomsbury.

Garcia, Manon. 2018. *On ne naît pas soumise, on le devient*. Paris: Climats.

Geiger, Ido. 2003. "Is the assumption of a systematic whole of empirical concepts a necessary condition of knowledge?" *Kant-Studien* 94.3: 273-298.

Geiger, Ido. Manuscript. "Kant on Aesthetic Ideas, Rational Ideas and the Subject-Matter of Art." Consulted June 10, 2020.

Gendler, Tamar Szabó. 2000. "The puzzle of imaginative resistance." *The Journal of Philosophy* 97.2: 55-81.

Geuss, Raymond. 1981. *The Idea of a Critical Theory: Habermas and the Frankfurt School*. Cambridge: Cambridge University Press.

Geuss, Raymond. 1996. "Morality and Identity." In Korsgaard (1996), 189-199.

Ginsborg, Hannah. 2015. *The Normativity of Nature: Essays on Kant's Critique of Judgment.*

Oxford: Oxford University Press, 15-31.

Glass, Ruth. 1964. "Introduction: Aspects of Change." In *London: Aspects of Change*. Edited by Centre for Urban Studies. London: MacGibbon and Kee, xiii–xlii.

Golder, Ben. 2015. *Foucault and the Politics of Rights.* Stanford: Stanford University Press.

Gooding-Williams, Robert. 2017. "Ideology, Social Practices, Anti-Black Concepts." Paper given at NYU Law and Philosophy Seminar, October 19, 2017. https://www.law.nyu.edu/sites/default/files/upload_documents/Ideology%20Paper3.1%20%28NYU%20Law%20and%20Philosophy%29%20%28002%29.pdfa

Gressis, Rob. 2010a. "Recent Work on Kantian Maxims I: Established Approaches." *Philosophy Compass* 5: 216-227.

Gressis, Rob. 2010b. "Recent Work on Kantian Maxims II." *Philosophy Compass* 5: 228-239.

Grier, Michelle. 2001. *Kant's Doctrine of Transcendental Illusion.* Cambridge University Press.

Gusfield, Joseph R. 1963. *Symbolic Crusade: Status Politics and the American Temperance Movement*. Urbana and Chicago: University of Illinois Press.

Guyer, Paul and Allen Wood. 1998. "Introduction." In Immanuel Kant, *Critique of Pure Reason*. Translated by Paul Guyer and Allen Wood. Cambridge: Cambridge University Press, 1-72.

Guyer, Paul. 1979. *Kant and the Claims of Taste.* Cambridge: Cambridge University Press.

Guyer, Paul. 1990. "Reason and Reflective Judgment: Kant on the Significance of Systematicity." *Noûs* 24.1: 17–43. https://doi.org/10.2307/2215611.

Guyer, Paul. 2004. "Perfection, Autonomy, and Heautonomy: The Path of Reason from Wolff to Kant." In *Wolff und die Europäische Aufklärung: Akten des 1. Internationalen Christian Wolff Kongresses*, 1:299–322. Halle (Saale): Georg Olms Verlag.

Guyer, Paul. 2011. "Kantian Perfectionism." In *Perfecting Virtue: New Essays on Kantian Ethics and Virtue Ethics.* Edited by Lawrence Jost and Julian Wuerth. Cambridge: Cambridge University Press, 194-214.

Guyer, Paul. 2013. "Progress Toward Autonomy." In Sensen (2013): 71–86.

Habermas, Jürgen. 1981a [1984]. *The Theory of Communicative Action.* Volume 1: *Reason and the rationalization of society.* Translated by Thomas McCarthy. Boston: Beacon Press.

Habermas, Jürgen. 1981b [1987]. *The Theory of Communicative Action.* Volume 2: *Lifeworld and System: A Critique of Functionalist Reason.* Translated by Thomas McCarthy. Boston: Beacon Press.

Habermas, Jürgen. 1983 [1990]. *Moral Consciousness and Communicative Action.* Cambridge: Polity.

Habermas, Jürgen. 1990. "Morality and Ethical Life." In *Moral Consciousness and Communicative Action.* Cambridge: Polity.

Hacking, Ian. 1982. "Language, Truth, and Reason." In *Rationality and Relativism.* Edited by Martin Hollis and Steven Lukes. Cambridge: MIT Press.

Hacking, Ian. 1985. "Making Up People." In *Reconstructing Individualism*. Edited by T.L. Heller, M. Sosna and D.E. Wellbery. Stanford: Stanford University Press.

Hacking, Ian. 1996. "The Looping Effects of Human Kinds." In *Causal Cognition: A Multidisciplinary Debate.* Edited by Sperber, Premack and Premack. Oxford: Oxford University Press.

Hacking, Ian. 1999. *The Social Construction of What?*. Cambridge: Harvard University Press.

Hacking, Ian. 2002. *Historical Ontology.* Cambridge: Harvard University Press.

Hacking, Ian. 2012. "Language, Truth, and Reason, 30 Years Later." *Studies in History and Philosophy of Science* 43: 599-609.

Harcourt, Bernard. 2020. *Critique and Praxis.* Columbia: Columbia University Press.

Haslanger, Sally. 2000. "Gender and Race: (What) Are They? (What) Do We Want Them to Be?" *Noûs* 34:1: 31–55.

Haslanger, Sally. 2007. "'But Mom, Crop-Tops Are Cute!' Social Knowledge, Social Structure and Ideology Critique." *Philosophical Issues* 17, The Metaphysics of Epistemology, 70–91.

Haslanger, Sally. 2011. "Ideology, Generics, and Common Ground." In *Feminist Metaphysics: Explorations in the Ontology of Sex, Gender and the Self.* Edited by Charlotte Witt. Dordrecht and London: Springer, 179–207.

Haslanger, Sally. 2017. "Racism, Ideology, and Social Movements." *Res Philosophica* 94.1, 1-22. http://dx.doi.org/10.11612/ resphil.1547

Haslanger, Sally. 2018. "What Is a Social Practice?" *Royal Institute of Philosophy Supplement* 82: 231-247. DOI: 10.1017/S1358246118000085

Hegel, Georg Wilhelm Friedrich. 1821 [2006]. *Phänomenologie des Geistes.* Edited by Dietmar Köhler and Otto Pöggeler. Berlin: Akademie Verlag.

Heinämaa, Sara. 1997. "What Is a Woman? Butler and Beauvoir on the Foundations of the Sexual Difference." *Hypatia* 12.1: 20-39.

Herman, Barbara. 1993. *The Practice of Moral Judgment.* Cambridge, MA: Harvard University Press.

Herman, Barbara. 1996. "Making Room for Character." In *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty.* Edited by Stephen Engstrom and Jennifer Whiting. Cambridge: Cambridge University Press, 36–60.

Herman, Barbara. 2007. *Moral Literacy*. Cambridge, MA: Harvard University Press.

Herman, Barbara. 2009. "Contingency in Obligation." *Nomos* 49, 17-53.

Hill, Thomas E. and Bernard Boxill. 2001. "Kant and Race." *Race and Racism*. Edited by Bernard Boxill. Oxford: Oxford University Press, 448-471.

Hill, Thomas E. Jr. 1991. "The Importance of Autonomy." In *Autonomy and Self-Respect*. Cambridge: Cambridge University Press, 43-51.

Höffe, Otfried. 1979. "Kants kategorischer Imperativ als Kriterium des Sittlichen." *Ethik und Politik*. Edited by Otfried Höffe. Frankfurt: Suhrkamp, 84–118.

Honneth, Axel. 1992 [1996]. *The Struggle for Recognition: The Moral Grammar of Social Conflicts.* Cambridge: MIT Press.

Honneth, Axel. 2000 [2007]. *Disrespect: The normative foundations of critical theory.* Cambridge: Polity.

Honneth, Axel. 2001. "Reconstructive Social Critique with a Genealogical Reservation: On the Idea of Critique in the Frankfurt School." *Graduate Faculty Philosophy Journal* 22.2, 3-12.

Honneth, Axel. 2007 [2009]. *Pathologies of Reason: On the Legacy of Critical Theory.* Translated by James Ingram et. al. New York: Columbia University Press.

Honneth, Axel. 2011 [2014]. *Freedom's Right: The Social Foundations of Democratic Life.* Translated by Joseph Ganahl. New York: Columbia University Press.

Honneth, Axel. 2014. "The Normativity of Ethical Life." *Philosophy and Social Criticism* 40.8: 817-826.

hooks, bell. 1984. *Feminist Theory: From Margin to Center.* Boston: South End Press.

Hornsby, Jennifer and Rae Langton. "Free Speech and Illocution." *Legal Theory* 4.1: 21-37.

Horwich, Paul. 1998. *Truth*. Oxford: Oxford University Press.

Hursthouse, Rosalind. 1999. *On Virtue Ethics*. Oxford: Oxford University Press.

Irwin, T.H. 2000. "Ethics as an Inexact Science: Aristotle's Ambitions for Moral Theory." In

Moral Particularism. Edited by Brad Hooker and Margaret Olivia Little. Oxford: Clarendon Press, 100-129.

Jaeggi, Rahel. 2009. "Rethinking Ideology." In *New Waves in Political Philosophy*. Edited by Boudewijn de Bruin and Christopher F. Zurn. Hampshire, UK: Palgrave Macmillan, 63-86.

Jaeggi, Rahel. 2013. *Kritik von lebensformen*. Berlin: Suhrkamp Verlag.

Jaggar, Alison. 1974. "On Sexual Equality." *Ethics* 84:4: 275-291.

Johnston, David. 1994. *The Idea of a Liberal Theory*. Princeton: Princeton University Press.

Kain, Patrick. 2004. "Self-Legislation in Kant's Moral Philosophy." *Archiv Für Geschichte Der Philosophie* 86.3: 257–306.

Kant, Immanuel. 1900-. *Kants gesammelte Schriften: herausgegeben von der Königlichen Preussischen Akademie der Wissenschaften,* 29 volumes. Berlin: G. Reimer (subsequently Walter de Gruyter & Co.).

Kant, Immanuel. 1996. *The Cambridge Edition of the Works of Immanuel Kant*. Edited by Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press.

Katsafanas, Paul. 2013. *Agency and the Foundations of Ethics: Nietzschean Constitutivism.* Oxford: Oxford University Press.

Kitcher, Patricia. 2003. "What Is a Maxim?" *Philosophical Topics* 31 (1/2): 215–43.

Klein, Peter and Ted A. Warfield. 1994. "What Price Coherence?" *Analysis* 54.3: 129-132.

Klein, Peter and Ted A. Warfield. 1996. "No Help for the Coherentist." *Analysis* 56: 118-121.

Kleingeld, Pauline, and Marcus Willaschek. 2019. "Autonomy Without Paradox." *Philosopher's Imprint* 19.6, 1-18.

Kleingeld, Pauline. 1993. "The Problematic Status of Gender-Neutral Language in the History of Philosophy: The Case of Kant." *Philosophical Forum* 25: 134–50.

Kleingeld, Pauline. 2007. "Kant's Second Thoughts on Race." *The Philosophical Quarterly* 57 (229): DOI: 10.1111/j.1467-9213.2007.498.x.

Kleingeld, Pauline. 2008. "Kant on Historiography and the Use of Regulative Ideas." *Studies in the History and Philosophy of Science*. 239.4: 523-528.

Kleingeld, Pauline. 2017. "The Principle of Autonomy in Kant's Moral Theory: Its Rise and Fall." In *Kant on Persons and Agency*. Edited by Eric Watkins, 61–80. Cambridge University Press. DOI: 10.1017/9781316856529.005.

Kleingeld, Pauline. 2018. "Moral Autonomy as Political Analogy: Self-Legislation in Kant's *Groundwork* and the Feyerabend Lectures on Natural Law (1784)." In *The Emergence of Autonomy in Kant's Moral Philosophy*. Edited by Stefano Bacin and Oliver Sensen. Cambridge: Cambridge University Press, 158-175.

Klemme, Heiner F. 2013. "Moralized Nature, Naturalized Autonomy." In Sensen (2013): 193–211.

Kohl, Marcus. 2015. "Kant on Freedom of Empirical Thought." *Journal of the History of Philosophy* 53.2: 301-326.

Kohl, Marcus. 2017. "Radical Evil as a Regulative Idea." *Journal of the History of Philosophy* 55.4: 641-673.

Koopman, Colin. 2013. *Genealogy as Critique: Foucault and the Problems of Modernity.* Bloomington: Indiana University Press.

Korsgaard, Christine M. 1996. *The Sources of Normativity.* Cambridge: Cambridge University Press.

Korsgaard, Christine M. 2003. "Realism and constructivism in twentieth-century moral philosophy." *Journal of Philosophical Research* 28, supplement: 99-122.

Korsgaard, Christine M. 2008. "Acting for a Reason." In *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press, 207-230. DOI: 10.1093/acprof:oso/9780199552733.003.0008.

Korsgaard, Christine M. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

Korsgaard, Christine M. 2018. *Fellow Creatures: Our Obligations to the Other Animals.* Oxford: Oxford University Press.

Kripke, Saul A. 1982. *Wittgenstein on Rules and Private Language: An Elementary Exposition.* Cambridge: Harvard University Press.

Kruks, Sonia. 2012. *Simone de Beauvoir and the Politics of Ambiguity*. New York: Oxford University Press.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Kukla, Rebecca. 2002. "Attention and Blindness: Contingency and Obligation in Moral Philosophy." *Canadian Journal of Philosophy* 28: 319-346.

Lambert, Thomas. 2019. "Nietzsche on creating and discovering values." *Inquiry* 62: 49-69.

Langton, Rae, and Caroline West. 1999. "Scorekeeping in a pornographic language game." *Australasian Journal of Philosophy* 77.3: 303-319.

Lara, Maria Pía. 2008. *Narrating Evil: A Postmetaphysical Theory of Reflective Judgment.* New York: Columbia University Press.

Larmore, Charles. 1981. "Moral Judgment." *The Review of Metaphysics* 35.2: 275-196.

Le Dœuff, Michèle. 1980. "Simone de Beauvoir and existentialism." *Feminist Studies* 6.2: 277-289.

Le Dœuff, Michèle. 1989. *Hipparchia's Choice: An Essay Concerning Women, Philosophy, Etc.* Translated by Trista Selous. New York: Columbia University Press.

Le Dœuff, Michèle. 1995. "Simone de Beauvoir: Falling into (Ambiguous) Line." In *Feminist Interpretations of Simone de Beauvoir*, edited by Margaret A. Simons, 59–66. University Park: Penn State University Press.

Lee, Ang, director. 2005. *Brokeback Mountain*. New York: Focus Features, DVD.

Lewis, David. 1979. "Scorekeeping in a language game." In *Semantics from different points of view*. Berlin: Springer, 172-187.

Longuenesse, Béatrice. 2001. *Kant and the Capacity to Judge: Sensibility and Discursivity in the Transcendental Analytic of the Critique of Pure Reason*. Princeton, NJ: Princeton University Press.

Lorenzini, Daniele. 2020. "On Possibilising Genealogy." *Inquiry: An Interdisciplinary Journal of Philosophy.* DOI: 10.1080/0020174X.2020.1712227

Louden, Robert B. 2000. *Kant's Impure Ethics: From Rational Beings to Human Beings.* Oxford: Oxford University Press.

Love, S.M. 2017. "Kant After Marx." *Kantian Review* 22.4, special issue on Kant and Marx: 579-598.

Makkreel, Rudolf. 2002. "Reflective Judgment and the Problem of Assessing Virtue in Kant." *Journal of Value Inquiry* 36: 205-220.

Makkreel, Rudolf. 2015. *Orientation and Judgment in Hermeneutics.* Chicago: University of Chicago Press.

Marx, Karl and Friedrich Engels. 1846 [1998]. *The German Ideology*. Amherst: Prometheus Books.

Marx, Karl. 1844 [1978]. "On the Jewish Question." In *The Marx-Engels Reader.* Edited by Robert C. Tucker. New York: W.W. Norton, 26-52.

Massimi, Michela. 2017. "What is this Thing Called 'Scientific Knowledge'? – Kant on Imaginary Standpoints and the Regulative Role of Reason." *Kant Yearbook* 9.1: 63-84.

Matherne, Samantha. 2013. "The Inclusive Interpretation of Kant's Aesthetic Ideas." *British Journal of Aesthetics* 53.1: 21–39.

Matherne, Samantha. 2014. "Kant and the Art of Schematism." Kantian Review 19.2: 181-205.

McDowell, John. 1978. "Are Moral Requirements Hypothetical Imperatives?" *Proceedings of the Aristotelian Society, Supplementary Volumes* 52: 13-29.

McDowell, John. 1998. *Mind, Value, and Reality.* Cambridge: Harvard University Press.

Merritt, Melissa McBay. 2018. *Kant on Reflection and Virtue*. Cambridge: Cambridge University Press.

Mills, Charles W. 2004. "'Ideal Theory' as Ideology." In *Moral Psychology: Feminist Ethics and Social Theory.* Edited by Peggy DesAutels and Margaret Urban Walker. Lanham, Md.: Rowman & Littlefield, 163-181

Mills, Charles W. 2005. "Kant's Untermenschen." In *Race and Racism in Modern Philosophy*, edited by Andrew Valls, 169–93. Ithaca, NY: Cornell University Press.

Mills, Charles W. 2014. "Kant and Race, Redux." *Graduate Faculty Philosophy Journal* 35 (1): 125–57. https://doi.org/10.5840/gfpj2014351/27.

Mills, Charles W. 2017. "Black Radical Kantianism." *Res Philosophica* 95 (1): 1–33.

Moi, Toril. 2008. *Simone de Beauvoir: The Making of an Intellectual Woman*. Second edition. Oxford; New York: Oxford University Press.

Moody-Adams, Michele M. 1994. "Culture, Responsibility, and Affected Ignorance." *Ethics* 104: 291-309.

Moody-Adams, Michele M. 1999. "The Idea of Moral Progress." *Metaphilosophy* 30: 168-185.

Moss, Jessica. 2014. "Right Reason and Aristotle: On the Meaning of *Logos*." *Phronesis* 59.3: 181-230.

Munzel, G. Felicitas. 1998. *Kant's Conception of Moral Character: The "Critical" Link of Morality, Anthropology, and Reflective Judgment.* Chicago, IL: Chicago University Press.

Murdoch, Iris. 1970 [1997]. *The Sovereignty of Good.* Reprinted in Iris Murdoch, *Existentialists and Mystics: Writings on Philosophy and Literature.* Edited by Peter Conradi. New York: Penguin, 299-386.

Murdoch, Iris. 1977 [1997]. "Literature and philosophy: A conversation with Bryan Magee." In

Iris Murdoch, *Existentialists and mystics: writings on philosophy and literature.* Edited by Peter Conradi. New York: Penguin, 3-30.

Neuhouser, Frederick. 1990. *Fichte's Theory of Subjectivity*. Cambridge: Cambridge University Press.

Ng, Karen. 2020. "Back to Adorno: Critical Theory's Problem of Normative Grounding." *Current Perspectives in Social Theory* 36: 49-59.

Nietzsche, Friedrich. 1882 [2016]. *The Gay Science: With a Prelude in German Rhymes and an Appendix of Songs.* Edited by Bernard Williams. Translated by Josefine Nauckhoff. Cambridge: Cambridge University Press.

Nietzsche, Friedrich. 1886 [1989]. *Beyond Good and Evil: Prelude to a Philosophy of the Future.* Translated by Walter Kaufmann. New York: Vintage.

Norton, Peter D. 2007. "Street Rivals: Jaywalking and the Invention of the Motor Age Street." *Technology and Culture* 48, no. 2: 331-59.

Nussbaum, Martha. 1986. *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy.* Cambridge: Cambridge University Press.

Nussbaum, Martha. 1990. "The Discernment of the Particular." In *Love's Knowledge: Essays on Philosophy and Literature.* Oxford: Oxford University Press.

O'Neill, Onora. 1975. *Acting on Principle: An Essay on Kantian Ethics*. Cambridge: Cambridge University Press.

O'Neill, Onora. 1987. "Abstraction, Idealization and Ideology in Ethics." *Royal Institute of Philosophy Lecture Series* 22, 55-69.

O'Neill, Onora. 1989. *Constructions of Reason: Explorations of Kant's Practical Philosophy*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139173773

O'Neill, Onora. 2002. "Instituting Principles: Between Duty and Action." In *Kant's Metaphysics of Morals: Interpretative Essays*. Edited by Mark Timmons. Oxford: Clarendon Press, 331–47.

O'Neill, Onora. 2018. *From Principles to Practice: Normativity and Judgement in Ethics and Politics*. Cambridge: Cambridge University Press.

Ortner, Sherry B. 2006. *Anthropology and Social Theory: Culture, Power, and the Acting Subject*. Durham, NC: Duke University Press.

Owen, David. 2002. "Criticism and Captivity: On Genealogy and Critical Theory." *European Journal of Philosophy* 10.2, 212-230.

Papish, Laura. 2018. “Kant’s Revised Account of the Non-Moral Imperatives of Practical Reason.” Ergo 5.11: 289-317.

Platts, Mark. 1988. “Moral Reality.” In *Essays on Moral Realism*. Edited by Geoffrey Sayre-McCord. Ithaca: Cornell University Press.

Pluhar, Werner S. “Translator’s Preface.” In *Critique of Pure Reason*. Translated by Werner S. Pluhar. Indianapolis and Cambridge: Hackett, 5-40.

Queloz, Matthieu. 2018. “How genealogies can affect the space of reasons.” *Synthese*. Online first April 5. DOI: 10.1007/s11229-018-1777-9

Queloz, Matthieu. 2019. “From Paradigm-Based Explanation to Pragmatic Genealogy”. *Mind*. Online first February 21. https://doi.org/10.1093/mind/fzy083/5359485

Quine, Willard Van Orman. 1960 [2013]. *Word and object*. Cambridge: MIT Press.

Rawls, John. 1955. “Two Concepts of Rules.” *The Philosophical Review* 64.1, 3-32.

Rawls, John. 1980. “Kantian constructivism in moral theory.” *The journal of philosophy* 77.9: 515-572.

Rawls, John. 1989. “Themes in Kant’s moral philosophy.” *Kant’s transcendental deductions*: *the three critiques and the* Opus postumum. Edited by Eckart Förster. Plato Alto: Stanford University Press, 81-113.

Rawls, John. 2000. *Lectures on the History of Moral Philosophy.* Edited by Barbara Herman. Cambridge: Harvard University Press.

Raz, Joseph. 1986. *The Morality of Freedom*. Oxford: Clarendon Press.

Reath, Andrews. 2006. *Agency and Autonomy in Kant’s Moral Theory: Selected Essays*. Oxford: Clarendon Press.

Recki, Birgit. 2001. *Ästhetik der Sitten: die Affinität von ästhetischem Gefühl und praktischer Vernunft bei Kant*. Frankfurt: Vittorio Klostermann.

Rescher, Nicholas. 2019. “Conceivability.” *Humanities Bulletin* 2.1: 8-19.

Richardson, John. 2004. *Nietzsche’s new Darwinism.* Oxford: Oxford University Press.

Ricoeur, Paul. 1965 [2014]. *De l’interprétation. Essai sur Freud.* Paris: Seuil.

Rouse, Joseph. 2001. “Two Concepts of Practices.” In *The Practice Turn in Contemporary Theory.* Edited by Theodore Schatzki, Karin Knorr Cetina, and Eike Von Sevigny. London: Routledge, 198-208.

Rush Jr, Fred L. 2000. "Reason and regulation in Kant." *The Review of Metaphysics* 53.4: 837-862.

Russell, Francey. 2019. Review of Melissa Merritt, *Kant on Reflection and Virtue. SGIR Review* 2.1: 60-71.

Sartre, Jean-Paul. 1943 [1993]. *Being and Nothingness: An Essay on Phenomenological Ontology.* Translated by Hazel E. Barnes. New York: Washington Square Press.

Schatzki, Theodore. 2001. "Practice Mind-ed Orders." In *The Practice Turn in Contemporary Theory.* Edited by Theodore Schatzki, Karin Knorr Cetina, and Eike Von Sevigny. London: Routledge, 50-63.

Schröder, Hannelore. 1997. "Kant's Patriarchal Order." *Feminist Interpretations of Immanuel Kant*, edited by R. M. Schott. University Park: Pennsylvania State University Press, 275–96.

Schroeder, Mark. 2014. "Hypothetical Imperatives, Scope, and Jurisdiction." In *Explaining the Reasons We Share: Explanation and Expression in Ethics*, Volume 1. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780198713807.003.0011.

Schwartz, Maria. 2006. *Der Begriff der Maxime bei Kant: Eine Untersuchung des Maximenbegriffs in Kants praktischer Philosophie*. Berlin: Lit Verlag.

Schwartz, Maria. 2006. *Der Begriff der Maxime bei Kant. Eine Untersuchung des Maximenbegriffs in Kants praktischer Philosophie.* Münster: Lit Verlag.

Searle, John. 1995. *The Construction of Social Reality.* New York: Free Press.

Searle, John. 2009. *Making the Social World: The Structure of Human Civilization.* Oxford: Oxford University Press.

Sensen, Oliver (editor). 2013. *Kant on Moral Autonomy*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511792489.006.

Shelby, Tommie. 2003. "Ideology, Racism, and Critical Social Theory." *The Philosophical Forum* 34.2: 153-188.

Shelby, Tommie. 2014. "Racism, Moralism, and Social Criticism." *Du Bois Review* 11:1: 57– 74 .

Silk, Alex. 2015. "Nietzschean Constructivism: Ethics and Metaethics for All and None." *Inquiry* 58: 244-280.

Spivak, Gayatri Chakravorty. 1988. "Can the Subaltern Speak?" In *Marxism and the Interpretation of Culture*. Edited by Cary Nelson and Lawrence Grossberg. Houndmills and London: MacMillan, 271-316.

Spivak, Gayatri Chakravorty. 2006. "If Only." *The Scholar and Feminist Online* 4.2. Special issue: "Writing a Feminist's Life: The Legacy of Carolyn G. Heilbrun". Accessed October 3,

2020. http://sfonline.barnard.edu/heilbrun/spivak_03.htm

Srinivasan, Amia. 2018. "How to Do Things with Philosophy." *European Journal of Philosophy* 26: 1410–1416. DOI: 10.1111/ejop.12409

Srinivasan, Amia. 2019. "Genealogy, Epistemology, and Worldmaking." *Proceedings of the Aristotelian Society* 119.2, 127-156. DOI: 10.1093/arisoc/aoz009

Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127: 109-166.

Street, Sharon. 2009. "In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters." *Philosophical Issues* 19: 273-298.

Street, Sharon. 2010. "What Is Constructivism in Ethics and Metaethics?" *Philosophy Compass* 5/5: 363-384. DOI: 10.1111/j.1747-9991.2009.00280.x.

Street, Sharon. Unpublished manuscript. "How to Be a Relativist About Normativity," 1-19. Accessed October 18, 2020. https://www.academia.edu/19887937/How_to_Be_a_Relativist_About_Normativity

Taylor, Charles. 1971. "Interpretation and the Sciences of Man." *The Review of Metaphysics* 25.1, 3-51.

Theunissen, Nandi. 2016. "Kant's Commitment to Metaphysics of Morals." *European Journal of Philosophy* 24.1: 103-128. DOI: 10.1111/ejop.12051.

Thomason, Krista K. 2020 (online first). "Wild chimeras: Enthusiasm and intellectual virtue in Kant." *European Journal of Philosophy.* DOI: 10.1111/ejop.12481

Thorndike, Oliver. 2018. *Kant's Transition Project and Late Philosophy: Connecting the Opus Postumum and Metaphysics of Morals*. Bloomsbury.

Tiisala, Tuomo. 2016. "Power and Freedom in the Space of Reasons." PhD diss., University of Chicago. ProQuest (10129652).

Tiisala, Tuomo. 2017. "Overcoming 'The Present Limits of the Necessary': Foucault's Conception of a Critique." *The Southern Journal of Philosophy* 55: 7-24. DOI: 10.1111/sjp.12224

Timmermann, Jens. 2000. "Kant's Puzzling Ethics of Maxims." *The Harvard Review of Philosophy* 8.1: 39–52.

Turner, Stephen. 2001. "Throwing Out the Tacit Rule Book: Learning and Practices." In *The Practice Turn in Contemporary Theory.* Edited by Theodore Schatzki, Karin Knorr Cetina, and Eike Von Sevigny. London: Routledge, 129-139.

Vaccarino Bremner, Sabina. Forthcoming. "'Until Art Once More Becomes Nature': Culture and the Problem of Unity in Kant's *Critique of Judgment*." *Archiv für Geschichte der Philosophie*. Accessed October 18, 2020. https://philarchive.org/rec/BREUAO

Walker, Margaret Urban. 2007. *Moral Understandings: A Feminist Study in Ethics.* Oxford: Oxford University Press.

Wallner, Ingrid M. 1984. "A New Look at J. S. Beck's 'Doctrine of the Standpoint.'" *Kant-Studien* 75.3: 294–316.

Webber, Jonathan. 2018. *Rethinking Existentialism.* Oxford: Oxford University Press.

Wolff, Christian. 1733. *Vernünfftige Gedancken von der Menschen Thun und Lassen, zur Beförderung ihrer Glückseeligkeit*, 4th edition Frankfurt and Leipzig: n.p.

Williams, Bernard. 1981. *Moral Luck: Philosophical Papers 1973–1980*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139165860.

Williams, Bernard. 1985. *Ethics and the Limits of Philosophy.* Cambridge: Cambridge University Press.

Williams, Bernard. 1996. "History, Morality, and the Test of Reflection." In Korsgaard (1996), 210-218.

Williams, Bernard. 2002. *Truth and Truthfulness: An Essay in Genealogy.* Princeton: Princeton University Press.

Winch, Peter. 1958 [2008]. *The Idea of a Social Science and Its Relation to Philosophy.* London: Routledge.

Winch, Peter. 1965. "The Universalizability of Moral Judgements." Edited by Sherwood J. B. Sugden. *Monist* 49 (2): 196–214. https://doi.org/10.5840/monist196549214.

Winch, Peter. 1987. "Who is my Neighbour?". In *Trying to Make Sense.* Oxford: Blackwell: 154-166.

Wittgenstein, Ludwig. 1953 [2009]. *Philosophische Untersuchungen / Philosophical Investigations.* English version translated by G.E.M. Anscombe, P.M.S. Hacker and Joachim Schulte. Chichester: Wiley-Blackwell.

Wollstonecraft, Mary. 1792. *A Vindication of the Rights of Women.* Edited by Eileen Hunt Botting. New Haven: Yale University Press.

Wood, Allen and Robert Louden. 2012. "General Introduction." In *Lectures on Anthropology by Immanuel Kant*. Edited by Allen Wood, Robert Louden. Cambridge: Cambridge University Press, 1-10.

Wood, Allen. 1981. *Karl Marx.* New York: Routledge.

Wood, Allen. 1991. "Unsociable Sociability: The Anthropological Basis of Kant's Ethics." In *Philosophical Topics* 19.1: 325-251.

Wood, Allen. 1995. "Exploitation." *Social Philosophy and Policy* 12.2: 136-158.

Wood, Allen. 1999. *Kant's Ethical Thought.* Cambridge: Cambridge University Press.

Wood, Allen. 2003. "Kant and the Problem of Human Nature." In *Essays on Kant's Anthropology.* Edited by Brian Jacobs and Patrick Kain. Cambridge: Cambridge University Press.

Young, Iris Marion. 1987. "Impartiality and the Civic Public: Some Implications of Feminist Critiques of Moral and Political Theory." *Feminism as Critique*. Edited by Seyla Benhabib and Drucilla Cornell. Oxford and Minneapolis: Polity and University of Minnesota Press, 56-76.

Zammito, John H. 1992. *The Genesis of Kant's* Critique of Judgment. Chicago: University of Chicago Press.

Zerilli, Linda G.M. 2016. *A Democratic Theory of Judgment*. Chicago: University of Chicago Press.

Zuckert, Rachel. 2017. "Empirical Scientific Investigation and the Ideas of Reason." *Kant and the Laws of Nature.* Edited by Michela Massimi and Angela Breitenbach. Cambridge: Cambridge University Press, 89-107.