Kant’s Search for the Supreme Principle of Morality

At the core of Kant’s ethics lies the claim that if there is a supreme principle of morality, then it is not a utilitarian or Aristotelian perfectionist principle, or even a principle resembling the Ten Commandments. The only viable candidate for the supreme principle of morality is the Categorical Imperative.

This book is the most detailed investigation of this claim. It constructs a new, criterial reading of Kant’s derivation of one version of the Categorical Imperative: the Formula of Universal Law. This reading shows this derivation to be far more compelling than contemporary philosophers tend to believe. It also reveals a novel approach to deriving another version of the Categorical Imperative, the Formula of Humanity, a principle widely considered to be the most attractive Kantian candidate for the supreme principle of morality.

Lucidly written and dealing with a foundational topic in the history of ethics, this book will be important not just for Kant scholars but for a broad swath of students of philosophy.

Samuel J. Kerstein is Associate Professor of Philosophy at the University of Maryland, College Park.
Kant’s Search for the Supreme Principle of Morality

SAMUEL J. KERSTEIN

University of Maryland, College Park
# Contents

<table>
<thead>
<tr>
<th>Acknowledgments</th>
<th>page xi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key to Abbreviations and Translations</td>
<td>xiii</td>
</tr>
</tbody>
</table>

Introduction: Derivation, Deduction, and the Supreme Principle of Morality

1.1 No Modest Claim 1

1.2 The Basic Concept of the Supreme Principle of Morality 1

1.3 Derivation and Deduction of the Categorical Imperative 4

1.4 The (Alleged) Gap in the Derivation of the Formula of Universal Law 7

1.5 Terminological and Thematic Clarifications 10

1.6 Outline of the Book 11

1 Fundamental Concepts in Kant’s Theory of Agency 16

1.1 Aims and Limits of the Discussion 16

1.2 Maxims: A Basic Account 16

1.3 Maxims and Other Rules of the Same Form 19

1.4 The Will 20

1.5 Determining Grounds of the Will 21

1.6 Acting from Inclination: Three Interpretations and Their Importance 22

1.7 Acting from Inclination in the *Groundwork* and in the *Metaphysics of Morals* 24

1.8 Material Practical Principles: Acting from Inclination in the *Critique of Practical Reason* 29

2 Transcendental Freedom and the Derivation of the Formula of Universal Law 33

2.1 Derivation in the *Critique of Practical Reason*: Allison’s Reconstruction 33
## Contents

2.2 A Thick Account of Kantian Rational Agency 34
2.3 Desire and Justification of Action 36
2.4 Practical Law and Justification of Action 39
2.5 Practical Law and the Formula of Universal Law 42

3 The Derivation of the Formula of Humanity 46
3.1 Outline of the Derivation 46
3.2 The Supreme Principle of Morality and Unconditional Value 47
3.3 The Unconditional Value of Humanity: Kant’s Argument 54
3.4 Korsgaard’s Reconstruction: Preliminaries 55
3.5 The Supreme Principle of Morality and Good Ends 56
3.6 From Good Ends to the Unconditional Value of Humanity: The Regressive Argument 59
3.7 The Failure of the Regressive Argument 65
3.8 Shortcomings in the Derivation of the Formula of Humanity 71

4 The Derivation of the Formula of Universal Law: A Criterial Reading 73
4.1 Main Steps of the Derivation on the Criterial Reading 73
4.2 Korsgaard’s Reading of the Derivation 74
4.3 The Structure of *Groundwork* I 77
4.4 The Failure of One Version of the Traditional Reading of the Derivation 77
4.5 The Challenge Posed by Aune’s Version of the Traditional Reading 78
4.6 From Duty and Moral Worth to Two Criteria for the Supreme Principle of Morality 80
4.7 Law as Motive: A Third Criterion for the Supreme Principle of Morality 82
4.8 The Criterial Reading and *Groundwork* II 86
4.9 Coherence with Ordinary Moral Reason: A Fourth Criterion 87
4.10 The Apriority of the Supreme Principle of Morality 89
4.11 Rejecting the Traditional Interpretation of the *Groundwork* II Derivation 91
4.12 Summary 93

5 Criteria for the Supreme Principle of Morality 95
5.1 Plan of Discussion: Focus on First Criterion 95
5.2 Moral Worth and Actions Contrary to Duty 96
5.3 Two Conditions on Acting from Duty 98
5.4 All Actions from Duty Have Moral Worth 104
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5</td>
<td>Only Actions from Duty Have Moral Worth</td>
<td>106</td>
</tr>
<tr>
<td>5.6</td>
<td>The Second Criterion and Its Grounds</td>
<td>109</td>
</tr>
<tr>
<td>5.7</td>
<td>The Third Criterion and Its Grounds</td>
<td>110</td>
</tr>
<tr>
<td>5.8</td>
<td>Relations between the Criteria</td>
<td>112</td>
</tr>
<tr>
<td>6</td>
<td>Duty and Moral Worth</td>
<td>114</td>
</tr>
<tr>
<td>6.1</td>
<td>Aims of the Discussion</td>
<td>114</td>
</tr>
<tr>
<td>6.2</td>
<td>Moral Worth and Helping a Friend from Duty</td>
<td>116</td>
</tr>
<tr>
<td>6.3</td>
<td>One Thought Too Many?</td>
<td>118</td>
</tr>
<tr>
<td>6.4</td>
<td>The Moral Worth of Actions Contrary to Duty</td>
<td>119</td>
</tr>
<tr>
<td>6.5</td>
<td>A Disturbing Asymmetry in Kant’s View of Moral Worth</td>
<td>119</td>
</tr>
<tr>
<td>6.6</td>
<td>Failure of Will or Unfortunate Event?</td>
<td>121</td>
</tr>
<tr>
<td>6.7</td>
<td>Moral Permissibility and Moral Worth in the <em>Metaphysics of Morals</em></td>
<td>124</td>
</tr>
<tr>
<td>6.8</td>
<td>The (Alleged) Transparency of Moral Requirements</td>
<td>127</td>
</tr>
<tr>
<td>6.9</td>
<td>Odious Actions and Moral Worth</td>
<td>129</td>
</tr>
<tr>
<td>6.10</td>
<td>Sympathy and Moral Worth</td>
<td>132</td>
</tr>
<tr>
<td>6.11</td>
<td>Summary</td>
<td>138</td>
</tr>
<tr>
<td>7</td>
<td>Eliminating Rivals to the Categorical Imperative</td>
<td>139</td>
</tr>
<tr>
<td>7.1</td>
<td>Aims of the Discussion</td>
<td>139</td>
</tr>
<tr>
<td>7.2</td>
<td>A Sweeping Argument against All Rivals</td>
<td>140</td>
</tr>
<tr>
<td>7.3</td>
<td>The Structure of Act Utilitarianism</td>
<td>145</td>
</tr>
<tr>
<td>7.4</td>
<td>Against Act Utilitarianism</td>
<td>146</td>
</tr>
<tr>
<td>7.5</td>
<td>Against Expectabilist Utilitarianism</td>
<td>148</td>
</tr>
<tr>
<td>7.6</td>
<td>Against Perfectionism</td>
<td>152</td>
</tr>
<tr>
<td>7.7</td>
<td>Kantian Consequentialism?</td>
<td>153</td>
</tr>
<tr>
<td>7.8</td>
<td>Against a Principle Akin to the Ten Commandments</td>
<td>155</td>
</tr>
<tr>
<td>7.9</td>
<td>Further Nonconsequentialist Rivals</td>
<td>158</td>
</tr>
<tr>
<td>7.10</td>
<td>Summary</td>
<td>159</td>
</tr>
<tr>
<td>8</td>
<td>Conclusion: Kant’s Candidates for the Supreme Principle of Morality</td>
<td>160</td>
</tr>
<tr>
<td>8.1</td>
<td>Kant’s Candidates and Criteria for the Supreme Principle of Morality</td>
<td>160</td>
</tr>
<tr>
<td>8.2</td>
<td>Two Formulas and the Basic Concept of the Supreme Principle of Morality</td>
<td>162</td>
</tr>
<tr>
<td>8.3</td>
<td>Two Formulas and Further Criteria</td>
<td>165</td>
</tr>
<tr>
<td>8.4</td>
<td>Two Formulas and Ordinary Moral Consciousness</td>
<td>167</td>
</tr>
<tr>
<td>8.5</td>
<td>Formula of Universal Law: Practical Contradiction Interpretation</td>
<td>168</td>
</tr>
<tr>
<td>8.6</td>
<td>Formula of Universal Law: Universal Availability Interpretation</td>
<td>171</td>
</tr>
<tr>
<td>8.7</td>
<td>Fundamentals of the Formula of Humanity</td>
<td>174</td>
</tr>
</tbody>
</table>
# Contents

8.8 Deriving Duties from the Formula of Humanity 177  
8.9 Formula of Humanity: Further Challenges 183  
8.10 Where We End Up 187  

Notes 193  
Index 221
Acknowledgments

This book would not have been completed without help and support from a variety of sources.

I would like to thank Terence Moore and Brian R. MacDonald of Cambridge University Press for their patience and expertise in guiding me through the publication process.


Most of the book was written during the academic year 1999–2000, which I spent as a Fellow at the National Humanities Center in Triangle Park, North Carolina. I would like to thank the National Endowment for the Humanities for supporting my stay there. The administrators and staff at the National Humanities Center could not have been more encouraging and helpful. In particular I would like to thank Karen Carroll, who edited an early version of my manuscript. (I would also like to thank Jane Strong for editing a later version.) Preliminary work on the manuscript was made possible by support from the University of Maryland, College Park, in the form of a General Research Board grant that relieved me from my teaching duties during the fall of 1996. I would like to thank the University of Maryland for this support, as well as for granting me leave to work at the National Humanities Center.
Acknowledgments

For their comments and criticisms of portions of this book, I would like to thank audiences at the British Kant Society Annual Meeting, the Central Division Meeting of the American Philosophical Association, the Midwest Study Group of the North American Kant Society, Duke University, the University of St. Andrews, and the University of North Carolina, Chapel Hill.

From early on I have been fortunate to have had outstanding teachers. I would like to thank Noël Carroll and Victor Gourevitch for their guidance, both philosophical and personal. I am grateful to Bonnie Kent who took the time to teach me not only how to work in the history of philosophy but to appreciate the importance of doing so.

I have learned a great deal about Kantian ethics from discussion and/or correspondence with many philosophers, including Paul Cohen, Michèle Crampe-Casnabet, Garrett Callity, David Cummiskey, Raymond Geuss, Stéphane Haber, Thomas Hill Jr., Dieter Schönheker, Ralf Stöcker, and Allen Wood. I owe a special debt of gratitude to Berys Gaut. Some central ideas in the book stem from our collaborative work, and Berys has been generous in encouraging me to develop them at greater length. Readers for Cambridge University Press, as well as two others, offered comments that have, I think, enabled me to strengthen several of my arguments. During my stay at the National Humanities Center, I profited from (often ambulatory) dialogue with many colleagues, including Ruth Grant, Michelle Massé, Louise McReynolds, Bernard Reginster, Daniel Sherman, Eleonore Stump, Timothy Taylor, and Marjorie Woods. I was especially fortunate to have been able to discuss philosophy with Thomas Christiano, who not only provided intellectual inspiration, but patiently helped me to work out some key points in the book. My friends and colleagues at the University of Maryland, especially Judith Lichtenberg and Corey Washington, have aided me at several points, both intellectually and personally, in carrying out this project.

I am deeply grateful for the help and support I have received from Rüdiger Bittner, Thomas Pogge, and Michael Slote. From the beginning, these philosophers have played essential roles in the book’s development. Each gave me valuable advice on my project as it unfolded, and offered trenchant and productive comments on the manuscript as a whole. My approach to Kantian ethics owes a great deal to each of them.

Finally, I would like to thank my in-laws John and Jane Strong, my parents Howard and JoAnn Kerstein, and especially my wife Lisa Strong, for their constant encouragement during the writing of this book.
Key to Abbreviations and Translations

Except for references to the *Critique of Pure Reason*, all references to Kant are to the Preussische Akademie der Wissenschaften edition of his works (Berlin: Walter de Gruyter [and predecessors], 1902). References to the *Critique of Pure Reason* are to the standard A and B pagination of the first and second editions. I list here the German title, academy edition (Ak.) volume number, and abbreviation for each of the works I cite. Under each entry, I specify the English edition I have consulted. The translations I employ sometimes vary from those of these English editions.

**Anth** *Anthropologie in pragmatischer Hinsicht* (Ak. 7)

**GMS** *Grundlegung zur Metaphysik der Sitten* (Ak. 4)

**KpV** *Kritik der praktischen Vernunft* (Ak. 5)

**KrV** *Kritik der reinen Vernunft* (1st ed. (A) 1781; 2nd ed. (B) 1787; Ak. 3–4)

**KU** *Kritik der Urteilskraft* (Ak. 5)

**KUE** *Erste Einleitung in der Kritik der Urteilskraft* (Ak. 20)
Key to Abbreviations and Translations

LE Vorlesungen über Moralphilosophie, “Moralphilosophie Collins” (Ak. 27)

MS Die Metaphysik der Sitten (Ak. 6)

Rel Die Religion innerhalb der Grenzen der blossen Vernunft (Ak. 6)

All of the English editions incorporate academy edition page numbering in their margins, except for the KrV and Rel. When I cite the Rel, I give the academy edition page number followed by that of the English edition.