

# Rightness as Fairness: A Moral and Political Theory (Palgrave MacMillan, 2016)

By Marcus Arvan

Overview and Chapter Abstracts

Now available at Palgrave MacMillan. Also available [on Amazon](#). For more, visit [marcusarvan.net](http://marcusarvan.net) (sample chapter available soon).

## Overview

Moral philosophy has long been marked by systematic disagreement. In *Rightness as Fairness*, I argue that in order to reliably arrive at moral truth, moral philosophy must be based upon seven scientific principles of theory-selection. I then argue that our best empirical evidence reveals morality to be a type of prudence requiring us to act in ways that our present and future selves can rationally agree upon across time. I show that this agreement—Rightness as Fairness—requires us to be fair to ourselves and to others, including animals. Further, the Four Principles of Fairness comprising this agreement reconcile a variety of traditionally opposed moral and political frameworks. Finally, *Rightness as Fairness* provides a uniquely fruitful method for resolving applied moral and political issues: a method of ‘principled fair negotiation’ that requires merging principled debate with real-world negotiation.



Some innovations of *Rightness as Fairness* are that it:

1. **Reconciles several leading moral frameworks** (consequentialism, deontology, contractualism, and virtue ethics) into a coherent whole. Rather than arguing that these different frameworks converge on a single moral principle (as some other theorists have attempted to establish), *Rightness as Fairness* argues that morality is a matter of *balancing* all four frameworks through Four Principles of Fairness, which in turn combine into a single criterion of moral rightness for properly balancing each.
2. **Reconciles several leading political orientations** (libertarianism, egalitarianism, and communitarianism), showing how all three are based on genuine ‘ideal’ and ‘nonideal’ moral truths that must be weighed against each other through processes of fair negotiation for settling the nature and requirements of domestic, international, and global justice on an ongoing basis.
3. **Provides a new method of ‘principled fair negotiation’ for moral problem solving.** Morality is shown to be partly a matter of principle, but also fundamentally a matter of actual negotiation. Sound answers to applied moral issues cannot (generally) be discovered through abstract moral argument alone, but must instead be partly created through ongoing processes of principled fair negotiation.
4. **Generates novel, nuanced analyses of applied ethical issues**, including the ethics of lying, suicide, trolley cases, torture, and treatment of nonhuman animals.
5. **Defends a new argument for the rationality of moral behavior**, based on the science of moral cognition, specifically our capacities for mental time-travel and modal imagination, as well as recent experimental findings on improving moral behavior.
6. **Explains morality's limits from within**, through a self-referential moral principle that recursively defines which of our actions are morally evaluable, as well as how “demanding” morality is.
7. **Satisfies all seven principles of theory-selection more successfully than existing moral theories.**

## **Chapter Abstracts**

### **Chapter 1. Ethics for the Twenty-First Century**

I argue that moral philosophy currently lacks a reliable method for distinguishing what is true about morality from what merely 'seems true' to different investigators. I then defend seven principles of theory-selected adapted from the sciences—Firm (Observational) Foundations, Internal Coherence, External Coherence, Explanatory Power, Unity, Parsimony, and Fruitfulness—as the most reliable method for distinguishing moral truth from 'seeming truth.' Next, using these seven principles, I argue that moral philosophy should be based on a simple, 'means-ends' instrumental theory of normative rationality: a theory that enjoys virtually universal support in everyday life and philosophical history. In the process, I show that a variety of alternative approaches to moral philosophy—intuitionism, constitutivism, *eudaimonism*, reflective equilibrium, moral realism, moral-language analysis, second- and third-personalism, etc.—all violate the first and most important principle of theory-selection: Firm (Observational) Foundations. I argue that only (A) instrumentalism and (B) other empirical facts regarding human cognition and moral psychology satisfy this first principle. I also argue that instrumentalism promises systematic advantages over other approaches on the other six principles of theory-selection as well. Finally, I address concerns that instrumentalism is not universally accepted, provides the 'wrong kinds of reasons' for moral behavior, and cannot establish morality's categorical normative force—explaining how the remainder of the book will disarm such concerns.

### **Chapter 2. The Problem of Possible Future Selves**

I show that our capacities of mental time travel—our abilities to imagine and care about our past, present, and future—generate a problem for rational decision-making for which there is no known solution: the Problem of Possible Future Selves. First, I show that in many cases of uncertainty, including paradigmatic moral decisions, we have an interest in both knowing our future interests and weighing them with our present ones, including our future interests regarding our past decisions. Second, I show that due to the unexpected nature of the future—including psychological change, transformative experience, and free choice—there appears to be no instrumentally rational way to know and weigh our future interests against our present ones. Finally, I argue that two possible solutions—probabilistic decision theory, and Michael Smith's suggestion that consistency is a constitutive, non-instrumental requirement of ideal rationality—both fail to solve the problem.

### **Chapter 3. The Categorical-Instrumental Imperative**

I argue that a new principle of rationality, the Categorical-Instrumental Imperative, solves the Problem of Possible Future Selves. I show that our present and future selves share interests in solving the Problem, and that these shared interests can only be satisfied if one's present self and future selves cooperate across time to forge and uphold a recursive, universal agreement that all of one's possible future selves can rationally accept given co-recognition of the Problem (in the present and future). I show that this 'universal agreement' with all of one's possible selves is not only intuitive—amounting to a strategy of 'being fair to oneself' and 'not putting one's future in jeopardy' that many of us already implicitly adopt to solve the problem in everyday life. I also show, in decision-theoretic terms, that the Categorical-Instrumental Imperative explains why unfair behavior is instrumentally attractive but nevertheless irrational. Specifically, I show that while the likely personal utility of immoral action may be high in Problem-cases, the overall expected utility of conforming to the Categorical-Instrumental Imperative is infinitely higher, as conformity to that principle can be rationally endorsed and upheld, in the present and future, by all of one's (infinite) possible selves. Indeed, I show that this argument formally verifies Immanuel Kant's famous claim at the conclusion of the Critique of Practical Reason that moral behavior is infinitely valuable, precisely insofar as moral principles reach out into 'worlds upon worlds.' Acting on the Categorical-Instrumental

Imperative is instrumentally rational, in Problem-Cases, precisely because an infinite number of one's possible future selves—across an infinite number of possible worlds—can rationally endorse acting upon it for its own sake. Finally, I detail how the Categorical-Instrumental Imperative recursively applies to itself, and must be utilized to determine which of our future actions it should apply to, thus drawing morality's limits from within.

#### **Chapter 4. Three Unified Formulations**

I show that the Categorical-Instrumental Imperative has several equivalent formulations analogous but superior to Immanuel Kant's formulations of his 'categorical imperative.' First, I show that insofar as our possible future selves can identify their interests with those of other human and nonhuman sentient beings, the Categorical-Instrumental Imperative can be restated in an equivalent 'Humanity and Sentience Formulation' which requires acting in ways that all possible human and nonhuman sentient beings could rationally agree upon. I then show that this second formulation entails a third formulation—the 'Kingdom of Human and Sentient Ends Formulation'—which requires acting on a universal agreement arrived at by abstracting away from the contingent ends of particular human and nonhuman sentient beings. Finally, I demonstrate these three formulations of the Categorical-Instrumental Imperative have advantages over Kantian ethics, particularly Kant's 'Categorical Imperative,' on all seven scientific principles of theory-selection defended in Chapter 1.

#### **Chapter 5. The Moral Original Position**

I show that the Categorical-Instrumental Imperative's satisfaction-conditions can be modeled using a thought-experiment similar to John Rawls' 'original position': a Moral Original Position that requires one to treat the interests of all human and nonhuman sentient beings as possibly one's own. First, I summarize Rawls' original position and discuss Rawls' Kantian, reflective equilibrium, and public reason rationales for utilizing it. Second, I summarize libertarian, feminist, communitarian, and cosmopolitan critiques of Rawls' theory. Third, I construct the Moral Original Position, showing how it models the Categorical-Instrumental Imperative's satisfaction-conditions and corroborates the aforementioned critiques of Rawls. Finally, I argue that the Moral Original Position must be utilized to resolve numerous points of contention between Rawls and his critics.

#### **Chapter 6. Rightness as Fairness**

I use the Moral Original Position to derive Four Principles of Fairness: (1) a Principle of Negative Fairness requiring coercion-avoidance and minimization as ideals, (2) a Principle of Positive Fairness requiring certain types of assistance to others as ideals, (2) a Principle of Fair Negotiation that requires negotiating conflicts between and costs related to the first two principles, and (4) a Principle of Virtues of Fairness that requires developing standing dispositions to conform to the first three principles, and applying the first three principles from such dispositions. I then combine these Four Principles of Fairness into a single criterion of moral rightness—Rightness as Fairness—showing how it reconciles several competing traditional moral frameworks (consequentialism, deontology, virtue ethics, and contractualism), and provides a fruitful new method of 'principled fair negotiation' for resolving applied ethical issues. Finally, I show that this method generates a compelling, nuanced approach to solving a variety of applied ethics—including the ethics of lying, suicide, assistance to others, development of one's natural talents, Trolley Problems, torture, world poverty, organ transplantation, and the ethical treatment of animals.

#### **Chapter 7. Libertarian Egalitarian Communitarianism**

I demonstrate that Rightness as Fairness reconciles three traditionally-opposed normative political frameworks: libertarianism, egalitarianism, and communitarianism. I begin by summarizing the moral attractions and critiques of each framework. I then show how Rightness as Fairness' Principle of Negative Fairness embodies a libertarian ideal of coercion-avoidance and minimization, its

Principle of Positive Fairness an egalitarian ideal of assistance, and its Principle of Fair Negotiation a communitarian concern for context-sensitive (e.g. personal and communal) costs and benefits—thus establishing each framework as containing genuine (but incomplete) elements of moral truth. Finally, I show that Rightness as Fairness requires iterated, ongoing fair negotiation to weigh and balance libertarian, egalitarian, and communitarian concerns (rather than principled argument or divisiveness) to settle, on an ongoing basis, what domestic, international, and global justice require.

### **Chapter 8. Evaluating Rightness as Fairness**

I contend that Rightness as Fairness fares better than existing moral theories on all seven principles of theory-selection defended in Chapter 1. Specifically, I argue that Rightness as Fairness has Firmer (Observational) Foundations than existing theories, grounding moral philosophy not in contested intuitions, but in observable facts about instrumental normativity and moral psychology; greater Internal Coherence, reconciling a variety of traditionally-opposed moral and political frameworks more successfully than rival theories; greater External Coherence, cohering with a wider variety of moral and non-moral facts than existing moral theories; greater Explanatory Power, Unity, and Parsimony, reducing morality to a form of prudence that explains a wide variety of moral and empirical observations, including recent experimental interventions that demonstrably improve human moral behavior; and finally, greater Fruitfulness, providing a more compelling all-purpose method for resolving applied moral and political issues than rival theories.