Reconstructing Reason:
An Investigation of Moral Rationalism and
the Normative Core of Critical Social and Political Theory

By

John D. Lundy

A Thesis
presented to
The University of Guelph

In partial fulfilment of requirements
for the degree of
Doctor of Philosophy
In
Philosophy

Guelph, Ontario, Canada

© John Lundy, March, 2012
This thesis is an analysis of notions of practical reason at the heart of moral rationalism, broadly conceived. In the Western tradition, at least since the Enlightenment, most moral philosophers and political theorists have thought their task was to develop credible alternatives to the traditional justifications of norms, principles, and institutions based on settled tradition or divine authority. The prevailing enlightenment tendency was to insist that the best way to figure out how to live would involve, instead or in addition to other methods, a significant role for some kind of thinking process called practical reasoning. Today we face the dilemma that globalization in the face of diversity has made it increasingly impossible to ignore the absence of any settled tradition which could act as a foundation for moral, social, and political practice, while both modern and ancient conceptions of practical reason have fallen under suspicion (some of it well-founded). I seek to answer the question: under these conditions, can we still expect to find, in reason, a legitimate basis for our social and political norms and practices? I suggest we can.
My approach is bracketed by what I claim to be two prevalent extremes. At one side there is the enlightenment and its scientistic heirs (particularly rational choice theory), and on the other, certain post-modern (over)reactions to the enlightenment failure, which allow rationality to lose its critical normative force. The constructive aim of this work involves drawing on leading contemporary thinkers who attempt to hang on to a connection between practical reason and normativity (particularly John Rawls, Jürgen Habermas, and Martha Nussbaum). They attempt this, I argue, by reacting to, and developing, landmark paradigms in the history of ideas about practical reason, particularly those of Aristotle, Hume and Kant. I then evaluate their attempts and synthesise their insights into a normatively critical account of moral rationalism, which I call the Discourse-Capabilities Approach, an approach that incorporates core insights of Aristotle, Hume, Kant, and their contemporary heirs, while avoiding what I claim are the main defects associated with each of these accounts.
Dedicated to the memory of my mother

Maida Lundy

and for my Grandmother

Sheila Lundy

who has been like a mother to me in her absence.
Acknowledgments

I would like, most particularly, to thank my advisor Omid A. Payrow Shabani along with the other members of my thesis committee, Karen Wendling and Renato Cristi for their encouragement and for helping to make my work stronger through their thoughtful and instructive critical engagements. In addition, I would like to thank my external examiner Simone Chambers for her enthusiastic interest in the arguments I have develop here and for her genuine philosophical engagement with my efforts. Thank also to all my friends and family, and most especially my partner Sabina Naziri, for their support throughout this project. I also acknowledge financial support from of the University of Guelph and the Government of Ontario in the form of the Ontario Graduate Scholarship.
## Contents

### General Introduction
1) Questions and Objectives 2  
2) Irrationality 9  
3) Towards a Reconstructed Reason 14  
4) Cosmopolitanism 19  
5) Overview 23

### Chapter One
A Brief History of Practical Reason I: Ancient to Modern

Introduction 31  
1) Reason from the Beginning: Aristotle 36  
   1.1) Nous 39  
   1.2) Phronesis 40  
   1.3) Eudaimonia 46  
   1.4) Weaknesses of the Aristotelian Approach 52  
2) The Age of Reason 54

### Chapter Two
A Brief History of Practical Reason II: The Enlightenment

1) Hume's Sceptical Reason 68  
   1.1) Hume's Reason: Analyzing the Concept 69  
   1.2) Reason and Conduct: Further Implication of Hume's Theory of Reason 73  
   1.3) Critical Remarks 79  
2) Kant: A New Paradigm in Practical Reason 83  
   2.1) The Unity of Reason 83  
   2.2) Theoretical Reason 84  
   2.3) Kant's Revolution in Ethics 86  
   2.4) Kant's Fundamental Moral Principle 88  
   2.5) Freedom 93  
   2.6) The Possibility of Autonomy 95  
   2.7) Weaknesses of the Kantian Strategy 100  
3) Conclusion 108
Chapter Six
Conclusions: Outline for a Reconstructive Approach to Moral Rationalism

1) Taking Stock
   1.1) Practical Questions & Rational Capacities:
       The Examined Life as the One Worth Living
   1.2) The Rational Will: Lessons From
       The Copernican Revolution in Ethics
   1.3) Capabilities & Grounding Experiences
   1.4) Pluralism & Public Reason
   1.5) Discourse Beyond the Philosophy of Consciousness

2) Synthesis
   2.1) Incorporating Equality
   2.3) Overview of the Outline
   2.3) Question and Answer: Elaborating the Outline
   2.4) A Final Note
A Note on Referencing

All citations are given as endnotes using Arabic Numerals, and come at the end of each chapter. Side notes to the text are found at the bottom of the page and are marked with Roman Letters. The page numbering for Aristotle references are that of the standard Berlin Academy Edition of the Greek text (1831-1870). When referencing larger sections of Aristotle, book and chapters numbers are given rather than page numbers. All page references to Kant refer to the standard Royal Prussian Academy* Edition of the German text. *(a.k.a. the “German Academy” (Post 1946) and the “Berlin Academy” (post 1993))
Human Madness is oftentimes a cunning and most feline thing. When you think it fled, it may have but become transfigured into some still subtler form...not one jot of Ahab's broad madness had been left behind; so in that broad madness, not one jot of his great natural intellect had perished. That before living agent, now became the living instrument. If such a furious trope may stand, his special lunacy stormed his general sanity, and carried it, and turned all its concentrated cannon upon its own mad mark; so that far from having lost his strength, Ahab, to that one end, did now possess a thousandfold more potency than ever he had sanely brought to bear upon any one reasonable object.

—Herman Melville, *Moby Dick*
General Introduction
This introduction has three principle aims: First, I offer a general characterization of the basic purpose, motivation, topics, and objectives of this work (covered largely in Section 1) as well as elucidating some broad features of the particular sort of approach I recommend (begun in Section 1 and continued in Section 4). Second, I preface my discussion with some preliminary arguments, particularly about irrationalist approaches and the ways in which they are antithetical to the moral rationalism developed in this thesis (Section 2) and about the cosmopolitan nature of the moral rationalism I seek to develop (Section 4). Third, I will offer a specific outline of the structure according to which I unfold this investigation (Section 5).

1) Questions and Objectives

I find our human, social world to be a rather irrational place. I take this to be a particularly troubling fact since, I believe, a decent world, a just world, would mean a more rational one. Here, in this thesis, I want to investigate that suspicion. It has been my hope that movement toward a better, more rational world is possible. My guiding motivation here has been to discover whether that hope is well founded and how, in the first place, it might be legitimate to criticize current practices and institutions as irrational. Ultimately the issue here is whether or not the practical moral concerns of life are amenable to illumination on the basis of some human capacity called reason.

In philosophy, in fact, I believe it would be fair to say in human life generally, there is really one central question: “what do we do—how should we live?” It’s a question that must always be answered in a practical sense—that is, we must act.
At bottom, this thesis will be about how to answer this most important and fundamental of all questions. Methods for deciding what to do could include following tradition, following some authority such as a powerful ruler or someone who claims to speak for a God or magical force, or possibly following unrestrained, primal, animistic instincts created over the course of our evolutionary history and coded for in our genes. The profoundly influential tradition called philosophy has tended to insist that the best way to figure out how to live would involve, instead or in addition to any other methods, a significant role for some kind of thinking process called practical reasoning.\(^A\)

This thesis is an attempt to tell the story of reason's connection to ethical values—the things we judge to be worthwhile in life—albeit a necessarily selective and abridged version of it. It is a story, some version of which has no doubt been told many times before; but never, in my experience, in a way that seemed wholly satisfying. Given its importance, it is a story we should never stop telling and trying to get right. I want to defend the notion that once humanity has broken free from superstition and the power of authority, the best hope for ethics is reason. This is

\(^A\) According to the standard analytic view, theories of practical reasoning are theories about how we figure out what to do, which can be contrasted with theoretical reasoning that is concerned with figuring out how the facts stand. While this view may recognize that there is certainly some overlap or interplay between the two (for example, to make decisions about what to do we likely need an accurate description of the facts of the situation we are in) this view does depend on a significant theory/practice, fact/value distinction—which is shown to be problematic by, for example, American pragmatism. If theory is necessarily value laden and therefore, on par with other practical activity then William James seems to be justified in claiming that the rationality criterion should be the same for beliefs and actions. The conventional view would also tend to have it that we could think of the principles of substantial moral, social, and political theories as the product or output and practical reason as the process. The investigation of practical reasoning is seen as more concerned with theories of how to arrive at principles (process) than the particular form of substantive moral principles (product). It may be, however, that product and process will not separate so neatly. These issues will resurface later in the thesis. (on what I call the standard view see chapter one of Varieties of Practical Reasoning, ed. Elijah Millgram (Cambridge Mass.: MIT Press, 2001)
why most fruitful social and political theory has looked here to find its normative core. By outlining major developments in the tradition of practical reason along with certain critiques of this tradition, I will first lay out the central problems occupying the rest of this thesis and clarify what I see as the horizon against which these issues are set. In doing so it is not my intention to contribute to any particular scholarly debates on the interpretation of these historical figures. Rather, I endeavour to tread a selective path, discussing general trends and major developments of thought which are important for understanding where we need to start from today if we are to make sense of ethical value.

In the Western tradition, at least since the Enlightenment, moral philosophers and political theorists have thought their task was to develop credible alternatives to the traditional justification of norms and principles based on settled tradition or divine authority. I will set out the terrain of my discussion by giving a broad overview of certain major, relevant developments in this tradition from ancient, through to modern and later modern developments. This will include the Platonic/Aristotelian view that to act ethically is to express our true nature and

---

A Relatedly, at this point I should also like to note that in this paper I am neither interested in, nor always fully qualified for, speculation about which interpretations may be more correct, that is, which interpretation most accurately and sympathetically thematises these thinker's work; rather, I am concerned with assessing the correctness of the actual views expressed according to the interpretation presented. That is to say, for example, I would simply argue that, viewed in a certain light, or focusing on certain aspects of some postmodern theory, it may appear to aim at a genuine reconstruction of reason (which I support) rather than pure deconstruction/abandonment (of which I am sceptical), but I will do so without the intention of claiming one interpretation is better than another as an interpretation.

B I will often speak of moral and political or social theory together, or interchangeably, since, although there may be times when it is necessary to make an analytical distinction between them, they must be seen as completely bound up with one another, since we cannot adequately think about what we should do outside our social and political context, that is, without considering our relation to others, and likewise since there's no way to evaluate social and political issues that does not lead us back to moral/practical questions about how we should live.
unique essence—which is reason—as well as the Kantian view that human freedom is expressed by following the categorical demands of our true rational selves. It will also include the influential family of views, most famously associated with Hume, called instrumental or means/end reason. Instrumental reason is about determining the most effective, efficient, and consistent way to achieve a set goal or given desire. Here, the rational act is defined as the one that actually leads to the desired goal. If one believes in the legitimate exercise of any form of rationality this will almost assuredly be one of them. However, many thinkers have tended to argue that this is the only possible form of rationality, meaning goals and desires are simply given—that is, inscrutable by reason.

The big question that remains is whether and how reason can go beyond instrumentalism, for, if this is all reason amounts to, I maintain that there is good cause to suspect that rationalization will lead to bureaucratization as described by Max Weber as social coordination, regularity, codification, systematization, predictability, and determinability by systems of accounting, notation, measurement, etc. likely purchased at the cost of actual human freedom.¹ Purely instrumental reason entails that all goals are based on nothing beyond subjective whim. Standing in no need of justification, ends are decided monologically, without reasons, beyond the reach of critical reflection, incapable of being challenged and revised. The worry is that such a reason will be reduced to the efficient mastery, regulation, and control of both nature and other individuals in order to realize unreflectively taken-for-granted personal or cultural goals.
I will further my elaboration of the horizon against which the contemporary investigation of forms of practical rationality must be set by presenting main currents of thought, inaugurated, most famously, by Nietzsche and culminating in the cluster of outlooks which are sometimes grouped under the notoriously slippery heading “postmodern,” currents which critique the tradition of rationality—a tradition stretching back to Socrates, the figure Nietzsche pinpoints as the beginning of the great decline of civilization. I will consider ways in which the tradition of rationality\textsuperscript{A} has been problematized. Many thinkers, from Hegel and Marx onwards, have stressed our cultural embeddedness in particular social relations. Weber and Frankfurt School critical theory showed the danger inherent in overstating cognitive virtues—equating good living with good theorizing to the detriment of variation, emotion, and individuality. Thomas Kuhn’s analysis has paved the way for an emphasis on the fact that what counts as a good explanation at one time, can change (often suddenly and dramatically) relative to historical context. Alasdair McIntyre and Jean-Francois Lyotard emphasize the degree to which standards or criteria are culturally relative. I agree that the idea of a single scientific method for knowledge (\textit{a la} Descartes), formal scientific algorithms and exact languages do not withstand sustained critique—nor does the particularly narrow internal consistency or self-

\textsuperscript{A}From the same Latin word \textit{ratio} we get the modern forms “reason” and “rationality.” In ordinary English speech people use these words as more or less synonymous. Applied in the practical domain of action both words describe something along the lines of acting in accord with well-founded beliefs or principles. For this reason I do not consistently distinguish between “reason” and “rationality.” Unless explicitly stated they are considered synonyms. Of course some theorists operate with more technical definitions according to which the terms are use to make a precise and consistent distinction of some kind. Rawls, being an important example of this in the context of my present investigation, will necessitate the use of a distinction, but I shall be clear when I am using the terms in a more technical sense.
interest maximizing accounts of rationality still offered by the un-self-reflective
disciplines of neo-classical economics or so-called “rational choice theories.”

In spite of this, I will go on to argue that the wholesale critique of modern
rationality detectable in thinkers such as MacIntyre and especially Richard Rorty—a
critique which has inspired a pervasive mood of relativism amongst much
academia—is unsatisfactory. Indeed this pervasive mood can have dangerous and
destructive consequences, since, insofar as it gives up on universal normative
standards, it lacks the resources to adequately criticize unjust social and political
arrangements and actually prevents or deflects genuine criticism of such
arrangements. I argue that criticism of rationality, insofar as it represents a
rejection of what reason has become and offers cautionary tales about the
pathologies of excessive intellectualism and the pretensions of final vocabularies
and grand metanarratives, may be instructive; but, insofar as it becomes a wholesale
rejection of the possibility of deriving norms from rational inquiry, they must be
rejected. I maintain that there is no alternative to criticism but more criticism.

Having characterized the current situation in terms of a conflict between the
broadly defined rationalism of traditional theory and the equally broadly defined
anti-rationalism of post-enlightenment thought, and finding both less than wholly
sufficient, I will go on to survey the contemporary intellectual climate highlighting
thinkers and traditions that recognize the fact that, while philosophy may not be
able go on as it has, it cannot be abandoned. I do not claim to present an exhaustive
overview nor do I feign a comprehensive expertise, but seek to characterize a
climate in which I see fruitful attempts being made.
The motive for this survey is largely to support my claim, not only that any normative theory that deserves to be taken seriously now needs to be grounded in reason, but also that most credible theories on the ground meet this criteria. There is a danger of being led astray by an inclination to overextend a polarizing view that sees a simple and pervasive reason/unreason division. Clearly, I do not want to deny that some views actually offer a wholesale critique of reason, but I want to stress that once these proposals are rejected a great deal of controversy remains. Since reason of some sort now plays a key role in a great deal of promising moral and political theory, the real question must move beyond simply: “Reason or unreason?” It should now become: “What is reason and how does it function within normative theory?” Further, however, I want to claim that any theory that deserves to be taken seriously needs to be responsive not only to the tradition of reason but also to many of the critiques of existing conceptions of reason. I suggest that thinking about the normative core of any social or political theory now needs to take place against something like what Richard J. Bernstein calls “The Ethical-Political Horizons of Modernity/Postmodernity” and this will largely take the form of answering the question: what is reason and how does it function as a basis of normative theory while being sensitive to the partial critiques of reason?

After my overview of the history and current state of debate the stage is set for me to unfold my proposed critical normative framework for a social and political theory based in reason—an approach which integrates what I see as the unique strengths of several different theories into a coherent perspective. Certain shortcomings of particular theorists within the tradition will be corrected by
combining insights from several different sources into a single moral framework, which I will return to presently. First, however, before I turn inwards to the development of an account of practical reason, I would like to devote a few preliminary remarks to the irrationalist approaches, the position they occupy beyond the margin of the broad family of views I have called moral rationalism, and the persistence of a generally rationalist tradition against these alternatives.

2) Irrationality

Today, it is not unusual to claim that reason is dead. Under the sway of Nietzsche's suggestion that reason is essentially just another version of God, the postmodern intellectual tide that has swept through academia and permeated the fabric of contemporary cultural life, has subjected reason to the same fate inflicted upon God, according to Nietzsche's bold proclamation. Many postmodernists\(^A\) are happy to embrace the corollaries of this death. They tend to be sceptical of foundations to human existence, grand historical narratives, concepts such as truth, objectivity and progress, and of all norms, universals, and authority. Instead, they tend to celebrate pluralism, discontinuity and heterogeneity, seeking political potential purely in the marginal and transgressive—the mad, deviant or perverse. Norms, like those of mainstream social life or any other form of authority, are seen as inherently oppressive because they force difference and novelty into uniform moulds.

\(^A\) Examples include a range of thinkers strongly influence by this strand of thinking in Nietzsche, beginning, perhaps most influentially with Martin Heidegger. More recent prominent examples include Jean-François Lyotard, Jacques Derrida, Richard Rorty and others.
One area I want to address is this tension between the long tradition which viewed human reason as the way to discover just norms, and the reactions against this tradition. Although some challenges require rethinking previous notions of reason I argue in favour of what could be called a revised, reason-based approach. The basic idea here is an old one, an idea which has been presented in many forms and faced many challenges. Post-enlightenment variants of historicism and contextualism which expose problems with the enlightenment rationalist ambition to find absolute grounds for certainty and objective truth, has led to a disenchantment with reason as a source of hope and beacon for moral progress. If reason is relative to particular languages and forms of life then there appears to be an unbridgeable gap between contexts that cannot be intelligibly translated into terms that make sense intercontextually. These kinds of worries have bred a pervasive fear that any talk of universal value will equate with conservative, Western, male, domination, and that universal reason is nothing more than an instrument of oppression and control embodied in scientism, capitalism, culture industry, and bourgeois political institutions. On this picture, a belief in the capacity of reason to lead us to just norms and better lives is naïve at best, and at worst a violent negation of difference and particularity. Counter to these suspicions I maintain that there could be no celebration of difference and the particular without the idea of a universal—in the form of universal human rights for example—that could encompass and safeguard this diversity. I submit, then, that such concerns call not for abandonment but rather a refinement of our understanding of reason. I suggest that a great deal of postmodernism, ironically, has become the ideal
theoretical counterpart of advanced global capitalism/neo-imperialism and the horrors it continues to wreak on humanity.

For these reasons I lament rather than celebrate the death of reason. I seek to join those who wish to breathe a new life into the concepts of reason, normativity and the possibility of both genuine progress and the legitimate use of power.¹ I seek to discover how reason can be put to use by unleashing its latent revolutionary/emancipatory potential for amelioration of humanity’s overwhelming condition of poverty, misery, alienation and truncated development. I wish to reconnect to a critical impetus I see in figures such as Marx and Marcuse, however I say “breathe a new life” rather than “revive” to emphasize the need to genuinely break through contemporary intellectualism, wallowing in the now decades old stagnating regurgitants of postmodernism, rather than simply following thinkers who would have us simply retreat to options now closed to us—viz. ways of understanding reason and doing philosophy that have not withstood scrutiny.

Here it becomes evident that the topic of this thesis, reason, the traditional instrument of philosophical investigation, strikes at the heart of philosophy’s very self understanding. In the wake of the assault on reason the world of philosophy has become a rather strange place (or at any rate stranger in some sense than it was before). In other academic disciplines such as history or biology, despite a wide range of topics and diverse methodological approaches one is struck with a sense that everyone is engaging in a roughly similar sort of endeavour. There is usually a substantial amount of agreement about, at the very least, what the discipline is and

¹ There are many examples here, the most inspirational examples of which this thesis will treat in detail: John Rawls, Jürgen Habermas, Martha Nussbaum, etc.
what sorts of things count as legitimate work in that field. The nebulous and
disparate world of philosophy, by contrast, is relatively far from any agreement on
what sorts of things philosophers can or should be doing, and this lack of common
mission has shown no signs of resolution. In fact, it is quite possible that the
biologist in his lab and the historian on the other side of campus have much more in
common with each other in their work than a thinker in a philosophy department
has with their colleague in the office across the hall. In part this is simply because of
a general fact about academia: that traditional disciplinary boundaries are largely
artificial delineations that often act as barriers to the free flow of investigations
driven by actual fundamental questions, but it is also because, philosophy is
particularly divided on fundamental philosophical question such as: what is
philosophy?

For some, philosophy continues to operate much as it always has from Plato
to Kant. They tend to operate on the basis of things like the pure platonic notion of
logos, indubitable self evident givenness, a priori necessity, metaphysical
foundations, and sovereign, atomistic, disengaged, self-transparent, rational
subjectivity. But, there are fewer and fewer of these thinkers left in either
continental or analytic circles owing to the critiques of reason and rational
subjectivity that have proliferated over the past century and pulled the rug out from
under traditional philosophy, leaving philosophers scrambling for a place to stand.\(^{A}\)

But there is much more at stake here than whether or not a handful of academics

\(^{A}\) For more on this see Terry Eagleton's *After Theory* (Toronto: Penguin Books, 2003) and the
collection edited by Kenneth Baynes, James Bohman, and Thomas McCarthy *After Philosophy: End or
Transformation?* (Cambridge Mass.: MIT Press, 1887)
will be able to secure grant money and carve out a comfortable home in the academy. The debate will have to be played out at a wider political and cultural level, for what is at stake is whether or not there can still be a distinction between truth and falsity, right and wrong, justice and injustice—distinctions that go beyond what individuals happen to be seduced into accepting as such at a given time. What is at stake is whether or not there is any ground for mediation of cultural contestation, whether or not progress is an illusion, and whether or not human existence is a futile passion. I can conceive of no other question or contention that could have a greater impact on our lives as individuals, our culture, or the fate of humanity.

There are those, and their influence runs deep in academia as well as popular global cultural forms, who say reason, and with it philosophy, has outlived its usefulness. People often call them philosophers but they tend be consistent enough not to self apply such terms, often not even considering themselves theorists but engaging in something much more like a practice. Leading proponents of this post-rational abandonment of philosophy may include Jean-Francois Lyotard, Richard Rorty and Jacques Derrida. But not all who accept the reality of human finitude, our inexorable situatedness in historical/cultural contexts, and the nexus of power, human interest, and knowledge, have seen the need to go so far as to completely abandon some notion of reason which would allow these facts to be partially transcended. But what would it take to reconstruct such a reason? What could this new reason look like? What role could it play in our lives and the lives of cultures
and political institution? These are foremost amongst the question that contemporary moral rationalism now needs to ask.

3) Towards a Reconstructed Reason

I argue that an account of practical reason, adequate to the objectives of moral rationalism, should be able to incorporate key insights from the major theoretical developments of Aristotle, Kant, and, to some extent, Hume. While no single contemporary theory seems fully adequate, I argue that several formulations go considerable distance towards incorporating the insights I identify. Ultimately, I will outline a proposal which aims to incorporate the most important insights of various previous accounts of moral rationalism, specifically by drawing on the Neo-Aristotelianism of Martha Nussbaum and the Neo-Kantianism of John Rawls and Jürgen Habermas.

I begin with Amartya Sen’s capabilities approach which overcomes the limited focal variables of previous theories of justice by broadening the focus to include any valuable states of being and doing—the constitutive elements of a human life. But, where the capabilities approach relies on a humanism that tends towards Aristotelian metaphysical, teleological, essentialism, I suggest moving towards a different sort of humanism, inspired more by critical theory and related to what has been called a science of human nature. A science of human nature is different from metaphysical essentialism in the sense that a theory of human nature is its aim not its premise. In other words, it is a method, not a final answer. And, like other sciences, the models which are its outcome, in this case models of human
nature, are fallible and open to revision and qualification in light of new evidence and explanations.

For Aristotle, as with Plato, it is not only that we have a responsibility to use reason because it is an effective way to discover what is worthwhile in life or because reason is an effective way of getting knowledge and knowing the truth about things that concern us will generally aid us in bringing about results that are desired; rather, this view is that we can indeed use reason as a means, but what we discover in this process is that reason itself is the ultimate end. Reason is not only a tool for discovering or bringing about the good life—the life of reason is the good life. Reason is the method or foundation of ethics and at the same time its final outcome or content.

I suggested moving away from these heavily substantive accounts of moral rationalism towards a more fallibilist and procedural conception of practical reason that would provide the normative transformative/revolutionary capacity for contestation, critique, and therefore moral progress, without the metaphysical baggage that comes with a fixed/thick/substantive conception of reason that prescribes a particular set of moral values such as the supposedly ultimate value of reason itself, which, would seriously stand in the way of the cultural translation that effective world citizenship requires. This view of reason does not guarantee full agreement or mutually satisfactory outcomes, but that is a strength of this view not a criticism, for any theory to promise such agreement indulges in fantastic daydreams. Global intercultural relations will never resemble a group-hug but they
need not resemble the bloody battlefield of perpetual war so long as we can create and maintain spaces for rational public debate and negotiation.

However, in the absence of an authoritative final answer, how, it must be asked, are we to deal with the wide range of disagreements about who we are and derivatively about what should count as valuable human capabilities? The answer, I propose, can be found in the science of human nature itself. The specific version I draw on to find this answer is what Habermas calls “reconstructive science.” Reconstructive science works similar to transcendental argument in the sense that Habermas reasons from the fact that we have social and communicative interaction to the preconditions necessary to make that possible. Those conditions are certain basic communicative competencies—essentially our ability to reason and engage in discursive justifications. Our model of human nature and even reconstructive science itself is meant to be revisable, fallible—continually open to challenge or contestation that must be reasoned out in discussion, as opposed to any final a priori foundationalism such as actual transcendental argument. So, the short answer is, although disagreements are certain to persist, the only way to deal with disagreement is through further dialogue. Further, however, in order for there to be a free, fair and open dialogue I follow Nancy Fraser in claiming that individuals must be free to participate on par with others in social life—what she calls participatory parity. I also follow Fraser in claiming that certain conditions, which can be understood as well-being achievements, must be met in order for participatory parity to obtain.
In these respects, I maintain that an approach broadly in line with recent, communicative-turn developments in critical theory will offer particularly fertile ground, owing to the fact that this tradition has tended to A) identify and criticize problems arising from a traditional theory conception of reason B) refuse to abandon an essentially constructive enlightenment project that bases norms on reason in favour of wholesale critique C) take seriously the challenges and alternatives of other reconstructive proposals, and D) have as its aim, the practical emancipatory goal of human development or self-realization. Specifically, I will argue this tradition is the best place to find resolution to the paradox of reason. This paradox is captured in the opening lines of Horkheimer and Adorno’s *Dialectic of Enlightenment*: “the Enlightenment has always aimed at liberating men from fear and establishing their sovereignty. Yet the fully enlightened earth radiates disaster triumphant.”3 This paradox is at the very heart of a great tension in human development which I will characterize. Indeed, the result of enlightenment rationality, overwhelmingly, has been to place us squarely in the iron cage described by Weber. It is no small wonder then that so much industrial-capitalist-age thought can be characterized as a rage against reason.

I will argue that Habermas’ central achievement was to set us on the road to resolving this paradox by articulating and disentangling the different forms of rationality operative in human social forms of life. I suggest, also, that the disasters of bureaucratic capitalism are largely the result of an unchecked proliferation of the internal logic of systems which operate according to an instrumental rationality that aims only at effective and efficient mastery and control. We can only criticize the
colonization of the lifeworld by instrumental systems rationality from the standpoint of a communicative rationality oriented towards achieving understanding. I will argue for the central significance of a communicative rationality and consider the claim that communicative rationality is in fact a universally valid rationality latent in the formal pragmatic structure of speech, which can fulfill the Marxian goal, in the sense that, this form of reasoning is meant to be “both immanent (not to be found outside of concrete language games and institutions) and transcendent (a regulative idea that we used to criticize the conduct of all activities and institutions),”⁴ (to borrow a sharp and succinct phrase from Hilary Putnam).

In light of this understanding it may be clearer what I had in mind at the outset, when I called the world irrational; I mean that individual and collective social action is often determined by unquestioned traditions and is coordinated through the use of violent oppressive force or through coercion, deceit and manipulation. Development, human emancipation, justice—would entail finding ways to direct and coordinate our action not primarily through the power of money, armies, or authoritative tradition, but rather on the basis of freely and consciously achieved agreement which recognizes no force other than reason, a force that can prompt us to change our beliefs and actions through argument, discussion and genuine understanding. Human beings clearly do have a technical interest in controlling the world around us but also a practical interest in understanding each other and an emancipatory interest in freeing ourselves from structures that dominate us by restricting our capacity for rational autonomy. Treating social
actors like any other physical object to be manipulated and controlled is incompatible with the freedom and flourishing of humanity. When we instead attempt to coordinate our action through understanding we implicitly take on the burden of redeeming our claims with respect to the reasons we have for them, and the only legitimate way we can come to share these reasons is through rational discourse and continued interaction.

Before concluding with a specific outline of the structure of this thesis, I want to make a few remarks about the cosmopolitan nature of the project.

4) Cosmopolitanism

If morality is based in reason and universal, then, in a globalized world, this must imply a version of cosmopolitanism. “Globalization cannot be limited to the political, economic and cultural domains but must above all be a normative enterprise: ‘if ethics is to function for the well-being of all, it must be indivisible,’” as Fred Dallmayr puts it. Critical and decisive challenges that potentially affect us all, such as: environmental damage, poverty, natural disasters, pandemics, serious conflicts, both armed and unarmed, as well as attitudes, psychological conditions, and patterns of individual behaviour which are potentially in tension with a form of life conducive to human flourishing and which are shaped by a truly global consumer culture, cannot be addressed if free rational discourse ends at state borders and is replaced by violence and coercion.

In the last thirty years the human world has realized a state of profound integration. Even the nation state itself is becoming increasingly less dominant in an age where technologically sophisticated global communication and transportation
networks have made it easy for the material of culture to flow freely across national boundaries. The human world has become astoundingly fluid, diffuse, hybrid, and permeable. This process of technological and socio-economic integration and intercourse is continuously forcing us to stand face-to-face with the other. A multiplicity of ethnicities, ideologies, orientations, cultures and subcultures are being thrust together at an unprecedented rate and in unprecedented ways, leading both to fruitful hybridity and, all too often, to violent or destructive contestation. If there is to be any significant degree of human flourishing, peace, or stability under these conditions it will require finding ways of getting along. Like it or not our lives are now connected to a global economy not only of money and material but of culture. Such a condition demands some version of a cosmopolitan ethic.

Many thinkers, albeit to various degrees and sometimes for different reasons, suspect the general principles of such cosmopolitanism are necessarily too abstract to actually bind people together in any practically meaningful sense and are too “bloodless to capture the moral imagination.” Our primary allegiance, they argue, must be to some stable cultural community, what Michael McConnell calls one of the “great traditions.” According to Richard Rorty there is no way to escape these traditions—the mere fact of human finitude means we are completely bound to tradition. The appeal here is always to some pre-political, pre-theoretical, pre-transcendental notion of ‘a people’—a primordial common denominator of ‘natural’ cultural groupings according to some combination of shared history, geography, tradition, custom, ethnicity, language, religion, bloodline, race or nation. The communitarian insistence that the particularities of our situatedness—what
Gertrude Himmelfarb calls "the givens of life: parents, ancestors, family, race, religion, heritage, history, culture, tradition—and nationality,"8—are genuinely constitutive of the individual, is true, if it means simply that we are historically and culturally positioned, that we do not choose the platform, the 'here' from whence we choose—that everything is, in some broad sense, culturally inherited. Of course this is true, we are not gods but socially and a spatio-temporally situated beings; but, situatedness does not entail a full-blown materialism. The cosmopolitan sceptic has an empirical claim that various exclusive primordial groupings have tended to determine allegiance and solidarity in the past and that we are all necessarily beings situated in some history and are therefore defined by certain allegiances, but, these empirical claims do not add up to a critique of the normative project of cosmopolitanism, which is pre-eminently concerned not with how things have been but with how things can and should be.

Further, the idea that each natural group might claim statehood and live autonomously has simply become a practical absurdity, not only because the idea of natural groups is becoming obsolete in a world where most people's identities are comprised of numerous cross-cutting allegiances, but also due to the fact that we are forced to share a biologically interconnected natural world, a geographically finite planet with limited resources, and a profoundly interconnected cultural and economic world. Structural conditions of our existence in a finite, diverse and interconnected world place on us the practical demand that we find some associative norms that can govern or provide a framework to negotiate the
coexistence of difference. It is difficult to imagine how we could be expected to respond to crisis on a global scale outside such a framework.

Promisingly, the re-imagining of belonging and membership is already well underway in practice. Many modern multicultural countries (including Canada) are held together less by the bonds of cultural and national identity than simply legal and political bonds embodied in a constitution. But, even a world comprised of multicultural liberal-democratic constitutional states, faces tension, crises and contestation that transcends state boundaries and threatens the possibility of human flourishing and the peaceful coexistence of difference. States are not hermetically sealed, and the fact is, the most pressing challenges we face are common threats to humanity which do not recognize national borders—environmental damage, poverty, war and natural disasters. Thus, I suggest that just as citizens of multicultural constitutional states have reconceived belonging to include the new bond of political association, it is now incumbent upon us to recognize that justice also demands that we push our understanding of belonging even further.

In exploring the possibility of a postmetaphysical humanistic ethic, then, I endeavour to be sensitive to the contemporary condition of a diverse yet highly integrated world. In other words, what I shall endeavour to develop is a cosmopolitan ethic which can maintain both normative/ethical recourses (necessary if we are to adopt ideals to define development and guide action) and eminent/materialist resources (necessary to diagnose current crises and future potentials for change) that together allow it to offer substantive guidance for
concrete decision-making processes most appropriate to the contemporary global context of a world which is politically, increasingly post-national, and philosophically, increasingly post-metaphysical—that is to say, a world that can no longer pretend that a shared moral framework based on either a philosophical or religious outlook that prescribes action based on a claim to possess final knowledge of the ultimate nature of reality, or the history or tradition of a particular group, is necessary or desirable with respect to the task of responding to the crises we jointly face as human beings.

5) Overview

Finally, I will now provide a specific yet concise guide to the structure of this thesis. In the first two chapters I set out to introduce the tradition of Moral Rationalism, broadly conceived, and analyze what I see as some of the most groundbreaking developments in its long and varied history. Later, I will go on to show how the most significant contemporary attempts to develop an account of practical reason are grounded in the approaches presented here. In chapter 1 I introduce the occidental origins of moral rationalism in the ancient Greek tradition, by focusing on the work of Aristotle, arguably the most influential and philosophically sophisticated outcome of this tradition. While offering criticisms of the form and function as well as the content and conclusions of this approach I try to present, as forcefully as I can, what I see as critical innovations of the Aristotelian approach, including core claims which have come to define the tradition I include under the heading moral rationalism. Later on I will consider how these developments might be retained in a contemporary theory of practical reason. In
chapter 1 I also introduce the enlightenment as an age of wide reception, deep influence, and profound proliferation of new and diverse ideas about practical reason.

In chapter 2, I will consider, in greater detail, two of the most significant and influential developments of this period, represented in the work of Hume and Kant. While I am sharply critical of Hume for taking a disastrously narrow view of practical reason (most explicitly, in this thesis, when it comes to my consideration of the 20th century influences of his approach) I strive to present his formidable arguments in a way that brings into relief some of the more sophisticated insights of his analysis of instrumental reason, which I argue need to find a place in any adequate account of moral rationalism. Next, while attempting to show the force behind some of the most significant arguments facing Kant’s moral rationalism I strive, also, to reveal the strength of the arguments Kant gives for several of the core aspects of his innovative approach to practical reason. An important concern in later chapters will be the attempt to incorporate these Kantian insights, while avoiding the Kantian mistakes, both of which I characterize here.

In chapter 3 I turn, first with a critical eye, to the contemporary state of moral rationalism. Here I will show the extremely inadequate condition of a great deal of current thinking in this area. I argue that a particularly influential sort of moral theory and theory of political economy associated with mainstream economics, has grown out of a particular Scottish/British tradition, central to which is an account of practical reasoning influenced largely by Hume. Drawing particularly on the theoretical framework of critical theory, and the arguments,
within economics itself, of Amartya Sen, I suggest that the dominant account of practical reason, given by so-called “rational choice theory,” is seriously flawed.

At this stage I will also challenge an influential range of reactionary views which represents one of the leading alternatives to the dominant account. I will also attempt to show that, notwithstanding the failures of the mainstream account, the core hope for a humanist ethic based in our shared capacity for reason can be defended against the radical critiques of this tradition. The main target of my discussion here will be Friedrich Nietzsche, likely the greatest pioneer and continuing source of inspiration for this tradition, as well as Rorty, one of the clearest and most forthright contemporary defenders of a Nietzschean critique. Having revealed some of the serious problems with mainstream accounts of practical reason, along with the poverty of a wide swath of irrationalist leaning alternatives, I argue the only viable solution involves reconnecting, in some way, with Aristotelian and/or Kantian accounts. From this point on I will be considering thinkers with whom my critical engagement will be primarily constructive in nature.

Finally then, this chapter will consider neo-Aristotelian approaches to practical reason. First, I briefly argue that the way certain communitarian-leaning thinkers, specifically Alasdair McIntyre, have taken up Aristotelian themes, entails the forfeiture of a normative core based in reason and lands these interpretations in a camp rather close to the explicit irrationalists. I then explain how the particular sort of neo-Aristotelianism, from which I gain constructive inspiration, is one which continues and develops the Aristotelian tradition of moral rationalism. Specifically
this will involve a consideration of Martha Nussbaum’s version of the capabilities approach. I argue that Nussbaum develops some of the most promising strands of Aristotle’s thought. In particular, I will examine how she develops the idea that the human good consists in our capability to function with respect to various universal facts of human existence in a way that allows us to realize a fully-human existence, one that fulfills the positive potentials of a rational being. However, I also suggest here that defending a normative account of a properly functioning human being, which can act as a stable foundation for defining virtue, proves problematic.

Continuing the search for inspiration, Chapter 3 and 4 turn to a consideration of certain important, contemporary attempts to articulate a Neo-Kantian account of moral value derived from practical reason. These are represented in the work of Rawls and Habermas respectively. Rawls reconstruction of Kantian rationalism is particularly important because he avoids the mistakes of Kantian formalism and articulates a way of incorporating consequentialist intuitions associated with instrumental reason while retaining the key conclusion of Kantian moral rationalism—that human beings qua rational actors, have a unique dignity which gives rise to categorical normative demands to respect their freedom and fundamental moral equality. Further, he recognizes the fact that the deep ethical disagreements, which are the outcome of individuals not-irrational worldviews, means that we cannot rely on a full ethical harmony realized on the basis of a single normative law that resides in the will of every rational individual. In spite of this recognition he holds out the Kantian hope that human affairs can be coordinated on the basis of shared reasoning; this, for Rawls, is public reason.
Rawls’ theory of practical reasoning explicitly incorporates elements which rely fundamentally on particular culturally inherited concepts which, I claim in chapter 4, make this account insufficiently critical. In chapter 5 I show that Habermas is particularly important for his concept of communicative reason which offers a universalist account of the sort of fundamental moral concepts Rawls takes from the public political culture of Western democratic states. In Chapter 6 I will propose that the necessary content which Rawls takes from culture, Nussbaum imports with a normative account of the person, and Habermas sees as a necessary presupposition of communicative action, is a single principle stipulating the fundamental moral equality of persons.

Finally, Chapter 6 will go on to propose a framework which seeks to incorporate the key insights of various previous accounts of practical reason, including fundamental moral equality. The synthesized view of practical reason presented in this framework has the following key features: 1) It includes instrumental reasoning, which is concerned with achieving consequences deemed valuable, as well as a capacity to reflect on our own nature and our ends or values. 2) It is public in the sense that the reasons given to justify decisions affecting other rational actors should not appeal to the private ethical self-understanding of any individual or group, but should appeal to what all parties affected by the decision could be expected to accept, on the basis of reasons, as being in the public good. 3) It is communicative, in the sense that practical reason is neither a formal operation nor a solitary endeavour. To say that specific normative principles or values are derived from reason means that they are the outcome of a social process. 4) And
yet, communicatively rational social processes are underwritten by certain presuppositions, beyond logical and semantic principles of reason, this includes the presupposition that human beings, as rational, should be treated as free and equal—a presupposition which introduces universal moral content to practical reason. 5) Moral discourse is also grounded by certain shared experiences in virtue of various universal aspects or conditions of human life. This thin account of human nature serves to concretize the sort of questions that need to be raised in the reflective procedures of practical reasoning; questions concerning the nature of good or virtuous functioning with respect to the various facts of human existence. Functioning achievements in these areas also turn out to be conditions for the participatory parity which makes dialogue rational.
Notes: Introduction

7McConnell, 80.
Chapter One

A Brief History of Practical Reason I: Ancient to Modern
Introduction

Reason itself is surrounded by unreason. It is bracketed on either side by two different variants of unreason. To the one side is the oldest kind of unreason—uncritical unreflective pre-philosophical unreason. Before the idea that human beings possessed some capacity to figure things out for themselves, however that is supposed to work, there was no way to determine, among other things, what we should do with our lives (ethics/the good life) or how we should coordinate our activities with one another (justice/the good state or society) without blindly submitting to some external power. Without philosophy, mysticism and superstition often went unchecked by argument—might and convention had to go substantially unopposed by right.

Without some clarification, I recognise that this characterization may strike some readers as a rather polemical valorisation of theory (and coming from a theorist perhaps a self-aggrandizing one at that). I qualify my account, then, in two important ways. First, I want to be clear that I am not claiming that the rise of philosophy as a field of study in ancient Greece or anyplace else was a sudden, epiphanous turning point representing a qualitative shift in human nature or human society. Quite the contrary, it is important for my overall argument (as I hope to make clear later) that the ability to problem solve, to discuss, to be persuaded by argument and evidence are innate capacities of our highly intelligent and social species which predate Athens. What I am claiming is that the Athenian school
represents a great, perhaps the greatest ever, quantitative leap forwards, towards
dialogic method and rational reflection and away from unreflective superstition and
traditionalism when it comes to answering questions about how we ought to live
together. This is not to say that Athenian philosophers represented a new class of
person who is not likely to hold unreasoned views (philosophers still do that), but it
does mean that they recognised, and explicitly put in a front-seat position, as it
were, a rational methodology for discovering and subjecting to scrutiny such
unreasoned beliefs rather than placing them beyond question.

The upshot here is that “pre-philosophical unreason” is not primarily
intended to describe an attitude which was temporally prior to philosophy but one
that is logically or conceptually prior to philosophy in the sense that it is deficient in
the analytic, self-reflective, and dialectic tools characteristic of that set of practices.
History is clearly not a straight and unbroken line of progress from unreason to
reason, societies have both progressed and regressed in this regard and pre-
philosophical unreason certainly exists today. Nevertheless, even though it is not
my sole or primary intention, I do allow that the “pre-philosophical” descriptor has
some temporal connotations since societies prior to major world-historical
rationalizing innovations such as the Athenian school or the enlightenment were
more inclined towards uncritical and unreflective forms of irrationality.\(^{A}\)

\(^{A}\) One could point here to the fact that many early hunter-gatherer societies were more peaceful, cohesive
and egalitarian that the allegedly more rational civilizations. I will maintain this is not owing to the fact
that they were more practically rational but largely to the fact that the increased size, complicity, and
diversity of more “advanced” civilization created more serious action coordination problems and because
the pace of instrumental rationalization and technological development vastly exceeded the development
and implementation of moral-practical rationality. I will return to these issues in later chapters.
The second point of qualification is that I use the term philosophy here in a rather broad as opposed to a technical academic sense. In this context “pre-philosophical” means an over-readiness to rely on external judgment, authority or power and a deficiency in self-reflectivity and the courage to think for oneself, not necessarily prior, logically or temporally, to the invention of an explicit methodology or formalized field of study. One does not need formal academic expertise to think about what is right or wrong, to be sure. What this account does entail, since practical reason is not the province of specialists, is that every person is called on to reason about moral-practical questions—to be, in one sense, a philosopher.

Not only has the unreflective or pre-philosophical form of unreason not left us, but it has been joined by a somewhat unwitting compatriot, an unreason of a somewhat more recent variety, on the front-flank of reason—the hyper-sceptical, self-deconstruction of reason. The progressively radical stages of reason’s critique of itself have led in some post-modern theories to an implosion of reason, where anything remotely resembling the old ideal of reason is seen as the ultimate naïveté. Although the more sophisticated views of thinkers associated with this second type of unreason may hope to oppose authority and tradition, they are, ultimately, unable to do so, for without reason they have nothing with which to resist them. Post-philosophical unreason and old fashion pre-philosophical reason amount, in large measure, to the same result.

Today we are beset by the forces of unreason: racism, sexism, homophobia, jingoistic nationalism, religious fundamentalism (Christian, Islamic or otherwise) and corrupt, tyrannical governance. But this is not a thesis about those forces. The
central motivation of this thesis is to try to answer the terrifying, all important question—“is there still something that can be opposed to the irrational forces that are threatening humanity?”

Unfortunately, the dominant view of a good life based in reason seems, in many respects, to have suppressed and degenerated humanity with greater speed and efficiency than unreason. The dominant game-theoretic and related ideas about reason draw on various enlightenment theories, but were perfected by economists working in the paranoid climate of the cold war and rely on a bleak and narrow conception of human beings as purely self-interested, probabilistically-calculating, utility-seeking machines, who only interact with each other in strategic, purely-instrumentalizing ways. This view has put faith in free market ideology and sanctioned corporate globalization under the administration of fiercely anti-democratic international economic organizations—the World Bank, International Monetary Fund and World Trade Organization which represent global governance based on debtor/creditor relations and military force rather than understanding or diplomacy between democratic nations. The result has been alienation, misery, intensified global poverty, catastrophic environmental damage, corruption, domination, inequality, inadequate education, militarism, war, terror, and a world unequipped to respond to natural disasters, pandemics, famine, the management and restoration of natural resources and environments, and other threats facing humanity as a species.

The first step towards negotiating a way out of this high stakes labyrinth of dead ends is to recognize that reason is not a simple unified idea. This is very good
news, because it means that we need not side with the wholesale rejection of the idea that we could base our lives, our societies, and our institutions on reason simply because some previous attempts to do so have proved problematic. This chapter and the one following it represent an attempt to show that basing ethics on reason is an idea with diverse manifestations represented in a richly varied history of development spanning thousands of years. In considering some of the main currents of thought in the various historical developments of this idea (from ancient Greece till the end of the enlightenment period) as well as some of the more serious challenges facing it, these first chapters aim to lay a foundation for subsequent considerations of more recent attempts at working out which parts of this tradition can be maintained and which strategies prove untenable. Criticizing reason from some irrational perspective seems problematic—an outright contradiction—but, if the history of ideas about reason is successfully shown to be diverse and problematic, then continuing a genuine self-critique of reason can make sense. One of the central claims of this paper is that the self-critique and development of reason can continue because of the strength of the core idea that normative judgments can be justified or redeemed rationally. If this core idea can be maintained then the primary theoretical task for normative theory is to disentangle this core from the problems faced in the ongoing development of practical reason.

The genealogical analysis I offer in both this and the immediately preceding chapter is not primarily aimed at correcting historical interpretation but is intended to be part of a conceptual analysis of the development of practical reason which is preparatory for understanding and addressing the problem of how to ground
morality in a contemporary context. I aim to layout the background or horizon against which contemporary debates about the possibility of rationally grounding ethics are set. All the philosophies discussed here address live issues and are deeply influential for contemporary proposals for how to go about normative theory and practice today. The main foci of these chapter are Aristotle, Kant and Hume, since, I claim, most views about reason's connection to ethics\(^A\) and politics, even today, are intimately related to one of these three major paradigms.

**1) Reason from the Beginning: Aristotle**

One of the most fundamental starting points of Aristotle’s philosophy is the distinction drawn between the natural and the artificial. Things that are caused by nature have within them an innate principle of motion or stationariness. Artificial things have no internal principle; hence, in so far as a thing is an artificial product, the sources of its production must come from without. Artificial products, such as beds, may happen to be made from natural products, and insofar as they are natural they do have an internal principle, but insofar as they are artificial products of art they do not. It is possible that if the wood of the bed were green and hearty it might take root and produce more wood, but one could never plant a bed and grow more beds.\(^1\) For Aristotle, each natural thing has its own unique internal principle—that which it is by, or according to, nature (e.g. the property of being carried upward is in the nature of fire).\(^2\) This is thought to be a reasonable starting point because it is held to be self evident that there are, in fact, natural objects that have their own

---

\(^A\) Ethics and morality will sometimes be used as synonyms, as per ordinary English usage. Later I will rely on more technical distinctions between the two, but I will be clear when I intend such distinctions.
innate principles of operation. In a sense, this is the stable, intuitive reference point around which all further philosophical reflections are oriented.³

When Aristotle comes to consider questions concerning what is good and how we should live he begins by recognizing man as a natural being and asks what is man “by nature”—what is the unique internal principle that drives him? His conclusion, “All men by nature desire to know.”⁴ This natural principle is what moves mankind from within and what distinguishes him from all other natural kinds. Now, when Aristotle asks “what is good?” he finds the usual answer, that pleasure, or what we might call happiness, is the ultimate good, to be unsatisfactory—for it tells us nothing about the content of that general term. Aristotle argues the good for all things consists in performing its natural function well. He offers analogies to various particular human activities. A sculptor, for example, has the function of sculpting. We would attach the adjective “good” to a sculptor who sculpts well. Man himself, according to Aristotle’s theory of human nature, also has a function—to know. And since this is not a function instrumental to some other purpose but the final or ultimate principle of man, it follows that knowledge is good in and of itself, and if the goal of human life is to know for the sake of knowing then it turns out that what it means for man to perform his function well is to live a life in which the activities of one’s soul “follows or implies a rational principle.”⁵

When Aristotle goes on to consider the very nature of knowledge and understanding itself, the idea of a natural principle again has a central significance, for it turns out that of all the species of knowledge, the highest form is precisely a
knowledge of the ultimate causes or first principles.⁶ “All men begin...by wondering that things are as they are, as they do about self moving marionettes...for it seems wonderful to all who have not yet seen the reason.”⁷ The ultimate reasons for things are the most difficult to discover because they are the furthest from the senses but they are also the most true knowledge, for it is “by reason of these, and from these, all other things come to be known.”⁸ There is not one universally-applicable, strict, or formal method for discovering these first principles and it is therefore incumbent on us to find the method that is appropriate in each case. Sometimes they are realized through induction or certain types of habituation, other times they may be directly perceived, but in all cases it is necessary to be as precise and rigorous as possible in these matters because of their foundational influence on all derivative knowledge. As Aristotle puts it, the beginning is more than half of the whole.⁹

In keeping then with Aristotle’s insight into the importance of beginnings I will, in this section, take time to consider in more detail Aristotle’s views on practical reason since it is he who is widely considered the beginning of serious thought on the topic, and furthermore because he may be considered, alongside Kant, the originator of one of the two main lines of thinking about reason directed towards human ends which have predominated discussions to this today.⁸

In Aristotle, for the first time we begin to see practical reason, which confronts the question “what to do?” systematically distinguished from theoretical

---

⁶ One could easily make the case that I could have begun this discussion with Plato. Indeed, despite his self distancing, Aristotle owes a great deal of his philosophy of reason, action and ethics to Plato, who deserves a separate discussion, but given the scope of this paper I must take Aristotle as the representative of this Greek Socratic tradition insofar as it sought to establish ethics on a rational footing.
reason, which is directed towards the question “what to believe?” Nous, for Aristotle, is the faculty of understanding, theoretical reason, or intuitive intellect. It is the part of the soul where reasoning (logos) itself dwells. The activation or activity of nous is study (theoria) – a theoretical contemplation for the sake of knowledge. Phronesis, on the other hand, is the name given to practical wisdom, intelligence or reason. To have the virtue of Phronesis is to have the ability to conduct one's deliberations so as to bring about the fullest realization, over a complete lifetime, of the supreme good—which Aristotle calls Eudaimonia—meaning, the fulfilled life, flourishing or well-being. Understanding the significance of reason in Aristotelian ethics will largely consist in examining the role of, and relations between; nous/theoria (theoretical reason and its activity); Phronesis (practical reason); and eudaimonia (the ultimate human good).

1.1) Nous

Nous is not best understood simply as a direct intuition of foundational first principles. For Aristotle, in the progress towards knowledge one begins on the empirical level. This level of experience and empirical research (episteme) bears the bulk of the epistemological workload. This means that we rely heavily on our own sense experience, and, while not requiring that all who aspire to truth be scientists, it may require that we have reliable access to their results. Secondly, dialectics contributes by clarifying principles derived from the physical world by experience. Ideas are refined through this process of discussion, argument giving, and conceptual clarification. Only after this process is it possible for nous to grasp first principles. Aristotle does not think a clear understanding of the necessary truths of
existence are something that could just happen to occur to someone—if it were that easy we would not require philosophical rigour. Although *nous* is described as involving a certain immediacy—as an intuition whereby one simply sees that certain first principle are necessarily true—it is an intuitive grasp underwritten by empirical research and dialectics. At a certain point we can see that certain first principles are necessarily true, but a significant amount of work is necessary to reach a position wherein one can apprehend such truth.

*Nous* is believed to reveal fundamental facts, many of which will be relevant to the ethical/political life. Most importantly, it is by way of *nous* that we are able to know the fundamental function, essence, or nature of things, and derivatively, to know what is good and what is bad, and to know the reasons why a thing ought to be done or ought not to be done. *Nous* does not, however, take the further step of prescribing action. This is the domain of *Phronesis*, a second and surely indispensable function of reason, if it be granted that ethics requires not only knowledge of various universal principles, chief amongst which is the knowledge of the highest good, but also requires knowledge of how to practically apply those principles in order to actually achieve said good.

1.2) Phronesis

*Phronesis* accomplishes this by supplying reasons that serve to guide our practical deliberations and thereby function as support or justification for our choices. “Choice” in this context does not mean simply choosing to believe that something is the right thing to do, for the conclusions of practical arguments are always actions for Aristotle, even if those actions were to be represented with a
What it means to practically justify a choice is to have reasons to conclude that the means chosen are the best way to achieve the end sought. Many different goods (*agonon*) may be sought through action but all are sought, at least by the virtuous person of practical wisdom (*phronesimo*), for the sake of the true ultimate good *eudaimonia*, which we will return to in more detail shortly. *Phronesis*, in other words, has no say over ends. It is purely instrumental—exclusively occupied with finding the appropriate means for the given end, in this case a function (*telos*) fixed by our nature. When we are presented with an action context, wherein we must choose some action or set of actions from amongst many possible actions, we are said to be in a situation that calls for deliberation. We deliberate, not about necessary truths or things fixed in the past that cannot be made otherwise by our decisions, but only about things that can be different, and things which we have the power to change or influence. For Aristotle, since what is good is an invariable, necessary truth about our nature, it has nothing to do with choice; thus, it is not something we could deliberate about. Deliberation—choosing actions—is the unique and exclusive domain of the practice of reason called *phronesis*. Also, since it deals, not with necessary principles, but contingent futural uncertainties that relate to doing and making things, *phronesis* is of necessity a much more inexact affair than *theoria*.

Despite being limited to the domain of deliberation and therefore means, Aristotle does however clearly distinguish *phronesis* from mere instrumental

---

^Take the following example: "You want a loan. The bank manager oversees all loans. You should talk to the bank manager." The conclusion of this syllogism, if understood as a practical argument, would not be the preceding sentence, nor the mere belief that the action it prescribes is the right action, but the actual act of talking to the bank manager.
cunning or shrewdness (deinotes).\textsuperscript{A} The difference being that instrumental calculation directed towards any end whatsoever deserves the name deinotes, whereas the intellectual virtue of phronesis only describes instrumental calculation which necessarily orients us towards the self-realization of eudaimonia. Insofar as deinotes is concerned, the fact that one’s purposes might be the most vile, heinous, and destructive ends imaginable is irrelevant so long as his or her monstrous purposes are efficiently and effectively achieved. Phronesis subsumes and surpasses deinotes.\textsuperscript{18} Phronesis, in general, includes all lesser and specific types of practical reasoning—moral as well as technical and prudential reasoning. The lesser forms of practical reasoning are called such only in a qualified way. A particular craft (techne), take for example the ability to build computers, involves the ability to deliberate well when selecting components and deciding how best to configure and assemble those parts so as to create a device most well suited to its intended function; but, no matter how well one practices the craft of computer making, if computers are not known to be the highest good in the universe or even the final purpose of a human life than one must conclude that the craftsperson or practitioner of the technique can be said to have deliberated well only in a limited way and merely with regard to some particular end. While a particular end may coincide with, or be part of, the ultimate end, it certainly need not be. For example, building the computer may have been a waste of time either because of an

\textsuperscript{A} This may appear to be a contradiction and indeed some (e.g. Raphael Demos, “Some Remarks on Aristotle’s Doctrine of Practical Reason,” in Philosophy and Phenomenological Research 22, no. 2 (1961): 153-162.) accuse Aristotle of being inconsistent on the question of whether or not phronesis can decide ends, although this does not seem to be widely accepted. I go on to follow the more predominant interpretations of commentators who, more charitably I think, present Aristotle as consistent on this point (John M. Cooper, C. D. E. Reeve).
overabundance of computers, or because more pressing matters should have been attended to instead, or, worse still, the computer may have been designed for the purpose of logistics related to the organization and administration of a genocidal death camp.

Even if one had somehow mastered not just building computers but every conceivable techne, one would still not, on account of that fact, qualify as practically wise in an unqualified sense because all techne is limited to guiding deliberations pertaining to activities that shape external matter, while the human good is not exhausted in such acts of making physical things. Practical wisdom, as such, must aim at what will be argued to be higher things—things which involve shaping, not primarily any external thing, but the character of the agent. Although the final good is not something we deliberate about, to be called a phronesimo, to be practically wise as such, without qualification, one must aim not at some specific good but at the good as such, without qualification.¹⁹ Hence Aristotle tells us “it is impossible to be practically wise without being good.”²⁰ This entails an important link between nous and phronesis. Practical wisdom needs theoretical wisdom to provide knowledge of the ends at which it must aim.

The ethical life—the truly good life—means knowing what is good and knowing how to achieve it, that is, knowing universal principles and how to act in order to achieve eudaimonia. Further, however, the good life also means acting in accordance with that knowledge in order to actually realize the good, and if that is to happen, being ethical also means that our passions and desires must be brought under the control of reason. It is not enough that reason should guide deliberation;
we must also “desire in accordance with our deliberation.” Everyone’s action is directed towards the realization of what they desire, what they think is good; for purposive or goal-directedness is thought to be internal to what action actually is. The virtuous person, however, is distinguished by the fact that they both know and desire what is truly good. Hence the view expressed in the famous metaphor that the passions properly function as a slave to reason.

The somewhat heavy-handed master/slave metaphor may be a bit misleading however. Aristotle says the intellect rules the passions with a “constitutional” not a “despotic” rule. The idea is not that desire should be dominated or completely suppressed in favour of a dispassionate cool and calculative governance of behaviour by the intellect. We must not imagine Aristotle akin to Mr. Spock. The virtuous person is a passionate one who must desire and enjoy virtuous activity. This happens when the passions are developed properly under the guidance of reason. Our natural powers, dispositions, and desires must be reflectively aimed or adjusted by reason. Reason and passion converge in a sort of negotiation process whereby desires become thoughtful and understanding becomes desirative. But, even if the basic model is meant to be more akin to a deliberation than a domination, for Aristotle, I see no way around the fact that it cannot be an entirely even-sided deliberation. I would suggest the most sympathetic explanatory model may be the pedagogical. Desire, the pupil, is not extinguished, is not without its originative contributions, nevertheless, there is no

---

^The logical and emotionally suppressed Vulcan alien from Gene Rodenberry’s Star Trek.
question that it is reason that must take pride of place in the roll of educator. A good desire is one that has been properly educated by good reasoning.

To summarize, Aristotle attempts to place ethics on a rational foundation in the following way: Firstly, the world is comprised of things, which have their own internal principle, that is, kinds of things simply are a certain way by nature. Human beings have the capacity to grasp the nature of things 1) fundamentally through empirical perception and scientific research, then, 2) refine and clarify these beliefs through dialogical argument and ultimately 3) to understand the first principles of things through pure theoretical reason—theoria. In this way we are able to discover what human beings are, by nature, and what in turn it would mean for a human being to function well, fulfill its nature, and become a good human being. Practical reason is our constant guide on the ethical path to goodness. It allows us to decide what to do—how to find the most effective way to achieve the good we seek. Since our passions and desires are receptive and malleable, possessing the capacity to be trained and shaped by reason, if we posses right reason our passions will become habituated to prefer the good. Thus, becoming ethical consists in the proper functioning of reason. We know the good through one function of reason, our action is guided through another, and finally we come to desire the good and our happiness comes to coincide with its achievement when our passions are shaped by the correct exercise of reason. We have yet to consider in detail however, what the final purpose of human beings (eudaimonia) is. It should be clear from what was said at the outset the eudaimonia itself is ultimately the activity of nous (theoria) but it
remain to be seen what specific reasons Aristotle gives for this conclusion and what its full implications may be.

1.3) Eudaimonia

The good life, for Aristotle, consists in following the virtuous path of his famous golden mean (mesotos). The virtuous character is always the middle way between two vices lying at either extreme. If bravery be the virtue, the extremes of cowardice and rashness would be the corresponding vices. The virtuous mean is not a mean, as Aristotle says, in terms of the object, but a mean relative to the agent. What’s good for Papa Bear may not be good for Goldilocks. If given a choice of porridge-temperature between 30° and 130° one should not expect the mean choice to be 80°; the mean for Goldilocks may be 37° and the mean for Papa Bear may be 127°. The mean is not chosen with respect to any quantitative or qualitative values represented by the actual range of options available. Thus understood, Aristotle himself admits this insight, taken alone, tell us nothing about how to act. The prescription that in all things one should not have too much or too little, but that which is just right, that is, most appropriate in a given circumstance, surely is not much of a prescription at all since it leaves unanswered the all important question: how do we know what constitutes the appropriate action in a given situation? Aristotle acknowledges the central significance of this question. His answer is that the mean is determined by what serves the supreme end (eudaimonia) which I will now freely call theoria. What this means is that, for example, facing danger in combat might be said to spring from a virtue of character (bravery) if we are defending a country in which theoretical study and contemplation flourished; but no
deed could be called courageous if it be in defence of some barbarian territory devoid of philosophers and scientists. If one were defending such a territory there would be no question of a mean; whether one acted reckless and overzealously, cowardly, or anywhere in-between the act would be unjust because it had the wrong aim. If we are fighting to defend a just and enlightened territory the correct way to fight is what best serves contemplation. If one were to run away or cower in fear when the enemy approached we would not be safeguarding the reason that dwells within the city. Likewise if one were to rush headlong into the barbarian hoards while the front lines were falling back to regroup one would face certain death without any appreciable contribution to the overall success of the battle. Contemplation would have been better serviced by staying with the group and preserving ourselves to fight in the defence of reason another day.

But again, why select this standard? The simple answer already given at the beginning of this chapter is that reason is what makes us distinct and therefore defines our nature, thus it is through contemplation that we fulfill the unique essence of our natural kind. It is also suggested that theoria is superior because it is self-sufficient. Moral virtues, such as justice, require an alter towards whom we may act justly, but the philosopher can contemplate the truth on his or her own. Aristotle also seemed to think that the contemplative life is simply a happier one, for “those who know will pass their time more pleasantly.” Aristotle, however, goes significantly further than all this. If a couch does its job well and is a comfortable couch it is at best a good couch. But, if a person does their job well they become

^ Aristotle may not have added “or her,” in this context. I do not discuss that issue here nor do I try to sort out when Aristotle is talking about human being (anthropos) or adult male humans (aner).
more than a good example of their kind, they become the greatest good—period.

This is on account of the fact that our reason, according to Aristotle, is Divine.

It is important here to take a moment to consider some relevant aspects of Aristotle's philosophy of mind from *De Anima*, which I think underwrite this picture of reason as Divine. Aristotle seems to have a view of the relationship between mind (or soul) and body which is neither materialist (viewing all states of consciousness as reducible to physical body states) nor dualist (at least in the sense of viewing states of consciousness as capable of existing independent of a body). He explains the relationship between soul and body in term of “matter” and “form,” a distinction which pervades most of his philosophical enquiries. The body is the material of which a human being is composed while the soul is the name given to the form of a human being (or any other natural living thing). The form is that which gives the matter its meaning. To take a simple example, if a nose were a living being its soul would be smell. Now it would seem a rather strange and falsely dichotomous question to ask whether the nose is identical to the act of smelling or does smelling exist separately from the nose? The question seems to misunderstand the nature of the form matter relationship. Consider Aristotle’s example of the wax seal. The shape of the seal (form) and the wax itself (matter) are clearly distinct since this material could easily exist in some other form, but on the other hand there is no reason to suppose the form of the seal could exist without its material basis in the wax. Likewise, it would not make sense, on this model of explanation, to say the soul is reducible to the body or that the soul could exist independently of it.
Despite the fact that Aristotle’s conception of the relationship between soul and body represents a novel third way, in the end he does seem to sign on to a version of dualism, albeit a different and much more restricted one than that of say Plato or Descartes, and he does this for a very important reason: I suggest it is, in part, this dualistic account of mind which allows him to assign transcendence to reason and equate it with the highest good. Aristotle’s dualism does not pertain to the entire soul, but only that part of it which concerns knowing and understanding viz. the mind. So, when Aristotle says, in book II, that “the soul is inseparable from its body, or at any rate...certain parts of it are”\textsuperscript{29} and that some parts of the soul “may be separable because they are not the actuality of any body at all”\textsuperscript{30} he seem to have left room for himself to consistently claim in book III that “while the faculty of sensation is dependent on the body, mind is separable from it.”\textsuperscript{31}

So, given Aristotle's novel solution to the materialist/dualist problem, what led him to suppose that the activities of mind are separate from the body? For Aristotle the mind functions in a way roughly analogous to the part of the soul that senses. Just as the sensitive faculty has the potentiality of receiving the sensible form of objects, the mind has the potentiality of receiving the intelligible form of objects. The mind, like the soul generally, since it has the potential of becoming identical with the object, is “in a way all existing things.”\textsuperscript{32} This has two important qualifications: 1) The mind is all objects not actually, only potentially, and 2) the mind is not the thing itself but only the form of it, which is to say the mind is never a stone, (form + matter) rather, it is simply the form of the stone. Moreover, Aristotle suggests that before the mind receives the form of the object it is not some material
ready to take on a form, rather it is a nothing. All that exists is a potential to become a sensible form, hence there is no need for a reliance on the physical body.\textsuperscript{33}

Because the mind is simply a capacity to think and judge and has no nature of its own Aristotle argues that an act of mind is not a bodily act. Before the particular acts of thought the mind is really nothing—at the very least nothing physical. If what we call mind is nothing but the potential for thinking, then it need not be blended with the body as is the case with the sensitive faculties, which require physical organs in order to receive the forms of objects. Mind is not the actuality of any part of the body i.e. the mind receives sensible forms unmediated by any act or condition of the physical body. As further evidence Aristotle cites the fact that the senses can be overwhelmed by powerful stimulation.\textsuperscript{34} Take an instance when we hear a very loud sound and then cannot hear well afterwards, such as stepping out of a loud concert, or when we are blinded by a great light. Thinking about an object which is highly intelligible, however, does not make the mind less able to think. According to Aristotle, this shows that the mind is not subject to the physical limitations of the body in the way senses are. It is because of this transcendent aspect of reason that Aristotle can say in the Nicomachean Ethics that "being connected with the passions also, the moral virtues must belong to our composite nature [mind and body]...[but]...the excellence of reason is a thing apart."\textsuperscript{35} By "thing apart" Aristotle means pure intellect, which he will equate with the Divine.

On Aristotle’s view, the Gods are understood to be the most blessed and happy beings. Moral virtues like justice, bravery, and temperance are beneath the Gods since they do not enter into contractual relations that need to be honoured,
they confront no risks or dangers that demand courage, nor do they possess bad
impulses or appetites that require restraint. Since the Gods take part in no action or
production their only activity can be intellectual contemplation, thus their only
virtues must be of an intellectual nature. So, if the Gods are not just good, but the
most blessed of all beings, and their only activities are contemplative, then it is the
intellectual virtues of humans that must be the greatest and most godlike. So
Aristotle tells us:

If reason is divine, then, in comparison with man, the life according to
it is Divine in comparison with human life. But we must not follow
those who advise us, being men, to think of human things, and being
mortal, of mortal things, but must, so far as we can, make ourselves
immortal, and strain every nerve to live in accordance with the best
ing thing in us; for even if it be small in bulk, much more does it in power
and worth surpass everything.

On this view, the intellectual virtue of contemplating well is not merely one
virtue among many; it is the supreme virtue, the for-the-sake-of-which of all else in
human life. The moral virtues are concerned only with that which is variable and
therefore less excellent. Despite this, Aristotle’s account of the good life takes keen
account of the fact that we are not gods or asocial, disembodied intellects by
including the active and practical in the eudaimon life. I do not think there is a
contradiction here. There are simply two senses of eudaimonia, what C. D. C. Reeve
calls “primary eudaimonia” which is divine theoria and “secondary eudaimonia”
which is practical activity. As Aristotle says:

For man, therefore, the life according to reason is best and pleasantest,
since reason more than anything else is man. This life therefore is
also the happiest. But in a secondary degree the life in accordance with
the other kind of virtue is happy; for the activities in accordance with
this befit our human estate [emphasis mine].
The eudaimon life itself is an inclusive end. Because of the kind of social, biological creature we are we need to eat, exercise, recreate, get along with each other etc. and the good life must involve all these activities because our ability to contemplate depends on them. So, although the life of contemplation (vitae-contemplativa) is primary, it is recognized that the active life (vitae-activa) is an indispensable component of the eudaimon life. The good life, then, consists in a range of goods properly ordered, that is, the good life consists in various external goods, which are chosen merely because of their instrumental value for realizing the activity of pure intellectual reason, and also many things that may be intrinsically good, but even these components of the good life are always ultimately choice worthy because of their relation to contemplation.

1.4) Weaknesses of the Aristotelian Approach

The Aristotelian account of ethics begins with theoretical reason (the instrument by which we discover what the good life for a human being is) and practical reason (the instrument by which we are able to structure every aspect of our life in order to orient it towards the realization of the good), and it ends with the realization that the greatest good is none other than theoretical reason itself. The significance of Aristotle's contribution, as virtual founder of the attempt to shift ethics onto a rational foundation, are inestimable despite the fact that it is marred by many serious difficulties, including significant vestiges of transcendent theology, which make this program deeply incredible as it stands in its original form. Many

---

\(^{A}\) By this I do not want to imply that there is anything wrong with religion or belief in God contributing in any way to ones ethical outlook, rather, I mean to suggest that moral rationalism is committed to maintaining that the whole of morality does not need to rest solely on faith in a divine
of these problems will be glaringly obvious to contemporary readers, and many of them will even be accepted, and accommodated for, by contemporaries still calling themselves Neo-Aristotelian. Some of these criticisms will come out over the course of this investigation, at this stage, however, I would like to address the most basic of these difficulties.

The first glaring question is; why should all other human goods be subordinated to any one good? Ultimately, Aristotle can only answer this question by appeal to a transcendent theology. Reason is the best thing in us because it is Godlike. As soon as we have reason to doubt the certainty of Aristotle’s contemplative Gods this picture begins seriously to lose its credibility. Moreover, the idea that the purpose of human life, in the final analysis, is to be more like the Gods means that Aristotle has not, in the end, actually succeeded in putting ethics on a rational footing at all. Aristotle may defend the primacy of contemplation by falling back to the claim that the contemplative life is objectively the most pleasant but here we would have to rely on the testimony of those who have experienced theoria and found it most desirable—i.e. the philosopher. It seems more plausible to interpret this merely as evidence for the power and influence of upbringing and ideology. Aristotle himself admits that one’s ideas about what constitutes the good life are influenced by how one has lived. In light of this, how is one supposed to know that the businessperson, who says life is about wealth, is wrong, and Aristotle, who has spent his entire adult life excelling in philosophy school, (either

---

power, that, in other words, some portion or aspect of ethics is accessible to us qua rational beings. To the degree that Aristotle fails to establish some normative standard on a purely rational basis he has failed to do what he appears to have set out to accomplish.

A in chapter three I will address how this problem also presents itself in various secular approaches.
the Academy or his Lyceum) is correct when he “discovers” that the values of study and truth that he has tirelessly cultivated and grown to loved are the true ones?

Secondly, Aristotle’s claim that our ultimate end does not originate in our desires and is not chosen, but rather, fixed by our natural purpose, clearly presupposes a natural teleology which has faced serious and sustained criticism.\(^A\) It is not clear why we should think that contemplation or any one fixed good is our true essence or even that things have true essences that we can discover. As discussed in the next chapter, Aristotle will be forcefully challenged by Hume on the assumption that it is possible for our passions to be educated by reason, and will be confronted by the counterarguments of Kantians who argue that he cannot properly account for the distinctive fact about morality, namely, that the value of moral action is not derivative, or in other words, moral action is not good or right because it leads to some good result, but is good solely in itself because it is an instance of an absolute rule. Against Aristotle, Kant will also argue that there is no reason to suppose that the morally right or virtuous act has a necessary connection to any desired result such as happiness. Further potential difficulties will stem from the radical critiques discussed in chapter three and the versions of moral rationalism (discussed in the later chapters) that try to take account of some of these concerns.

2) The Age of Reason

It is apt that the enlightenment period (centered in the eighteenth century) is also called the age of reason on account of its fierce new criticisms of appeals to force, dogmatism, and all unexamined presuppositions. In the words of Axel

\(^A\) There is no space here to specifically address this criticism but it is an issue that I will touch on in subsequent chapters.
Honneth “what is unique to enlightenment is its immanent relation to a criterion of rational validity which acts as a standard against which opinions and convictions can be upheld by rational examination.” And, as Tom Rockmore puts it “one of the common threads which links together otherwise disparate analysis of modernity lies in the reliance on the traditional philosophical instrument, reason.”

Since the medieval period much of Europe had taken its moral guidance from Catholicism according to which moral rules and ideals were taken to be the will of a perfect God, and this will was revealed by God himself in the words of the Holy Bible. The ultimate ideal of a moral life was exemplified in the person of Christ, whom we are meant to emulate. Thus, imitatio christi could be called the Christian fundamental moral principle. The moral law was thought to be binding for us because it is backed by the power and authority of an omnipotent redeemer who will reward obedience and punish transgression. But this moral basis had already become problematic in the wake of the protestant reformation that splintered the Christian faith into various sects. Even amongst those who held to transcendent religion there was no agreement on how to interpret the revealed word of the Bible and the moral rules it contained. In the absence of agreement about fundamental questions, such as what it meant to imitate Christ, religion was beginning to look untenable as a basis for a public morality that could be shared by all. This fire was fuelled further by the fact that the idea that a divine book that could give us direct access to a transcendent God’s eye perspective was not standing up well to enlightenment scrutiny.
Unfortunately, however, reconnecting with the rational ideas of the classical Greek age was not a viable solution either. This had, broadly speaking, been the approach of the Renaissance period and it too fell under suspicion. Thus, the “age of reason” is also a period in which reason intensifies its self-criticism. The developments of this period largely consist in “enlightenment about the boundaries and limitations of human rationality.”\textsuperscript{44} Doubts arose surrounding the Aristotelian idea that through reason we can apprehend our objectively correct ends. The old Greek ideas about reason require that subjective human reason be measurable against some objectively reasonable things. The world has a natural order and natural laws which we perceive. In this way we directly apprehend the true idea or essence of things and our moral laws receive ontological confirmation from the rational order of being.\textsuperscript{45} On this view, as shown in the previous section, human beings have a calling or a natural role which can be deduced from their place in the teleological order or being. But the idea that things have teleological significance seems to depend on the assumption that the universe is rationally ordered by a wise creator—an assumption now held up to the light of critical enlightenment reason.

The idea that we can simply grasp the ultimate nature of reality, or that it is revealed to us from beyond, begins to look less and less likely, leading either to the abandonment of religion and metaphysics or the attempt to found them completely anew. Enlightenment thinkers now question whether the objectively reasonable is relevant to action and indeed whether it is even possible to discern what is objectively reasonable.\textsuperscript{46} Such doubts, however, were simply the birth pains marking the passage to a new, more self-critical and more subjectively oriented
reason, thought to be appropriate to an enlightened secular age. The enlightenment thinkers found their task to be that of developing a moral point of view from within, that is, they sought to find binding universal norms without the resources of a transcendent perspective (traditional religion) or the assumption that humans are imbued with natural purposes (traditional metaphysics).

With the empiricists came the rise of the “mechanical model of reason.” Such views tend to support the idea that we are equipped with a capacity for knowledge to tell us about how the world is, but knowing how the world is—facts—never allows us to deduce how the world should be—values. Thomas Hobbes develops a mechanistic model according to which reason is simply an ability to calculate—to follow an algorithm. Practical reason is nothing other than calculating consequences which in turn excite the appetites or aversions that motivate action. Since reason does not speak to ends there is nothing absolutely good or bad. The practical and theoretical reasons which Aristotle sought to distinguish are collapsed into one mechanistic theoretical reason and its prudential application.

Francis Bacon develops a similar instrumental account of reason where things only have a value relative to a given aim, and Blaise Pascal exemplifies how reasoning about expected value returns can then be done according to a probabilistic framework in his famous “wager” argument for believing in the existence of God. Deciding whether or not to believe in God is a matter of comparing what we likely stand to lose or gain from believing vs. the probable risks and advantage of not believing. Believing in God is thought to be the best thing because it is more likely to pay off for us. The utilitarians, Jeremy Bentham and John Stuart
Mill, tried to represent desire according to a single variable. Reasoning could then be seen as a maximizing activity according to a single utility scale. For Mill it was the general, necessarily impersonal, utility which ethics demands we peruse but later economic thinking, in the wake of Adam Smith, would drop that idea, so that being rational simply means seeking one’s own utility, which supposedly leads to an equilibrium, and in the end represents the best way to serve the general good. The Empiricist movement also led to unprecedented understandings of how the material world works, and, relatedly, to an equally unprecedented technological mastery of the forces of nature. Many thinkers were understandably in awe of these achievements which lent credibility to empiricism as an epistemological model for all human knowledge.

This new way of looking at morality, owed fundamentally to Hobbes, grounded morality in us as opposed to any authoritative source outside of us, whether in the form of God or a natural order. Specifically, for Hobbes, morality is grounded in our own desires or self-interests. We realize that submitting to a chosen human authority, the Leviathan, will allow us to escape the war of all-against-all found in the state of nature. Ethics is therefore identical to the law imposed by the sovereign. A consequence of this view, possibly a problematic one is that moral agency extends only to reciprocating partners in the social contract. The so called “free rider problem” points out that individuals may follow the sovereign’s rule when one has to or when it suites that individual, but when it suites one’s purposes and one thinks they can get away with it, that person would be right to
break the rules. This shows that no agreement between interested parties can itself give rise to any categorically binding obligations at all.\textsuperscript{48}

Also, for Hobbes, we are still necessarily subject to an external rule, only now it is a human who creates the law, and it is one to which we submit for the sake of our own desires. So, whereas an unenlightened medieval subject submitted to the King for no reason at all, except for an unthinking faith that the king was a divinely appointed representative whose rule was the will of the almighty, an enlightened Hobbesian would submit because the subject knows he or she is better off having a king than living in total chaos. Even Locke, who argued, against Hobbes, that we have certain inalienable natural rights, thought that most people needed significant external direction about what morality requires. In general, most early enlightenment thinkers (seventeenth century) thought that although morality was based on human desires it still had to be externally imposed.\textsuperscript{49}

A notable exception was the great continental rationalist Benedict de Spinoza who thought that ethics was about freely expressing our own nature. The key to understanding Spinoza’s conception of freedom is to be found in his distinction between active and passive affects. By "affect" Spinoza has in mind something close to, but slightly different from, "emotion" as one typically understands it today. Spinoza defines "affect" or "emotion" (affectus) as both bodily states and their corresponding ideas. In the \textit{Ethics} Spinoza also tells us that affect can be either "activity" or "passivity."\textsuperscript{50} Affect is activity in the case where something takes place, either within us or in our external environment, such that we can clearly and completely understand that change as issuing from the individual alone, or in
Spinoza’s language: when we are the adequate cause of said effect. Conversely, affect is passivity whenever we are not the adequate cause, that is, when we are passively affected by something external to our own essence. Emotion only remains passive—something that happens to us—insofar as we are ignorant of the nature of these emotions and the ways in which they influence us. Without an adequate knowledge or understanding of ourselves in terms of our emotions and their causes, we are subject to a kind of blindness that actually leaves us devoid of agency. In this state humans do not act but suffer; one is subject to external stimuli that push, pull, or shape the individual—this is the condition of human bondage wherein the individual is merely a powerless suffering slave of the emotions. The individual is causally determined by forces other than itself; it cannot act according to its own nature but merely reacts to the variety of mysterious/un-comprehended internal and external factors that push it about. Spinoza thinks this condition of passive suffering needs to be remedied with rational knowledge.

On Spinoza’s view, whenever these forces cease to be mysterious—whenever we form a clear and distinct idea (what Spinoza calls an adequate idea) of an affect—that affect ceases to be simply some blind force acting upon us. Knowledge of the affect is synonymous with bringing it under our power of activity, such that they can now be understood as an unfolding of our own nature. Instead of a passive lump being pushed around, the enlightened individual becomes like a vector in the sense that all we need to look at or understand in order to see where the thing is going is the immanent force of the thing itself. Thus, activity is the power to be your own adequate cause, and that is the same thing as knowledge, which, since it frees us
from the suffering state of bondage to the passive emotions, is also *virtue* and *happiness*. This activity/power/knowledge/virtue/happiness is also the very meaning of freedom for Spinoza; for freedom in self-determination. Since there is no way of escaping Nature, which is a completely causally determined system with no room for contingency, freedom is not opposed to determinism, it is simply indwelling causation. To be free is to have the self-knowledge that is the power to cause things to happen according to our own nature. For Spinoza, perfect or total activity or freedom is only possible for God /Nature but, because of their capacity for knowledge, human beings can approximate, more or less closely, this ideal.

Although it took very different forms, the eighteenth century would vigorously pursue the idea that morality must not be imposed on us. While undeniably original, it seems to me, Spinoza, in many ways, prefigures the themes of autonomy and reason which dominate the rest of the enlightenment. Like Spinoza some eighteenth century empiricists, such as David Hume and the 3rd Earl of Shaftsbury, would argue that morality comes from the individual and also like Spinoza they base their ethics on something like feelings instead of self-interests as Hobbes proposed. Hobbesian contractualism relied on the empirically false claim that everyone will exclusively pursue their own material self-interest, in other words, he denied the fact of natural sociability. Although he borrows a great deal from Hobbes instrumental account of rationality, Hume, by contrast, has the advantage of a more sophisticated account of the ways in which human beings are capable of benevolent or sympathetic feelings for others.
Hume’s attempt to account for what we call reasoning by the functioning of the imagination, essentially amounts to a bold attempt to reconceive that concept. The result is a picture of the mind in which reason and passion cannot struggle for control of the will. Hume’s is very likely the most sophisticated enlightenment version of an empiricist view of reason and will be considered in detail in the following chapter, so, for present purposes, suffice it to say that by making ethics an outgrowth of human feelings of approval and disapproval Hume remained committed, with Hobbes, to an instrumental view of practical reason.

The enlightenment would have to await the work of Immanuel Kant for an attempt to reinstate—on what, in retrospect, might be called a sort of post-metaphysical ground—the Aristotelian idea that reason can actually speak to ends which are not simply given by self-interest (Hobbesian Contractualism), feelings of approval and disapproval (Hume) or even the greatest utility for the greatest number (utilitarianism). Against all instrumental conceptions of practical reason Kant argued for the priority of the right over the good, where the right was understood as a law of pure reason that, in accord with Rousseau and before him Spinoza, must be self applied. Instead of the will being bound solely through the agent-relative self understanding of the subject, a knowledge set to which they have privileged access, Kant thinks the will, or human capacity for choice, can be bound or influenced by insight—that is, by objective reasons. I will say no more about Kant here since, along with Hume, Kant is one of the “the two most original eighteenth century moral philosophers”\textsuperscript{51} and will therefore be discussed in some detail in the next chapter.
I could not end this overview of enlightenment practical reason without a mention of Kant’s most influential successor and the last great enlightenment defender of reason—Hegel. G. W. F. Hegel’s criticism of Kant, to which I will have reason to return below, is essentially that Kant’s formalism is empty. The substantive content of morality must come from the vocabularies, practices, and institutions of our society—there is no critical perspective beyond history. At first this may seem like a return to Hobbes, for whom a shared moral law or public reason is something merely constructed or entered into on account of the fact there is no shared reason or law of nature other than the self-interest and instrumental calculations which led us out of the state of nature. This may also sound at first like a cultural relativist abandonment of the standards of reason, but perhaps surprisingly, nothing could be further from the truth.

The reason for this lies in the fact that Hegel also rejects the dominant prioritization of the individual self-consciousness as the subject of ethics in favour of a collective subject called Geist—the collective human spirit or mind—which is expressed in the aforementioned aspects of our social context that are developing dialectically through history according to an inner logic. For Hegel the ideas of a society will exhibit certain internal contradictions which require a development of those ideas into new ones which, in turn, throw up new problems driving further development. The overthrown ideas are seen, not as irrational, but necessary stages in a rational development process—ultimately a single mental totality—through which Geist comes to know itself. Once all contradictions are resolved, and the human species is fully self-aware, reason will have realized itself and history will
come to a close. Freedom, according to this picture, is not autonomous self-legislation triumphing over impulses, rather freedom exists in connecting one’s individual desires and beliefs with the cultural history and social institutions which embody and are involved in the rational evolution of the human spirit which continues to overcome its contradiction until the dialectical logic is complete.\textsuperscript{52} Ethics is ultimately a matter of plugging into the rational development of \textit{Geist}.

I hope to have demonstrated not only how Moral Rationalism took center stage in enlightenment thinking but also the fact that enlightenment practical reason is far from a simple or unified concept. In order to engage more fully with the strands of enlightenment moral rationalism of most lasting influence and relevance for the present investigation, I continue now with a more detailed examination of Hume and Kant’s respective accounts of practical reason.
Notes: Chapter One

2Aristotle, Physics, 193a.
3Aristotle, Physics, 193a.
4Aristotle, Metaphysics, 980a 1.
5Aristotle, Nicomachean Ethics, 1098a 6-9.
7Aristotle, Metaphysics, 982b 11-19.
8Aristotle, Metaphysics, 112.
9Aristotle, Nicomachean Ethics, 1098b 5-9.
15Aristotle, Nicomachean Ethics, 1112b 11-15.
16Aristotle, Nicomachean Ethics, 1141b 5-15.
17Aristotle, Nicomachean Ethics, Book I, Ch. 3.
19Aristotle, Nicomachean Ethics, 1142b 25-34.
20Aristotle, Nicomachean Ethics, 1144a 35-37.
21Aristotle, Nicomachean Ethics, 1113a 11-14.
22Aristotle, Nicomachean Ethics, 1176b 21-25.
24Aristotle, Politics, 1254b 3-5.
26Aristotle, Nicomachean Ethics, 1138b 1-35.
28Aristotle, Nicomachean Ethics, 1177a 22-25.
29Aristotle, De Anima, 413a 2-5.
30Aristotle, De Anima, 413a 5-9.
31Aristotle, De Anima, 429b 3-5.
32Aristotle, De Anima, 431b 21.
33Aristotle, De Anima, Book III Ch. 4.
34Aristotle, De Anima, 429a 30-429b 5.
35Aristotle, Nicomachean Ethics, 1178a 17-24.
36Aristotle, Nicomachean Ethics, 1178b 7-24.
37Aristotle, Nicomachean Ethics, 1177b 30 & 1178a 2.
38Reeve, 97.
39Aristotle, Nicomachean Ethics, 1178a 6-10.
40Aristotle, Nicomachean Ethics, 1178b 31 and 1179a 5.
41Reeve, 195.
44Honneth, 693.
46Schnädelbach, 9.
50Baruch Spinoza, Ethics (Hackett; Indianapolis, 1992), 103.
51Schneewind, 150.
Chapter Two

A Brief History of Practical Reason II:  
The Enlightenment
1) Hume's Sceptical Reason

Reason, for Hume, concerns the links between ideas. A piece of reasoning is a chain of ideas, where each idea somehow bears a relation in the mind to the next. One important function of Hume's whole empirical psychology developed in his A Treatise of Human Nature is to account for the association of ideas. There is much to support the interpretation that the status of reason was the central concern of David Hume’s career. In the abstract to A Treatise of Human Nature, Hume describes himself, saying “if anything can entitle the author to so glorious a name as that of an inventor, tis the use he makes of the principle of the association of ideas, which enters into most of his philosophy.” For Hume these principles are “the only link that binds the parts of the universe together, or connects us with any person or object exterior to ourselves...they are really to us the cement of the universe.”

Hume was also especially preoccupied with the relation between reason and common life—issues such as: How are we to get on with the business of life and philosophical inquiry given some of the sceptical conclusions we are lead to by certain lines of legitimate philosophical reasoning? Can reason influence the will? Can it be employed to discover principles of morality? Hume is famous, perhaps most famous, for the very restricted role he ascribes to reason. His conclusions about the role of reason in our lives were indeed radical and original, however it is at the very least an oversimplification to think that the thrust of Hume's arguments about reason are only negative claims about the faculty of reason’s ability to

---

\(^A\) John W. Danford goes so far as to say “the central preoccupation of his career was the place of reason in Human life.” (J. W. Danford, David Hume and the Problem of Reason (Yale University press; Chelsea Michigan, 1990))
generate truth, motivate action, determine right and wrong, etc. Hume was also deeply dissatisfied with the traditional explanation of what reason was in the first place. So, before we can understand or criticize the role Hume ascribes to reason we must understand, what Hume considers his greatest innovation—his constructive account of reasoning—which will necessarily lead us through several core themes of Hume’s Treatise.

Hume’s views on the role reason plays in our lives are far from implausible, and despite significant limitations I suspect any correct theory of reason has, at the very least, something important to learn from his analysis.

1.1) Hume’s Reason: Analyzing the Concept

In the sizable wake of Aristotle, philosophers had traditionally posited a powerful, independent, and irreducible faculty called reason which had the function of apprehending truth. Hume was very sceptical of the broad appeals to this, rather mysterious, truth revealing light. He wanted to know how reason works. He wanted a rigorous empirical explanation for how we get from one idea to another and how we arrive at beliefs. This meant the explanation must involve, first and foremost, the science of human nature, not merely the speculations of formal logic. Hume was dissatisfied with anything like Aristotelian Syllogistics as a basis for an account of reason. Deductive validity alone was of little use to explain our actual reasoning. For example, if we take the premise “Sabina goes for a walk” we are naturally led to a belief in the idea “Sabina moves.” This is not a formally valid argument because it is not true in virtue of the arguments form (SW → SM). If we change the content of the variables the argument does not remain valid (Sabina goes
to the art gallery therefore Sabina is the Prime Minister of Canada). The argument is only true semantically, that is, one can only recognize it as true once the content or meaning of the terms being related is taken into account. It seems Hume thought that an empirical methodology revealed that inference actually has more to do with the content of the ideas being related than with pure form.³ In the Treatise, Hume sets out to explain the principles and operations of reasoning not in terms of syllogistics, but in terms of the elements of the mental world and the way they are related according to his new empirical science of human nature.

Reasoning, on Hume’s account, is about finding certain relevant ideas, putting them in the right order and seeing that certain relations hold between them. The faculty of reason or the understanding is involved in the activity of making inferences. At the broadest level, inferences come in two forms and there are, therefore, two types of reasoning viz. demonstrative and probable.⁴ Demonstrative refers to inferences which result in knowledge. Here the premises are certain and this certainty is transferred to the conclusion. In probable inferences the premises are not all certain, and this uncertainty infects the conclusion. While probable reasoning results in beliefs or opinions it is not considered knowledge, properly speaking.

If two ideas are intuitively related, that is, both ideas and the relation between them are immediately conceived, this would not be reasoning but simply a case of direct perception. However, when one idea is related to another not by

³ Sometimes Hume seems to draw the same distinction with a different use of terminology, reserving the term "reason" for demonstrative reason while referring to probable reasoning as "the understanding."
direct intuition but via one or more intermediate ideas (where there is an intuitive perception of the link between each idea in the chain) then this would be a case of demonstrative reasoning. The possible applications of reason in this demonstrative sense were very limited for Hume. Just how limited, will be seen when I come to discuss the specific types of relations of ideas.

Hume’s conception of probable reasoning or the functioning of the understanding is significantly more complex. Hume begins his *Treatise* by laying out the elements of the mental world and the various relations in which they can stand to one another. His proposed mental landscape does not include any separate faculty responsible for making probable inferences, instead all such relations of ideas end up being explained in terms of the functioning of the imagination. In order to understand what Hume has in mind by probable reasoning it is first necessary to take a survey of his mental landscape. For Hume, the mind is made of distinct mental objects called *perceptions*. Perceptions are either *impressions* or *ideas*. What Hume has in mind by impressions would include all of our sensations, passions and emotions. These are basic or original facts about us. Nothing can truly be known about what causes them. Ideas, the other type of perception, are copies or faint images of impressions. All simple ideas are causally dependent on some impression. There can be no innate, spontaneous, or uncaused ideas on this view. Simple ideas can, however, be associated by the imagination to form complex ideas.

Ideas can relate to other elements of the mental world in only three ways: 1) impressions cause ideas 2) ideas naturally relate to certain other ideas, and 3) ideas can be philosophically related to each other. In order for the mind to function
Hume only needs to attribute two powers to it: 1) the ability to recall an idea once it has made its first appearance after being caused by an impression, and 2) the ability to combine or compare that idea with other ideas. Hume argues that all probable reasoning can be accounted for by the functioning of this rather limited mental machinery without the need to invoke a separate faculty of reason. Hume held that his science of human nature, which consists in an empirical phenomenological investigation of the interaction of perceptions, including relations of ideas which might be called chains of reasoning or inference, will take the place of explanatory work traditionally ascribed to a separate faculty named reason or the understanding. I will now attempt to say something more about what, specifically, the relations between ideas are and how they come to be, for only then will we be able to make any sense of the various sorts of reason Hume does recognize.

When, according to Hume, our minds are involuntarily led from one idea to another and the same simple ideas regularly come together to form complex ones, this natural association of ideas is not a matter of reason—that is, we simply behold or intuit that the ideas are related without any need for argument or intermediary ideas. The mind also naturally infers causes from effects and vice-versa; in this case a species of inference is required. We make these causal inferences in cases where past experience have shown certain things we call causes are constantly conjoined with certain other things that happen after them, which we call effect. This has to do with the functioning of the imagination according to habit and custom not a specialized truth discerning power of the mind. The beliefs or opinions which arise from these inferences are not considered demonstrably certain, only probable.
In addition, our imagination allows us to wilfully associate ideas together in various other ways. These arbitrary unions are what Hume calls *philosophical relations*. Of these relations, those of number and quantity are dealt with in the formal demonstrations offered by the sciences of algebra and arithmetic. These branches of mathematics are, for Hume, reason in the very strictest sense. This is the only time when inferences are demonstratively certain and yield actual knowledge.

So, in a nutshell, ideas can be related in different ways; one of the ways the mind is led from one idea to another is called reason. Demonstratively certain reasoning is only ever responsible for one of these ways—relations of number or quantity—and even then only sometimes. This is reason in the purest or strictest sense and is limited to a rather small and isolated role in the whole of actual human life and belief—algebra and arithmetic. Causal reasoning is involved with relations of cause and effect. It is not certain but probable. Having touched on some relevant uses Hume makes of the concept of “reason,” I go on to consider the relation between reason and the passions, reason and practical life, and the social/ethical implications of Hume’s conception of reason.

1.2) **Reason and Conduct: Further Implication of Hume’s theory of Reason**

Traditionally it was thought that our passions were often at odds, or in conflict, with reason. This, according to Hume, is a profound misunderstanding of what actually happens. On his view, if we understand the true nature of both reason and passion it would be understood that it is nonsensical to speak of any possible
conflict between the two. This move represents a break with most conventional thoughts about reason, both ancient and modern, since at least Plato and Aristotle.

Unlike ideas, our passions are not copies of anything more basic or fundamental. Passions are a species of impressions, and are, as such, original facts about our existence. It would not be appropriate to say that such original facts of our existence are either opposed to or in accord with truth or reason. A Proposition is the sort of thing which can be true or false, reasonable or unreasonable because of its capability to correspond with realities, such as when a proposition states that something exists which actually does exist, or that ideas are related in ways in which those ideas actually are related. A passion or emotion simply is what it is, in some sense. It cannot be compared with something else to see if it is an accurate representation of it, in the way that an idea can be compared with the object or sensation it is meant to represent.⁶

Neither is it the case that reason can cause or determine our passions, according to this account. Reason can lead us to truth and to form beliefs but the discoveries of reason can never be the actual origin of any passion whether a desire or aversion, that is to say, it can never give rise to a practical judgment. We have a propensity to desire that which produces pleasure and avoid that which causes suffering (either in ourselves or, owing to our capacity for sympathy, in others). There is nothing outside certain positive or negative feelings which can motivate our desires. Reason can play a role in directing our action, but it is always a role which comes after motive. Without an antecedent passion no type or amount of reasoning could give rise to any action. Only a passion to avoid negative feelings
and achieve positive ones can lead us to act. Once the initial goal of action is supplied by the passions, reason comes into play in a very significant way, but it can never step outside its bounds. Its role is always directive; reason can only point a passion to its proper object or to the proper means to its object. Hence, Hume stands conventional wisdom on its head with his infamous pronouncement that “reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them.”

Once we feel a desire or aversion based on a prospect of pleasure or pain, demonstrative reason can influence our conduct by directing us to that which will satisfy our end. Knowing the truth about things proves useful in getting what we want. Likewise, causal reasoning, by supplying information about the link between cause and effect can also afford us a means to fulfilling the ends supplied by our passions. But, according to Hume, causal reasoning has the ability to excite the passions in a way that demonstrative reason cannot. The idea seems to be that demonstrative reason is a priori and only functions on the level of ideas. Causal reasoning on the other hand is able to relate a present impression with the idea of some other idea constantly conjoined with it. It is thus able to produce a belief in the empirical existence of some new unobserved thing. Causal reasoning is able to produce belief in something which is a potential object of desire or aversion and in this way is able to excite the passions. By “excite,” Hume does not mean reason supplies any judgment concerning ends themselves, only the existence of objects corresponding to those ends.
There are only two possible senses in which we could even say irrationality is connected with the passions. One would be the case where a passion is founded on the supposition of an object that does not exist. The other would be a case in which a passion gives rise to an action that is not a sufficient or appropriate means to the end sought.  

Hume takes his reason-limiting conclusions a step further by arguing that it is passion or feeling alone that determines the normative spheres of aesthetics and most importantly morality. Moral blame or approval is always a feeling as opposed to an act of knowledge. Hume argues that moral qualities are not discoverable properties of objects but inventions of the mind and thus reason is incapable of determining the distinction between good and evil. Moral judgments are no more capable of being true or false than the passions, and for the same reasons. Moral judgments are just facts about the way we feel. There is nothing that they can more or less accurately correspond to on the basis of which we might describe them as true or false. In Hume’s words “’tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger.” Strictly then we must consider Hume a non-cognitivist.

The real basis for the categories of good and evil are our sentiments of approval and disapproval which attach to a given action or situation. We have a tendency to approve of that which causes pleasure and avoid that which causes pain not only for ourselves but for others, owing, as mentioned above, to what Hume calls sympathy (a capacity to enter into the suffering of others as if it were our own—something perhaps closer to what we would call empathy). So, for Hume the
foundation of moral value is public utility—a tendency to promote the interests of our species. In applying this standard of public utility Hume has not distinguished himself from many thinkers in the cognitivist camp. What distinguishes Hume is the fact that he does not claim this standard is grounded in any abstract or universal standard such as rationality, but rather in an appeal to the specific fabric and constitution of our species or the particular circumstances of a given community. Thus, while it is not contrary to reason to prefer the scratching of my finger to the destruction of the entire world, it is not in the interest of public utility, thus it is not in keeping with sentiments we quite naturally entertain, and thus it can be called inhumane and immoral, even though immoral does not mean the same thing for Hume as it would for a cognitivist.

Reason does have a function in ethical and political conduct in precisely the same way that it does in directing personal conduct towards the fulfillment of the passions. Once the motivating sentiment of moral approval or disapproval are present, the activities of reason play an indispensable role in disclosing the most useful, efficient and appropriate mean for maximizing what we consider good and avoiding what we consider evil. It is only through reason that we can actually assess what impact on public utility any given course of action might have.

So then, how is it that we so commonly come to the conclusion that there is a genuine conflict between reason and the passions? This comes about because of a confusion of certain calm passions with reason. It is argued that what is actually going on is a battle between certain violent desires and emotions and other more calm and peaceful ones which have sensations which are “not evidently different”
from those associated with the operation of the faculty of reason.\textsuperscript{11} An area of genuine weakness in Hume’s view may have been highlighted by Stanley Tweyman’s argument that this common confusion of passion for reason is inconsistent with Hume’s claim that it is evident simply from introspection that conduct is motivated by passion not reason.\textsuperscript{12} Tweyman claims Hume is simply begging the question when he says “’Tis evident...that the impulses arises not from reason, but [are] only directed by it.”\textsuperscript{13} Hume has not explained how we can be sure certain passions only seem like reason:

If the feeling associated with reason is in fact similar to the feelings of certain calm desires and tendencies then when Hume is trying to establish that reason cannot dictate actions his ground for this claim cannot be an assertion of introspective certainty or self evidence.\textsuperscript{14}

There is no space here for a real response to this line of criticism but I would offer the suggestion that although Hume’s arguments on this issue seem to beg the question, I believe his conclusions do have some support when we consider, not simply the local argument, but place Hume’s conclusion within the broader context of his entire philosophical project. That reason cannot dictate action is clearly not evident to anyone who considers the question carefully; (in fact it may not have seemed evident to anyone until it was “evident” to Hume) however, in light of Hume’s systematic arguments about the functioning of the mind, including his new understanding of precisely how reasoning is supposed to function, there is at least some reason to suppose the passions are beyond its sphere of possible influence. If this is right, it shows merely that significant criticisms must be targeted deeper, at Hume underlying account.
1.3) Critical Remarks

As covered in the last section, on the classical picture, the mind is able to use the faculty of reason to survey reality and discover what is of objective value. This act of rational discovery triggers desires which motivate actions in pursuit of those values. For Hume our aims or desires are already given and not subject to the influence of reason. Reason takes on an instrumental role in which the rational action is understood as that which has the best chance of satisfying one’s desires. On this view values are real but radically relative to the goals, abilities, circumstances etc. of each individual. What actions are appropriate to achieve these subjective goals is an objective matter and is the proper domain of reasoning of various types.

Now, if we grant, with Hume, that value is a subjective matter, it would be consistent to make the observation, that there is a great deal of intersubjective agreement about what sorts of things we actually do desire. Certainly many of our passions might be highly unique or specific but many others are shared virtually universally: the desire for continued existence, self respect, self expression, acceptance in a community of some sort, some degree of autonomy, food, shelter, sunlight etc. The recognition of this intersubjective agreement has led some thinkers, who endorse much of the Humean account of reason—John Rawls for example—to introduce an idea of basic or primary goods, which are not discovered to be good by reason but become so by agreement. However, might not reason itself

---

^A In the specific sense that it is actually the case that we value certain things—that our preferences are facts about us—but clearly not in the sense that things in the world have some objective property of being valuable.
be counted amongst such goods? I maintain that Aristotle overstated the case by claiming it is most fundamentally human nature to desire knowledge.\textsuperscript{15} Clearly, however, many people do desire to know—not a desire to know the correct or most effective way to fulfill some other end, but simply desire knowledge and the path of reason itself. Surely, it is at least possible that amongst a person’s many passions one might find a love of understanding. This possibility is not deeply incompatible with a Humean type approach to reason. However, I believe it would destroy the simple picture of dichotomous roles for passions as motivator and reason as strictly directing passions towards their objects, if reason were one of these objects.

Hume’s framework must be amended if it is to account for the fact that even if our passions are not arrived at by a process of reasoning, it can so happen, and often does, to at least some degree, that reason is one of our passions and therefore has the potential to motivate our action in a way Hume did not seem to acknowledge. The passion for reason can be so strong that it will cause us to seek truth and understanding even at the cost of frustrating many of our other interests. (Witness common sentiments such as “I just want to know the whole truth no matter how painful it will be.”)

While, as my preceding argument suggest, I believe Hume’s conception of the role of rationality is in some respects too restrictive of the actual place of reason in our lives, at the same time I find that when he does appeal to reason his standards of rationality seem too strict in some sense. For Hume:

\begin{quote}
nothing can be contrary to truth or reason, except what has a reference to it, and as the judgments of our understanding only have this reference, it must follow, that passions can be contrary to reason only so far as they are accompanied with some judgment or opinion.\textsuperscript{16}
\end{quote}
For example, Hume states an affection can be called unreasonable in this sense when it is “founded on the supposition of the existence of objects, which really do not exist.” Now, suppose you went to see a lecture about David Hume. At the end of the lecture you wait your turn, and at the appropriate time rise from your seat and address a carefully formulated question about rationality to the speaker (on account of interest stimulated by this paper). But while you are distracted with your attempt succinctly to pose your important query, a cunning and malicious prankster snatched your wooden stool and tossed into the roaring fireplace at the back of the room. By the time you finish your exchange with the speaker your stool has been reduced to coals. When you go to take your seat, before realizing what has happened, you fall crashing to the floor. Now, assuming that the question you asked made sense, I do not think it would be correct for anyone to describe you as irrational on the basis of these events, despite the fact that your desire to sit down on the chair beneath you, and the resultant action of attempting to sit on it were based on the supposition of something which no longer exists. Why? Because, on the basis of the evidence available to you at the time you had good reason to believe your chair would still exist and, moreover, remain just as you had left it. You had no good reason to suppose something like this would happen. The Humean picture does not account adequately for the possibility of reasonable but false beliefs.

Hume has also faced even more serious doubts because the radically sceptical conclusion of his philosophy, which cannot account for the objective validity of natural science, seem to undermine the empiricist project. Also, while it seems plausible that small communities could be held together and coordinate their
activities on the basis of norms derived from feelings of sympathy and trust, it is difficult to imagine how such feelings could be sufficient for the creation of norms that could hold complex societies together.\textsuperscript{18} In international diplomacy, nations, states, even large cities—we do not know the people with whom our actions must be coordinated. It is unclear how natural feelings of warmth and solidarity could govern our actions towards people we do not know. In fact such natural feelings of solidarity for others often conflict with what we would expect justice to ask.

Suppose my best friend asks for my last cork to keep his favourite wine fresh. I may feel strongly inclined to give it to him. Yet, when someone tells me that a bunch of foreigners in a leaky boat need my cork to save their craft I may feel indifferent to their plight and therefore disinclined to let them have it. But one would expect justice to entail that the people facing drowning have a stronger claim on my cork that my wine loving friend.

While I have reservations beyond the preliminary critical remarks developed here, there is no room for a prolonged critical engagement at this juncture. Hume's instrumentalist account is important for this thesis primarily because it stand in the background of a range of contemporary accounts with which I will be directly engaged in Chapter 3. There I will develop several lines of criticism directed at Humean inspired accounts as they appear in the 20\textsuperscript{th} and 21\textsuperscript{st} century contexts. For now I turn my attention to the enlightenments other most profound theoretical formulation of practical reason.
2) Kant: A New Paradigm in Practical Reason

Naturally, given the breadth, complexity, and sometimes (it must be said) unclarity of Kant’s work, and given also Kant’s nearly universal recognition as one of the most central and significant philosophical innovators in history, it is no wonder there exists a considerable range of interpretations of his critical architectonic. I will not focus here on competing interpretations, for this is a thesis about practical reason not about Kant. As is the case with the other philosophers discussed in this chapter my consideration is not aimed at forwarding an interpretation.

2.1) The Unity of Reason

Reason for Kant is ultimately just one thing. Reason is a faculty which seeks the necessary and unconditioned which, for Kant, is *a priori*. Practical reason and theoretical reason are different functions or applications of reason in general. The role Kant ascribes to reason has given rise to two innovative shifts in thought—Kant’s “Copernican revolutions.” The fist revolution is epistemological, concerns the

---

\textsuperscript{A} In the present discussion of Kant’s notion of practical reason I endeavour to follow, more or less, what I take to be a fairly traditional or received view in Kant scholarship. Specifically, that means taking the ethical formalism of the *Groundwork of the Metaphysics of Morals* as central and indeed, as its title implies, foundational for the later ethical treaties. I do this because 1) this interpretation has been tremendously influential 2) it is probably the interpretation according to which Kant is the most original, and most importantly 3) I believe we have much to learn from both it and its critics. I should like to note, however, that a range of thinker as influential and diverse as John Rawls and Gilles Deleuze have interpreted Kant in a different way: specifically, not placing such emphasis on his formalism, and, relatedly, tending to see less continuity in Kant’s moral philosophy, understanding the *Critique of Practical Reason* and the *Metaphysics of Morals* (written after the significant new theoretical developments of the *Critique of Judgment*) to be somewhat more discontinuous with the *Groundwork* and finding within these later works Kant’s most fruitful insights. I have some sympathy with these interpretations but do not discuss them directly, though relevant bits of such accounts may be seen to come out later when I consider certain neo-Kantian views on practical reason.

\textsuperscript{B} I think it worth explicitly noting Kant’s use of “reason” which is one of several key Kantian concepts used in a rather confusing way. It has two uses or meanings 1) *reason in general* also called judgment which includes *understanding* (theoretical reason) and *reason* (practical reason), and 2) *Reason in particular* which is practical and contrasts with understanding. Once aware of the distinction, usage is usually clear from context but I will generally qualify “reason” with “practical,” “theoretical,” or “in general,” to avoid confusion.
role of theoretical reason in perception, and is closely related and analogous to
Kant’s ethical revolution, in part, on account of the ultimate unity of reason. Thus, I
begin with Kant’s first revolution.

2.2) Theoretical Reason

Kant reasons from the fact that we experience the world in the way that we
do to the conditions that must obtain for this to be so. On this account, experience
begins with sensibility, the passive and receptive faculty of representation which is
affected through an immediate relation to objects. However, for Kant, there is more
to experience than appearances; experience, is also a matter of understanding.
Actual experience is never the simple sensation of an undifferentiated manifold; it
encompasses the activity of conceptualizing intuitions. The concepts necessary for
this purpose arise from a form of reason or judgment called understanding. The
understanding is that part of reason in general which unifies intuitions and makes
them into objects of possible thought. The concepts of the understanding relate to,
or have correlate counterparts in appearances. This is what distinguishes them
from “concepts of reason” (called “ideas”) which have no object of experience,
relating, not to any possible object of experience, but to the totality of all objects of
experience in general. (Reason does not unify appearances under concepts but
unites the rules/concepts of the understanding under principles.) Because it
relates to the sensible in this way, the understanding is the appropriate sphere to
deal with theoretical judgments (science/empirical experience), which contain a
reference to an "is" or "is not" as opposed to practical judgments which contain an
"ought" or "ought not" and fall under the domain of reason.
Some concepts are empirical—laws derived from experience by comparison, reflection or abstraction—but the concepts of fundamental importance for getting experience off the ground, are the *a priori* concepts of the understanding (the categories). They are laws that the subject subconsciously yet spontaneously/constitutively applies to the raw, undifferentiated, and unintelligible data of the manifold which issues from the purely receptive part of the mind. Understanding furnishes the individual with a set of rules that can (and must if one is to have experience) be used to give synthetic unity, (organization/intelligibility) to appearance. Only once concepts are applied to intuitions do we have “phenomenon,” the true correlate of our experience.

In recognizing the role of both sensibility and understanding Kant’s transcendental idealist account of experience is meant to be a middle position that combines the insights and overcomes the limitations of the two opposing accounts of experience offered by rationalism and idealism on the one hand (Descartes, Berkeley), and materialist empiricism on the other (Hume, Locke)—accounts which had dominated and divided earlier enlightenment thought on epistemology and ultimately ethics. Against a thoroughgoing empiricism that maintains experience is purely receptive, things-in-themselves are perceived directly, and space and time are counted among the independent things given in themselves, Kant stresses the constituting role of the spontaneous subject in experience. Pace the various forms of empirical idealism that maintain minds make or constitute matter, and despite his own emphasis on the role of the subject in experience, Kant argues for an empirical realist view of knowledge and experience that is objective and grounded in a
receptive capacity to experience, though he does so in a highly and uniquely qualified way. Not only were these views thought to be compatible but Kant argued that empirical realism was only possible given a transcendental, or as Kant sometimes calls it, critical or formal idealism. On this view phenomena are constituted by subjective structures (formal idealism) and within these structures (the transcendental conditions) it becomes possible for reality to be “immediately perceived” (empirical realism).\textsuperscript{22} This empiricism is radically different than Hume’s in that it takes account of the active role of the mind in forming experience.

Although it stops short of full blown idealism, Kant’s first Copernican revolution is the shift of epistemological priority from the object to the constructive role of the subject in experience. Further, in the \textit{Groundwork}, the role Kant ascribes to reason, this time when applied in the realm of action and value, yields what has legitimately been called a second Kantian-Copernican revolution.\textsuperscript{23}

\textbf{2.3) Kant’s Revolution in Ethics}

In no philosophical system is there a more direct and unqualified link between reason and morality than Kant’s own. Just as theoretical reason makes possible objective experience from subjective impressions, reason’s practical application is thought to yield objective normative standards in the realm of human action. For Kant, to be ethical is no more and no less than to be rational. If right action simply means rational action, then there is no longer any need for a substantive account of human good whether it be derived from an objective account based on theological/transcendent grounds (Christianity, Plato, Aristotle) or from a subjective account based on desires/preferences or shared beliefs (natural law
theory, utilitarianism, communitarianism, Hume). It is not that we use reason to pursue some other virtuous ends, rather, our virtue consists in obeying the moral law which is the expression of our own pure reason. No external authority, but a moral law within the rational self—this is Kant’s revolution in ethics. If successful, it means that we have within us universal moral standards that can govern our lives and harmonizes our interactions with each other without the need to rely on a true human telos or any such supposed superstition. This requires some explanation.

To begin with, practical matters concern action and things achievable through action. Insofar as we conceive of ourselves as rational beings we must conceive our action as undertaken for the sake of some value determined by reason. Indeed, we would not be inclined to call something an action that was not undertaken for the sake of some value. The sun coming up is not an action—it is simply an event. A worm crawling up out of the rain soaked earth is not taking an action either, but merely exhibiting a behaviour determined by instinct. Since reason seeks after the unconditioned, when applied to practical matters (action)—it must seek unconditional value. But the only value we find in things seems to be conditional not absolute. In other words, there is always an “if” that goes along with our valuations. (E.g. “It is a great cake if you like chocolate.” or “The collapse of the bridge was a terrible loss, if you have an interest in crossing the river.”) And where value is conditional all imperatives—all things that we should do—must be conditional as well. (e.g. “You should avoid the border crossings if you want to avoid a huge hassle.”) If no universal value is found then there can be no categorical imperatives—nothing that we should or should not do absolutely, without
conditions. Further, if there is no unconditioned value or corresponding categorical imperative then strictly speaking there is no practical reason—there is only prudential reasoning about means to subjective material ends we already happen to have (i.e. Hume). Such instrumental reasoning merely concerns natural casual/effect—asking, what is the most likely course of action to achieve a given result “Y.” In the absence of absolute value such reasoning would have no terminus. “X” is chosen for the sake of “Y”, “Y” is chosen for the sake of “Z”, and so on without an end.

Kant’s radical insight is to side with the thrust of early enlightenment thought and locate value not in the object but in the subject, while at the same time continuing to maintain the absolute categorical nature of moral law. On this view, there are not things out there in the natural universe which have value in themselves. Therefore, it cannot be our task to go around discovering what things or states are truly good and which are not, or what our proper role is within the divine order of things; rather, it is we who confer value upon things through the act of valuing them. As Christine M. Korsgaard puts it, for Kant “what makes the object of your rational choice good is that it is the object of your rational choice.”

No longer do we need to look outside ourselves for the ultimate good, for all rational being have the capacity to make things good through their own rational choice.

2.4) Kant’s Fundamental Moral Principle

A thing which a particular rational subject happens to value is called a subjective end because its value is not necessary but contingent on some persons valuing it—it is connected to the private purposes of a particular individual.
Nevertheless, insofar as it is the object of a rational choice, it is objective in the sense that having value conferred on it by one, makes it valuable for all. In other words, as a rational being you make things valuable by choosing them, but other rational beings have the ability to make things valuable through their choice as well. This is not a subjective moral solipsism where what’s good for you is different than what is good for me on account of the fact that any rational valuation of something makes it valuable not just for them but objectively so. You might like Cave Mold Beetles (which are endangered). That does not mean I have to like them; I may think they are disgusting. However, this would mean I have to respect their value by, for example, not trying to hasten their extinction.

Now it would seem that the things each of us chose to value would inevitably conflict with one another and therefore values must clash. But, Kant would claim this is impossible since to value things which are not compatible with the system of things others could value is contrary to the unconditionality-condition internal to reason itself. Thus, if we choose something which is not compatible with others’ ability to freely choose, that choice would not be a rational one, and as such would not confer value. So, for instance, no matter how much one likes torturing people, it could not become valuable. Because there are many rational beings capable of having rational interests that confer objective value, Kant argues that these interests must be subject to a formal universalizability constraint. This need for harmonizing one’s choices with the ability of others to chooses is captured and formalized by Kant in his first formulation of his one fundamental moral principle: “act only according to that maxim through which you can at the same time will that it should
become a universal law”\textsuperscript{27} (the formula of universal law). This is the one and only imperative that is categorical.

This fundamental principle functions as a sort of proviso which limits, in very significant ways, what actions we can rationally/morally pursue. Essentially it is a test which will only allow maxims which further a “systematic harmony of purposes in the individual and in the human race,”\textsuperscript{28} as H. J. Paton describes it. This principle does not give laws or commands; it is a procedure for adopting or rejecting maxims. Maxims are subjective principles of action and are presupposed by action.\textsuperscript{29}

Whenever we do something, and we know what it is that we are doing, we are operating according to some subjective principle. When we take a cookie from the refreshment table we may be operating according to the subjective principle “have a cookie whenever the opportunity arises.” Whatever subjective principles guide our actual individual actions are called maxims. We are free to value what we want and act accordingly subject only to the constraint that our maxims are universalizable—do not conflict with other rational wills.\textsuperscript{A}

Reason, in general and in the strictest sense, is logic, which concerns necessity—what must be the case.\textsuperscript{30} Indeed, for Kant, the form of practical reason is the formal principle of logical consistency.\textsuperscript{31} The categorical imperative is a binding principle of pure reason, and its formulation in terms of universal law is meant to express a purely formal constraint. Suppose I happen to like cookies. When I arrive at the refreshment table I would like eat ten cookies, but I know that cookies are

\textsuperscript{A} This, incidentally, is not a constraint, strictly speaking, but freedom itself, since it is an expression of our own rational nature, not something imposed by anything external. We will return to this point in detail below.
also a material interest of other rational being at the reception. If the first people to
the table ate ten, many others who like cookies would not have any opportunity to
have any. Kant would consider this to be a contradiction, because acting on such a
maxim could not be made consistent with others rational maxims. It could not be
willed to be an objective universal principle because it is not a principle all others
could share—it not possible for everyone to eat ten cookies. Only a universalizeable
maxim which could be adopted by all, such as “take only that amount of cookies as
would be compatible with others interest in the cookies” is rational. Thus taking
one or two cookies might be a universalizable, rational, and moral action whereas
eating ten cookies is inconsistent (with other objective values), irrational and
immoral.

On this view, the morally good life is one in which we test all maxims and
accepts only those which are rational—free from contradiction. But, not only does
the moral law have form (logical consistency) it also has matter—an end in itself.
The matter of the moral law is expressed in a second formulation of the categorical
imperative: “Act in such a way that you always treat humanity, whether in your own
person or in the person of any other, never simply as a means, but always at the
same time as an end” (the formula of humanity). Kant is adamant however that
this is not a second categorical imperative, but simply a different way of stating the
one fundamental principle of the moral law. However, since we (including Kant)
normally understand form and matter to be distinct, if the idea that these are two

^The formula of the kingdom of ends is essentially an elaboration of this formula of humanity.
ways of stating the same principle is to be intelligible, Kant has some serious explaining to do.

Things take on objective value because a rational will values them. This entails that we respect and recognize as worthy not only the things we valuate but the things any rational will values, which, in turn, entails the formal requirement that our maxims be compatible with a universal system of values. On this conception, the capacity for rational choice and its value conferring function has a special status—it is the ultimate source of value, and as such it must have a special dignity. The things that are valuable because we value them have a price, meaning something of comparable worth can be found, but rational beings, as the source of value, have a unique and incomparable worth. Instead of a price they have dignity. Everything of value depends on a valuer with the one necessary exception of the rational will itself—the fountainhead of all value. When we test our maxims to make sure they are universalizable, we are, by that very activity, respecting the equal dignity of all rational agents. Thus, for Kant, seeing rational choice as the source of value is equivalent to making humanity itself our necessary end. In other words, the (formal) principles and actions which rational beings find universalizable—that is, consistent with other rational beings—will be exactly equivalent to the (material) principles and actions that never treat another person as a mere means, but—as autonomous independent sources of value—accords such persons the status of ends unto themselves.
2.5) Freedom

To summarize, pure reason applied to human action expresses itself as the moral law which is captured by a single (variably expressible) categorical imperative that has both form (universalizability) and matter (the kingdom or commonwealth of all rational wills). But, Kant then goes on to argue that it is this moral law which is the only adequate expression of human freedom, claiming that the “making of laws must be found in every rational being himself and must be able to spring from his will.” Insofar as we are rational beings the idea that objective moral obligations derive from reason is the same as saying they derive from us as individuals. What Kant, indeed, what most reason based approaches to normativity have generally been trying to avoid is the result that obligation can only derive from some freedom-limiting external authority. These approaches face the difficult task of explaining why the imposition of some alien authority on a rational being is binding or legitimate. For Kant we are not only the subject of the moral law but also its author. The only imperative that can be legitimate for us is the one we create freely for ourselves, or in other words the only reason we could be “subjected” to a law is if the law itself were simply a free expression of our true self.

Freedom, for Kant, borrowing from Rousseau, is autonomy—the ability to be self-legislating, spontaneously self-caused, or the origin of our own actions. In a sense, this freedom is the justification for the rational subject’s status as the necessary end of morality. Unlike objects or animals a rational being is an end in itself because it has this special ability to direct itself in accordance with its own rational principles. Therefore, it would be an affront to the dignity of such a being to
illegitimately apply any external standard to direct or determine their capacity for free choice (note the influence of Spinoza here) as though a person were a mere natural thing like any other to be manipulated or used. For the will to be determined by anything external would be heteronomy in Kant’s language. The categorical imperative is a principle of reason which is binding because it is a self-imposed law. Thus, for Kant, being moral or doing one’s duty means the same thing as being rational which also means the same thing as being free in the specific sense of autonomous.

Now, Kant’s notion of autonomy is not merely a negative version of freedom which demands that one is left alone to do whatever one wants. It is, further, a positive conception which requires that one is moved by conscious purposes or reasons that come from within. On this view, “within” means issuing from our will—our rational selves. While action determined by an external authority, such as another person, is heteronymous, so is action determined by our own material drivers, instincts and appetites. One must recall that, for Kant, a person in the sense of some particular biological organism is not an end in itself, only a person in the sense of a power of autonomous choice is an end in itself. Therefore, humanity is an end in itself not qua species but only insofar as it is a kingdom or community of rational wills. Thus, freedom is not free in the sense of getting what you want, as in “I want to eat those cookies because they look sweet and tasty and it will make me feel good to eat them.” That is the opposite of freedom. If I base my action on such desires, that would simply amount to me being pushed around by contingent, subjective, empirical desires—by the casual laws that
govern the natural world. In effect it would mean giving up on being my own cause and letting myself be banged around by molecules—allowing the dignity of a free will to be reduced to a hunk of matter. To be truly free is to be an original, spontaneous cause of our own action; and this is the only conduit through which value can enter the world we experience. Recall that the world we experience (even after being organized by consciousness) comes devoid of value; it is just cause and effect—a bunch of particles knocking about in space and time. The only thing that could be moral is something that causes itself—something that is autonomous—and that something is a rational being who is governed by laws of practical reason, not the laws of natural cause and effect; and this pure practical reason can become a force in the material world through the medium of human action—the free and spontaneous action of a rational being.

2.6) The Possibility of Autonomy

At this point, given its foundational significance, one might expect to hear Kant’s proof for the fact that humans are actually the source of spontaneous supernatural action entirely separate from the casual laws that govern the phenomenal world, but instead Kant admits that human freedom is unprovable. From a theoretical point of view it is simply impossible to know, either way, whether we are free or not. Earlier empiricist approaches suggested our notion of freedom is an illusion because all that is observable in nature is empirical cause and effect. But, according to Kant’s scheme, there is no reason to think that the laws governing appearances apply to things in themselves. Thus, it would be pure dogmatism to think that a human qua rational ego is nothing more than a human
qua appearance. But, not only is it possible that we are free, it is necessary that we assume we are if we are to act—it is, in other words, a practically postulate.

When we go around doing things we all have the feeling and the idea that we are making our own decisions. We simply could not operate, that is, conceive ourselves as acting, without this assumption. If we did not suppose that our reason was the source of our principles and actions then we would be merely cogs in a system and we could not regard our action as our own or regard ourselves as moral agents. Since Kant takes it as sufficient for his purposes that freedom is necessarily presupposed by all rational beings in there action, his argument is meant to be preeminently practical as opposed to metaphysical.38

Although one must not look for its theoretical proof, in order to understand how Kant grounds ethics in reason, it is necessary to probe deeper into this notion of autonomy and the conditions for its possibility. Autonomy simply means that one’s actions are guided by the will. Before proceeding it is necessary to take a moment to clarify another terminological confusion, this time regarding Kant’s notion of the will. Simply put, for Kant “will is nothing but practical reason.”39 Will is decidedly not a capacity for choice as per our more ordinary non-technical use of the word. Kant uses different words for “will” (wille) and a “capacity for choice” (willkür) respectively. Part of the confusion must stem from the fact that some of Kant’s translators, including Norman Kemp Smith, have sometimes translated both terms as “will.”40 Once aware of the distinction it is generally clear form context which concept is intended. Will, (wille) as a rational “power of self determination,
independent of any coercion through sensuous impulses” is precisely that which ought to determine the capacity for choice (willkür).

The action of a perfectly rational being would be based on pure will. For them their subjective principles of action (maxims) would be identical with objective principles of action. For them the moral law would not have an imperative character. There would be no sense in the phrase “what a perfectly rational being should do” there is only “what a perfectly rational being does do,” which is necessarily right. Thus, what a perfectly rational being will do is what an imperfectly rational being (such as a human) ought to do. For us, a maxim’s rational character, while a motivating force, does not stand alone in its ability to influence our capacity to act (willkür). Our maxims are not always identical with the objective principles determined by reason independent of the sensible world, that is, they are not always autonomous. For human beings the rational/moral action is objectively necessary (it is an absolute truth that it is what we should do) but subjectively contingent (we might not do it because our action is subject to determination from external influences), to use Kant’s language.

When the capacity to act is bound by some object or end other than the will itself it becomes heteronymous. This can happen in two ways which previous moral theories have tried to develop into a basis for normative judgment. First, when our subjective principles are empirical—based on the pursuit of happiness or following feelings. Hume and the utilitarians try to base ethics on such empirical principles. The human faculty of desire includes feeling—the ability to take pleasure or displeasure in a given representation. When such feelings or impulses are allowed
to determine our capacity to act, we lose our freedom and act immorally. On Kant’s view, even when we act out of benevolent feelings (such as Hume’s sympathy) for others this is not part of virtue. The moral action is always done purely out of respect for the moral law—which lies within and is the expression of true freedom; it is never done because it makes us happy or satisfies some feeling. The fact that pleasure may happen in some cases to accompany a moral action does not detract from its goodness, but such feelings must not be the motivation—the fact that they are commanded by the moral law must be sufficient.43

A second type of heteronymous principle is based on the pursuit of a rational Ideal of perfection. For Kant, the problem here is that it is vague and does not tell us anything—“to be moral, be perfect.” Or else morality is contained in the definition of “perfection” and the argument becomes circular. One cannot use an ideal of perfection to define morality unless one already has an idea of what morality is. In its religious variant, the argument is that to be moral one must follow the command or example of God who is morally good, but to say that God is morally good presupposes we know what that is.44

In contrast to these approaches, Kant sharply distinguishes duty—action out of reverence for the pure ought of the moral law—with acting for the sake of any consequence such as happiness or achieving some, supposedly objective, externally-imposed ideal of perfection. For example, we should tell the truth simply because the categorical imperative commands it, not because it makes us feel good or fulfills some natural purpose. Unlike Aristotle, more often than not, what reason requires will conflict with what desires and impulses push a person toward—even a virtuous
person. Freedom/moral duty/reason entail resisting these inclinations. If we are to be free, rational agents who are capable of acting on our own principle—making our one decision for our own reasons—we must overcome all heteronymously determined action. This entails that our will must be an uncaused cause—a special original cause irreducible to the laws that order appearances.

Kant claims this is possible only if there are two worlds—the phenomenal world we experience and the noumenal world which underlies it. Onora O’Neill is probably right to call this the most controversial and fundamental feature of Kant’s moral philosophy. Whether these worlds are meant to be two separate metaphysical realities or two aspects of a unified world is no clear, and a controversy I will leave aside. What is clear is that theoretical reason, as outlined above, applies, *a priori*, to sensible intuitions which are caused by things in themselves that are not themselves known, but must be assumed to be the cause of the sensible manifold.

Similarly, practical reason applies *a priori* to free human action. Since free human action is purely rational it is not part of the sensible world that is known through sense and understanding; rather, it is part of the intelligible world of things in themselves that can be conceived but are unknowable through experience. Thus, practical reason gives laws for things in themselves whereas theoretical reason applies only to phenomenal experience. The moral law “not only points to a purely intelligible world but determines it...permitting us to know something which belongs to that world.” In other words we do not know the supersensible through sense, but we are aware of its law. This is why Kant can be said to prioritize reasons
practical application over its theoretical (perhaps for the first time). Humans are finite, rational beings. Insofar as we have a finite, spatio-temporal existence we are subject to the laws of cause and effect. If we get hit by a bus we get flattened regardless of what we happen to think about it. But insofar as we are will, we are subject to a different set of laws—the laws of pure practical reason which apply to the intelligible world and govern the kingdom of rational wills. Insofar as we are subject to the empirically conditioned laws that we apply to our experience, we have a heteronymous character; insofar as we follow the laws of pure reason, which govern the supersensible world, we are free.

2.7) Weaknesses of the Kantian Strategy

Now I will outline how Kant’s strategy has been made problematic by a range of serious challenges, an issue that will come up in subsequent chapters addressing neo-Kantian reformulations. I will focus here on difficulties that strike at the core of Kant’s general strategy of deriving ethics from pure reason. It is worth noting that many of the serious problems with Kantianism may leave the normative core of the theory intact. In other words one could broadly agree with Kant’s reasons for conceiving of morality as the fullest expression of reason while being critical of Kant’s specific view of morals. For example, Kant is often, and for good reason, criticized for supporting a system of rigidly insensitive rules.\textsuperscript{47} We end up with absurd consequences such as being morally forbidden from telling even the smallest lie to save any number of innocent lives because Kant’s universal law “cannot follow the sinuosity’s of experience...it rolls over them like a bulldozer and hopes for the best.”\textsuperscript{48}
In many contexts we need rigid formal rules as rough and ready approximations of ethics, but morality itself should not be cast in the mold of law, it should have more room to adapt to individual cases. No principle of action can be generalized to cover all cases; therefore, all moral rules must be subject to exception. Rules should be derived from general and universal principles, and, pace Kant, exceptions to those rules should be required in cases where the rule fails to embody the ideal of the general principle. This criticism, however, need not damage Kant’s general justificatory strategy if one were to reinterpret universalizability in a less rigid way, one which does not expect it to generate a static set of maxims that prescribe uniform action across the board of human interaction. What Kant failed to realize is that the general principle that we should coordinate our activity with other ends (rational beings) and their rational interests such that we can exist together as a harmonious community of ends, may be generally served by a policy of honesty, but when ends unto themselves are being destroyed in order to adhere to such a maxim the rule ceases to serve the principle.

Unfortunately, however, other criticisms cut closer to the bone. One of the most serious problems is the charge levelled by G. W. F. Hegel and many others, that Kant’s formalism is tautological, and therefore empty. Deriving a substantive moral law from the formal principles of pure logic seems like a straight-out logical impossibility. Insofar as Kant’s moral meta-rule is meant to test for consistence it does not seem to rule out anything, as Hegel claims in *The Phenomenology of Spirit*. The only maxim that contains a logical contradiction would be something as immediately absurd as “tell the truth and do not tell the truth.”
“Murder everyone you meet” or “destroy the world if given the opportunity” contain no logical contradiction.

The clearest type of case that Kant offers, in which a maxim is supposed to contradict itself, is when the consequences of the universal adoption of a maxim fails to preserve the means for acting on that principle. For example, lying if universally adopted would destroy the institution of honesty on which the deceitful act of lying is predicated. Nevertheless, the universalized maxims “always lie” or “lie whenever it suits your interests” are not themselves logical contradictions; everyone would simply lie until the institution of truth-telling is destroyed and then they would accept the consequence of not being able to lie anymore. This however, does not make the maxim a self-contradiction because the consequence of destroying the possibility of future lying does not contradict the maxim itself. The maxim “preserve the institution of truth telling and always lie” may be a contradiction, but the class of such contradictory maxims is extremely narrow and would not support the robust moral theory Kant envisaged.

While it is true that our self-interest may contradict the interest of other rational beings, I believe we would have to call this a material contradiction not a formal one. If such contradictions are supposed to be a problem, morally speaking, it would have to rely on the idea that a rational will is an objective end unto itself. Of course Kant does not try to prove this separate claim because he thinks that the form and matter of ethics are equivalent, but that does not seem to be so. The first formula of the categorical imperative (formula of universal law) requires nothing since virtually any maxim can be universalized whereas the formula of humanity
requires quite a lot. Even if we grant that the formal constraint of universalizability operates the way Kant says it does, the most we get is the idea that one must only act on principles that leave intact others' capacity to act, in other words, we are formally required to grant each other negative freedom. But, if humanity is an end in itself that would mean, not only that we do not act in ways that directly contradict other rational wills, but that we actively work to foster the conditions necessary for the capacity of rational choice. As Onora O'Neill puts it, the formula of humanity “requires that we support one another’s (fragile) capacity to act and to peruse their particular ends.”53 The idea that formal, logical rules are equivalent to substantive rules would seem to require a sort of conjuring trick that T. K. Seung calls “transubstantivation.”54 If Kant’s substantive ethic cannot be ultimately formal then he has failed to ground ethics in reason, at least in the way he supposes himself to have done.

Another way to get at the same problem is to notice that the Kantian idea of autonomy is that our true, rational self generates its own maxims, and yet the categorical imperative is only a meta-rule for testing maxims we are already motivated towards. If maxims must come from our bestial, material self we are back with Hume and it is hard to see how the principle of autonomy can be the measure of an ethical life. On the other hand, if not material impulses, where are maxims supposed to come from in the first place? No matter how hard Kant squeezes his ethical formalism it does not seem to generate substantive principle that could guide action.
A further but closely related problem concerns the separation of phenomenal and noumenal realms and the reliance of ethics on the metaphysical privileging of the later. For Kant, since no end is given to us by nature, whether that be fulfilling desires and inclinations (which he calls happiness) or achieving some empirically or metaphysically determined state of perfection, we must, as rational beings, be able to set our own ends. Insofar as we are self-caused, rational beings acting beyond the laws structuring the phenomenal world, we are supernatural. Freedom, reason, and ethics are predicated on our being a part of a mysterious unknown world of things in themselves. The Noumenal kingdom of ends to which we belong, Kant calls, a mystical body (Corpus Mysticum).\(^{55}\) He openly acknowledges the consequence that ethics could be based on nothing other than satisfying material interest if there were not this intelligible world beyond all sense.\(^ {56}\) Despite Kant’s ingenious and meticulous arguments, the idea that there is such a mysterious “world beyond” is itself a problematic idea. There is no room to do justice to all the relevant fundamental details of Kant’s critical architectonic but I need to say a few things about this problem. Kant’s ethics must rely not merely on the idea that there are necessarily unknown things in themselves which underlie our intuitions, but, further, on the superior reality or metaphysical privileging of that realm. As Kant puts it:

\[
\text{The law is not valid for us because it interests us (for this is heteronomy and makes practical reason dependent on sensibility—
that is to say, on an underlying feeling—in which case practical reason could never give us moral law); the law interests us because it is valid for us as men in virtue of having sprung from our will as intelligence and so from our proper self; but what belongs to mere appearance is necessarily subordinated by reason to the character of the thing in itself.}^{57}\]

Even if one grants that we are both finite (natural) and rational (supernatural) beings Kant does not seem to have an answer to the question: why is it that the law governing my will as a member of the intelligible world ought to govern in spite of the fact that that I am also a part of a sensible world? Kant does not want his ethic to rely on transcendent metaphysical assumptions of the sort that seem to be required here, but how else could one make sense of this privileging? Reason prescribes moral law. Moral law is only valid because it springs from our “real” selves and our “real” self is a supernatural will connected to a mystical noumenal body that has priority over the world of sense.

Not only does this privileging of an unknowable supersensible world need to rely on suspect metaphysical assumptions, it may also harbour dangerous social and political consequences. Kant’s sharp separation between our character, which stems from our rational will and for which we are morally responsible, and our temperament, which stems from the personal drives and impulses of our empirical selves and for which we are not morally responsible, means that we can ignore peoples actual expressed wishes in the name of the interests of their true self. When it comes to true freedom, what the empirical self wants is not the important thing, but, rather, what the purely rational self would want. Thus, positive freedom seems to come into direct conflict with negative freedom. One’s own aims and desires, as well as those of other peoples can and should be repressed and acted against in the name of a proper self. People are capable of autonomy but do not always act autonomously. If we determine that someone’s beliefs, goals, or actions are heteronomous we are free to trample over and repress them, if not duty bound to do
so out of respect for the dignity of their “real self.” This line of criticism is famously and influentially developed by Isaiah Berlin.  

One way Kant might be able to mitigate the impact of such criticisms is the idea of the “highest good” which he briefly adopts in the second critique and then abandons. The notion of the highest good replaced the moral good as the ultimate end. Here Kant maintains the same view of the moral good expressed through adherence to the categorical imperative but acknowledges that the natural good of happiness is also a legitimate goal for humans being insofar as we are finite beings. In this way, the absolute priority of the intelligible world is dropped. The highest good consists of both moral virtue and happiness as a reward of virtue, which must be directly proportioned to it (why they should be proportioned is not clear).

Unfortunately, this variation of the theory leads to even deeper problems. Since happiness and obeying the moral law are conflictual aims, Kant solve this antinomy by introducing two metaphysical postulates God and immortality. Since Kant maintains that there is no natural connection between virtue and happiness, God must intervene to ensure their proportion. Also, because achieving the highest good of fulfilling the moral law and having a perfect happiness proportioned to it will take an infinite amount of time, we must be immortal. It is this sort of argument that led Nietzsche to his apt description of Kant as a fox who cunningly escapes his captivity, only to turn around and slink back into his old cage. With this move Kant’s ethic loses its bases in human reason and freedom, resting, instead, squarely on the metaphysical assumption that there is an immortal soul and benevolent God. Instead of morality being able to prove God, as Kant originally intended, he seems to
return to a version (albeit a complex one) of the age-old transcendent strategy of deriving morality from religious beliefs.

Finally I should like to note that the criticisms Kantian or Aristotelian practical reason approaches face must be doubts about the functioning of reason in the practical sphere as opposed to doubts about reasons ability to motivate (content scepticism vs. motivational scepticism). It is necessary that practical reasons also be a motive (internalism requirement). That is why it does not make sense to ask, why should I do what I have a reason to do?\(^A\) If reasons could not motivate us then there would be no practical rationality. Given this psychological demand of the internalism requirement one can point to cases where a supposed reason fails to motivate action which is supposed to show a problem, namely, that reason cannot be the source of ethics because it does not motivate us. But, this does not refute practical reason. Just because reasons are capable of motivating does not mean that anyone who understands a reason will automatically act on it, for people are capable of irrationality. The fact that people make mistakes or are moved by things other than reason only shows that our rationality is imperfect. If one wants to refute practical reason they must do more than claim reason fails to motivate on particular occasions; they need to show that there are in fact no processes or operations of reason which yield conclusions about action.

Hume, for example, does not believe that reason can motivate the will (motivational scepticism) but this depends on the idea that reason is limited to the operations Hume claims for it (content scepticism)—that is, reason as the faculty

\(^A\) This issue will be revisited in connection with Habermas (Chapter 5 Section 2.9)
that judges truth and falsity with respect to representing the relation between
abstract ideas (math, logic) or objects (empirical science). And, of course, passions,
for Hume, are original existences not capable of being true or false, and knowledge
of the relations between things furnishes no universal conclusion regarding moral
action. If Humean scepticism is to be forceful we must accept the deeper theory that
reason is strictly limited to representing relations between things. 62

3) Conclusion

I hope to have shown, in outline, how the idea that reason can act as a basis
for normative theory has developed in the thought of some of the key theorist in the
history of this idea; and, in the process, to have shown that the concept of practical
reason is itself a complex notion. I also hope that the major alternatives discussed in
the last two chapters are clear, but I will conclude by summarizing some of the
important differences and similarities between them.

Both Hume and Aristotle give pride of place to theoretical reasoning. For
Aristotle practical reason leads us towards a perfect life, which is, above all, one of
theoretical contemplation. For Hume there is no practical reason per se, only the
employment of theoretical reasoning to aid us in our quest to satisfy our individual
desires. Kant holds that theoretical reason is less important, for theoretical reason
produces the concepts we use to bring sense under rules, whereas practical reason
goes beyond the phenomenal world and concerns itself with the "fundamental law
of a supersensible nature." 63

Hume and Aristotle also agree that realizing desires plays a central role in
ethics, albeit a very different one in each case. For Aristotle the pleasant and
desirable will coincide with the good in the virtuous, practically-wise person. For Hume there is strictly speaking no good or virtue, simply our desires, and we employ reason to satisfy them. Kant, on the other hand, understands reason as a constraint which filters or regulates our desires. Virtue consists in following reason even when it conflicts with our desires. Thus, a virtuous person for Aristotle will desire the good and act skilfully to realize those desires, whereas a virtuous person for Kant is not responsible for their desires, no matter how terrible, provided the autonomous will does not succumb to their influence.

For Aristotle, as with Kant, practical reason is one unified thing. Within our formal, rational capacity lies practical reason, within which Kant distinguishes the moral reason of the pure ought, expressed in the categorical imperative, from the technical and prudential reasoning expressed in merely hypothetical imperatives, whereas skill and prudence are as much a part of a virtuously ordered life as a good will, for Aristotle. For Kant skill and prudence are nothing but amoral powers or instruments which can be use by a bad will just as well as a good one.

For Aristotle virtue can reach a relatively stable and sustainable state—the harmonious perfection of the whole person with the greatest good of theoretical reason at the pinnacle. For Kant, virtue is always locked in an unremitting battle with the sub-rational heteronymous influences which must be pitilessly resisted by the practically rational will. For him, we follow reason simply because reason is freedom—our will expressing itself—not because we discover that it is our function within some natural order that we must fit into, whereas for Aristotle the “science which knows to what end each thing must be done is the most authoritative...and
this end is the good of that thing, and in general the supreme good in the whole of nature."

Having considered these major historical developments in the conception of practical reason and having identified some of the problems these accounts face, I will now offer an account of the current field, arguing that much of the dominant understanding of practical reason is woefully inadequate. Later in the next chapter I will begin arguing that key insights of Kant, Aristotle, and (to a lesser extent) Hume have continued to define many of the more fruitful contemporary articulations of moral rationalism, and ultimately that these insights might be consistently combined and developed into an adequate contemporary understanding of moral-practical reason.
Notes: Chapter Two

2Hume, 417.
4Owen, 76.
5Owen, 82.
6Hume, 266.
7Hume, 266.
8Tweyman, 136.
9Hume, 267.
10Hume, 267.
11Hume, 268.
12Tweyman, 132-133.
13Hume, 266.
14Tweyman, 133.
15Aristotle, 108.
16Hume, 267.
17Hume, 267.
18Habermas, 14.
19Seung, xvi.
22Kant, 1st Critique, 371A.
23Seung, xii.
25Guyer, 24.
26Seung, 109.
29Kant, Groundwork, 421n.
31Seung, xii.
32Bubner, 224.
33Kant, Groundwork, 429.
34Kant, Groundwork, 436.
35Guyer, 25.
36Kant, Groundwork, 434.
37Kant, Groundwork, 448.
38Kant, Groundwork, 448-449.
39Kant, Groundwork, 412.
41Kant, 1st Critique, A534/B562.
42Paton, 26.
43Paton, 19.
47 O’Neill, 182.
48 Gibson, 5.
49 Seung, 102.
52 O’Neill, 182.
53 Seung, 107.
54 Kant, *1st Critique*, A812/B840.
55 Paton, 45.
56 Kant, *Grownwork*, 123.
57 Paton, 45.
59 O’Neill, 180.
60 Korsgaard, 5-6.
61 Korsgaard, 7-25.
62 Kant, *2nd Critique*, 74.
63 Murphy, 263-293.
Chapter Three

Facing the Contemporary Dilemma
Introduction

So far, I have considered in some detail what I identified as the three main currents of thought on the topic of practical reason represented by Aristotle, Hume, and Kant respectively. No attempt to understand the contemporary debate would gain much traction without a clear picture of these landmark statements; for the most significant contemporary attempts to understand practical reason, as I hope to demonstrate, draw their major inspiration from these thinkers. In this chapter I want to move the discussion into contemporary terrain by doing three main things.

1) I want to begin by discussing what I think it fair to characterize as the dominant mode of thinking about practical reason in the 20th and now 21st centuries. Although popular, it is a mode of thinking which may not have indisputably dominated in academic philosophy nor always in the popular imagination, but it has certainly dominated amongst those economists, politicians, policy makers, and economically élite groups who have had the most profound influence on the shape of social, political, and economic practices and institutions which affect the everyday lives, activities, and relationships of the majority of people. This is a way of thinking related to a school of thought that became known as “rational choice theory” and it owes a profound intellectual debt to the Humean tradition.

I then want to suggest why this view of practical reason, a view that has come to have such a hold on our world, is deeply and fundamentally inadequate. My aim is to show that this influential way of thinking has inherited some problematic Humean assumptions and has taken this tradition in a cruder and less adequate direction than Hume’s own position and, certainly, a direction that fails to improve
on the previously discussed difficulties with Hume. I hope to challenge this understanding’s problematic theoretical assumptions and, by looking at the implications of this impoverished way of thinking in a 20th century global historical perspective, to point to some of the disastrous outcomes of putting these ideas into practice. The main conclusion is that alternatives must be sought.

2) In addition to the targeted criticism already discussed in connection with the three main philosophical paradigms (chapters 1 & 2) and those targeted at the now dominant rational choice model (chapter 3, section 1) late modern and contemporary thought has seen the rise of a series of significantly more radical critiques levelled at the entire notion of reason, truth, and objectivity in general, and ultimately the possibility of grounding normative principles in some form of practical reason. This line of thinking, forcefully inaugurated by Friedrich Nietzsche has been developed in a variety of different directions from Heidegger and Derrida to Foucault and Lyotard. As exemplars of this kind of critique I will briefly discuss the source of these radical critiques in Nietzsche, as well as a particularly clear and forceful contemporary defender, Richard Rorty. This line of thinking has become increasingly dominant in academia, if not in philosophy then profoundly so in other social science and humanities disciplines such as literary and cultural theory, anthropology, and history to the point that it now represents one of the most popular alternatives to the inadequacies of the dominate approach.

The purpose of this discussion is twofold. First, and most straightforwardly, I want to suggest why this form of critique should not be pursued as a viable

---

^1 I will also make brief reference to an important thinker, perhaps seemingly less radically, but still, I think, part of this camp—Alasdair Macintyre.
alternative. Second, I want to introduce a series of concerns which form part of the horizon against which contemporary reassertions of moral rationalism are set. In general, a common driving concern of the radical critiques is our apparent inability to transcend the contingencies associated with the unavoidable particularities, the situatedness, of our concrete historical, personal, and cultural perspective. Although I will not endorse the more or less radical turning away from reason as a viable solution, I do want to recognize some of the legitimate problems these critiques raise for moral rationalism. An important task for the remainder of this work will consist in assessing how contemporary theories of practical reason address some of these concerns.

3) Finally, having characterized (in objective 1 and 2) the contemporary dilemma—in short the gross inadequacy of the dominant paradigm of thought regarding practical reason, coupled with the inadequacy of a family of radical critiques which constitutes one of its leading alternatives—I will go on to consider solutions to this dilemma. As is clear already, or so I hope, my position is that a solution to this dilemma will involve reconnecting in some way with Aristotelian and/or Kantian ideas about reason and ethical value. Specifically, in this chapter I will make a first start towards this end by briefly considering a promising Neo-Aristotelian approach—Martha Nussbaum’s version of Capability Ethics. I will suggest that, while promising, this strategy may still be dogged by certain problems and blind spots. This will set the stage for my foray into the territory of solutions to carry on in the direction of contemporary approaches which draw their most explicit inspiration from certain core insight of the Kantian strategy—approaches
which impress me as particularly fruitful and which comprise a major area of investigation, vis-a-vis contemporary sources of resolution, within the present work. The next two chapters (4 & 5) will be devoted to a consideration of the two most profound contemporary attempts (that I am aware of) to reconstruct a version of moral rationalism on the basis of Kantian insight—John Rawls and Jürgen Habermas respectively.

I should like to note, finally, that this chapter moves particularly quickly over a wide range of topics. My purpose in objective 1 and 2 cannot pretend to be a fully adequate critique of the schools of thought considered therein. This would require a level of engagement which vastly exceeds the scope of this chapter. My purpose here, more modestly, is to map out what I have called the contemporary dilemma, the potential solution to which I then seek to explore. In order to afford this thesis its focus on said solutions, a brief characterization of what I take to be a dilemma must suffice in place of the detailed analysis which these competing schools of thought, no doubt, deserve. Similarly, the engagement with capability ethics admits to being highly selective. I seek only to briefly highlight some promising feature of the approach and point to some problems which necessitate, on my view, also looking to Neo-Kantian approaches for answers.

1) The Dominant Model

Individual freedom—historically, one of the abiding concerns of moral and political thought—might fairly be described as the ultimate political goal of our age. In order to tailor our behaviour and arrange our institutions and interpersonal relationships in such a way as to realize human freedom we would need to know
something about what it means for a human being to be free. In other words, understanding the meaning of human freedom entails an enquiry into human nature, and this has usually centered, in some way, on humanities unique capacity for reason. Because of the close connection between these two concepts, the paradox of reason identified by Horkheimer and Adorno could easily be recast in terms of a paradox of freedom. That is, like reason, the pursuit of freedom has repeatedly been transmogrified into its opposite. Witness today, a U.S. lead campaign to bring its brand of “freedom” to the world; a campaign that has taken the form of an aggressive militarism which has fuelled the rise of counter militarisms (“terrorism”) particularly those inspired by antidemocratic, authoritarian variants of Islamism, while at the same time, in an attempt to combat this backlash, many Western democracies are rapidly rolling back long standing laws designed to safeguard individual freedoms. In large measure, the reason for this can be traced to narrow and misguided ideas about reason which have trapped us into constricted, alienating and dehumanizing ideas about what it means to be free.

The narrow ideas about reason which are responsible for the paradoxical predicament of advanced capitalism are largely an heir of the Humean side of enlightenment thought, and were developed and formalized by game theorists working in the fearful and paranoid cold war period of the 1950’s. In this section I will outline some of the massively influential developments in this line of thinking about reason and make an argument about the impact they have had on human freedom.
1.1) **Rational Choice Theory**

After WW II, still largely under the influence of the British economist John Maynard Keynes, Western governments believed their role included significant steering performances designed to manage the economy and protect citizens from the economic chaos and uncertainty of unregulated markets which was thought to be the cause of the great depression. Unchecked, the self-interest at the heart of capitalism was thought to be dangerous, but intelligently regulated by state bureaucracy it could be harnessed for the benefit of all.

In the 1950 the tide of economic thinking began to shift towards the *laissez-faire* economics of Friedrich von Hayek. According to Hayek, the idea that society could be rationally planned or managed by governments for the public good was fundamentally flawed, since government officials and bureaucrats are not altruists interested in the public good, but, like everyone else, self-interested actors scheming to achieve personal advantage. Thus, big government, especially government interference in markets, inevitably led to tyranny and the end of freedom. The Soviet Union was taken to be the case in point.

The good news, however, according to Hayek, was that—so long as we can avoid having ideologues and zealots seduce us with romantic dreams of a society rationally managed by people interested in the public good who are then in a position to abuse their power to enslave us—we have nothing to fear from the self-interest which drives free markets, because markets themselves are a sort of self-directing automatic system. All people need to do is act to maximize their own self-interest, which is all rationality actually enjoins. Hayek's widely out of favour view,
considered discredited for most of the first half of the 20th century, began to receive support from the emerging field of game theory.

Purportedly, rational choice theory (sometimes called decision theory or rational decision theory) seeks to study and explain the behaviour of rational actors, whether these be individuals or groups, and in so doing, shed light on social, political, and legal institutions, policies and practices. Rational choice theory in its contemporary form began to take shape in the early 1940's through the collaborative work of mathematician John Von Neumann and economist Oskar Morgenstern, who together pioneered the field of game theory and its application to economic behaviour.

When rational individuals make choices in the context of a purely non-rational environment their choices can be modeled by classical economics in a relatively straightforward way. However, since human beings are profoundly social creatures, rational individuals are more often than not choosing within a social context, which is to say a context wherein the action that one rational agent will choose depends, not merely on the expected response or outcome in a natural, physical world, but also on what one can expect other rational choosers to do. This can be characterized as a strategic situation. The idea behind the game theoretic approach in rational choice theory is essentially that in studying strategic games (those in which players aim at the same clear and quantifiable goal predefined in the games parameters, and in which a given player must take another player or players interests and strategies into account) and the rational strategies for performing in
such games, we can develop a useful model for understanding all human interactions that are game-like, which is to say, strategic.

Rational choice theory directed its attention not only to individual choice in social contexts but, in the form of social choice theory and the closely related public choice theory, to the rational decisions of groups. Pioneering work in this field was done in the early 1950’s by the economist Kenneth Arrow. The basic approach was to model the rational decisions of strategically acting individuals and then try to aggregate those together in order to discover the rational decisions on a group level. Essentially public choice theory studied traditional problems of political science from the perspective of game theory, modeling the behaviour of politicians and voters as self-interested individual actors. Many of the central claims of contemporary economics would depend on the view of rationality at the heart of this school of thinking, one that Amartya Sen calls an “immensely popular”¹ approach used to “draw far-reaching policy conclusions in economics, politics and law.”²

Game theory was further developed, applied, and popularized by John Nash, a mathematician working for an American military think tank called the RAND Corporation. At the time RAND was working on nuclear war strategies to defeat the Russian Communists. Nash started to mathematically analyze games such as poker and prisoners-dilemma type games in order to devise the most rational strategy to employ—that is, how to work out what the other person is most likely thinking and how to respond in order to maximize our own advantage. What Nash calculated was that in games such as the prisoner’s dilemma where we have a choice between
trusting or not trusting each other for a potentially greater advantage, the best strategy is always to double-cross the other player and never to trust them. It also turns out, when everyone is free to pursue their own interest and employ a rational strategy of calculating what others will do in an effort to maximizing individual advantage, the result is always a stable, balanced system known as the Nash equilibrium—the mathematical demonstration of which subsequently earned Nash the Nobel Prize in Economics in 1994.

Applied to the cold war this meant neither the Americans nor the Soviets had any reason to trust each other by entering into non-proliferation or disarmaments agreements. Whichever side honoured such an agreement could be double-crossed by the other who surreptitiously ramps up production of its ballistic missile arsenal. In fact, regardless of which side honours the agreement, they would become the ones responsible for making things much worse for both sides by creating a power differential which greatly increases the chance of a nuclear war. But, if no one trusts each other and each side maintains as many nuclear weapons as possible, the result will be the stability and relative security of a distrustful stalemate.

Nash thought he could take his idea much further and apply the principles of his nuclear war strategy to the whole of human behaviour and show that a stable and prosperous society can be maintained on the basis of self-interest, suspicion, and treating others like adversaries to be outsmarted and defeated in a game. While one cannot simply stomp on Nash’s arguments with an ad hominem attack, I believe, It is also relevant to mention that not only was the cold war a dark and suspicions climate, especially at a military strategy organization like RAND, but John Nash
himself happened to be a paranoid schizophrenic who was deeply ill at the time he developed his theories, believing, for example, that everyone who wore a red tie was a communist spy. Nash’s mental illness may partially explain the paranoid and distrustful attitudes which seem to infect fundamental assumptions underlying his theories. But, regardless of the explanation, these assumptions are problematic. Nash’s equilibrium equations only work if we assume that humans are isolated, suspicious, strategically and instrumentally rational beings, constantly calculating how to defeat others in order to realize our sole motivating drive—the satisfaction of personal self-interest. Nash himself saw that his assumptions were empirically falsified. When the games were tested on the secretaries at RAND they always cooperated with each other. In fact in experimental trials the only people who played the game according to the rational strategy Nash predicts are known psychopaths and economists themselves. There is now extensive evidence that people do not act in the systematically self-interest maximizing way called “rational” on these models.

Despite such problems, game theoretic ideas about freedom through markets continued to gain momentum. In the 1970’s these ideas inspired, and in turn received further support from a new paradigm of thought in the biological sciences—the gene perspective. According to Richard Dawkins, a leading proponent of gene perspectiveism, the primary target or unit of biological evolution through natural selection is the gene itself. All life is a struggle or contest to persevere, through survival and reproduction, played out by the individual genes themselves. A game theoretic analysis can then be used to understand the competition between
self-interested genes. Organisms are ultimately just machines used by genes as a strategy in their game of survival. All human feeling, culture and behaviour including things that appear to be altruism, are driven by instructions coded in our genes, which aim solely at the ultimate purpose of their own replication. Our behaviour is essentially an epiphenomenon—purely a function of our genes struggle to maximize their own advantage. In a way game theory had risen to the level of a sort of theory of everything in the biological and social sciences, applicable on the most fundamental level to genes, on the medium scale to organisms, and on the macro level to groups of organisms (superorganisms) and ultimately human civilizations.

In this intellectual climate, prominent biologists, even those who criticize Dawkin’s gene perspectivism as overly simplistic, such as E. O. Wilson, who works on understanding social insects that function as profoundly integrated units (called superorganisms) by “analyzing the self-organization of colonies based on simple rules of individual worker behaviour,” could unhesitantly aspire to some rather lofty goals. For instance a goal Wilson and his collaborator Bret Hölldobler put thusly: “we hope to assist in establishing more clearly the relevance of sociobiology to the general principles of...systems theory.” Simple algorithms are seen as the key to understanding, not just ant colonies, but all complex systems including human social system.

This way of thinking even influenced thinkers who saw themselves on the political left, but who, nevertheless, found themselves, even if unwittingly, the unlikely intellectual allies of the conservative economists of the political right. A
case in point is the psychiatrist R. D. Lang who came to believe that the schizophrenia of his patients as well as a host of other human problems where actually being caused by the situations of their family life. Lang used game theory to analyze how power and control functioned in daily family life. He argued that the family was not the caring and nurturing environment it had been naively taken to be, rather, it was a strategic arena where acts of “love” were used as weapons to manipulate and control one another. The family was often a toxic and damaging environment which is, by and large, a mechanism for controlling people. This meant even those who claim to love us should be treated with suspicion. Lang became a central figure of the anti-psychiatry movement and an inspiration for a countercultural movement distrustful of institutions claiming to work on behalf of our best interest.⁴

Late 20th century theorists on the right and the left often converged in their criticism of those who held power in societal institutions like the state and even the family. In both cases the critique is carried out in the name of negative personal freedom. Informed by the rational choice theory of rationality (the so called homo economicus view of human nature) the notion that other people, including political representatives and civil servants, could have wider goals and act out of interests previously held to be important, patriotism for example, was now thought to be naïve. Politicians and bureaucrats can never actually be concerned with the public good, only helping themselves.

⁴ A good example is Lang’s early work The Divided Self: An Existential Study in Sanity and Madness (1956 Toronto: Penguin Books, 1990). His work is also chronicled in a 1972 verite style documentary directed by Peter Robinson called Asylum.
This was not widely seen as a catastrophic result. As economists, such as James Buchanan argue, zealots and ideologues who think they know what is best for people and cannot be dissuaded from pursuing grand visions for society are actually very dangerous, while “rational,” self-interested people can predictably be influenced by incentives. We do not have to fear self-interest because we can design, or simply allow for the quasi-natural course of, certain systems wherein the force of self-interest is harnessed for mutual benefit. This picture has two key practical conclusions: 1) less government and 2) systematic analysis and a narrowly defined “rational” administration of government based on incentivization.

I will briefly elaborate what I have in mind here. 1) If we are *homo economicus*—bound together primarily as individuals rationally competing to satisfy self-interest requirements, rather than as citizens or community members, whatever that may entail—markets can largely replace politics as a means for coordinating our interaction. Free markets are thought to respond to people’s needs better than democracy. 2) The idea here is that the corrupting potential of subjective values and emotions must be purged from governments and any remaining government-delivered services. This is achieved by implementing systems of incentives and “rational” objectives with clear quantifiable performance targets. With everyone—elected officials, doctors, soldiers, teachers—operating predictably as isolated self-interest maximizing machines, things like government bureaucracy and the economy, can then be scrutinized objectively using systems analysis. In the 1980’s Ronald Reagan and Margaret Thatcher, employing the help of U.S. economist and systems analyst Alain Einthoven, (another RAND nuclear war
strategist) began enacting these strategies by overhauling the scope and function of government. In the following sub-subsections I will take at closer look at some of the assumptions underlying these shifts.

1.2) **An Immanent Critique**

Above, I characterized rational choice theory by stating that it “purports” to be a discipline which studies what it is for actors to make rational choices. By presenting itself in this light “rational choice theory,” through an act of self-definitional fiat, has set itself up as the special field of study concerned with the workings of practical reason. Clearly this is misleading. First, as I hope is evident already, “rational choice theory” is not so much a field of study (e.g. metaphysics) as it is a particular view one might hold about the topic under investigation by a given field (e.g. Cartesian Dualism).

This however is also somewhat misleading. Within a given discipline, what one generally refers to as a “view” or a “position” on some topic which can be called the subject domain of the field of study in question, generally consists of an articulation of a way of conceptualizing or making sense of that topic coupled with a set of reasons why one ought to conceptualize the topic in this way— in short, an explanation why one ought to take this view. Typically this is done with reference to competing conceptualizations of the fields subject domain which are argued to be less adequate. Rational choice theory is not a “position” in this sense. While it does contain a characterization, or family of related characterizations, of practical reason it scarcely, if at all, defends that view and certainly fails to engage in any meaningful way with the long and sustained history of theorizing about practical reason. Thus,
rational choice theory is less like a theoretical position on the nature of rational choices, and more like an empirical research program based on a set of assumptions about rational choice. Strangely, a theory of rational choice is not the aim, nor even a component, of this “discipline” whose subject domain is purportedly rational choice. Instead its aim is to model, predict and then regulate social, political, legal, and economic interactions by applying a set of undefended assumptions about rational choice. Why rational choice theory appears more like a field of study than the articulation of a particular theoretical position may be accounted for by the fact that it is actually more like a research program which applies a form of analysis based on assumptions about rational choice.

Just because rational choice theory begins from assumptions which it is not overly concerned with defending does not entail that those assumptions are incorrect or not well founded on the basis of argument provided elsewhere. Indeed these assumptions are not drawn from thin air; to a large degree, they are drawn and developed, (not always consistently) from the Scottish Enlightenment and British philosophical traditions (particularly empiricist and utilitarian e.g. Hobbes, Locke, Smith) in which they were defended with arguments. While borrowing assumptions from great thinkers might represent an improvement over arbitrary or convenient assumptions it clearly remains insufficiently critical. The point here, however, is that by taking on rational choice theory one indirectly engages a host of other thinkers in a long tradition—perhaps the most important of which is Hume, who offered a systematic and well defended account of reason which is pivotal for a
whole tradition of thinking about practical reason generally and contemporary rational choice theory in particular.

For some reason the link is very often stressed between rational choice theory and utilitarianism. While lines of influence do exist here, there are major differences, chief amongst which is the fact that for utilitarians, such as Bentham and Mill, to be practically rational, specifically to act morally, meant acting in a perfectly disinterested way in the interest of the overall or general social good, an idea quite alien to mainstream economics. Rational choice theory is much closer to Hume, who may have been an important precursor to utilitarianism, but, despite sometimes being labelled as such, was not strictly a utilitarian himself. The principle of utility is not normative for Hume. He did not employ pleasure and pain in order to address the question; what aught we to do? Hume simply appealed to pleasure (agreeable feeling) and pain (disagreeable feelings) in order to explain why we approve of certain things and disapprove of others. For Hume, as with rational choice theory, reason is wholly instrumental to subjective desire. Of course, for Hume, we are carried by our natural capacity for sympathy (often de-emphasized in rational choice theory) to a concern with the public good. This, however, has nothing to do with reason and everything to do with natural sentiments we happen to have as well as the customary standards that make up the fabric of our community. This made Hume a sort of communitarian and proto-sociobiologist. Both Hume and rational choice theory explain choice, even seemingly other-regarding choice, ultimately in terms of egotistic feelings of pleasure and pain. Thus, Hume can be regarded as the grandfather of the dominant self-interest paradigm of
rationality. Further, since all rationality can do is serve as an instrument for our basic passions, on this account, Hume was led to defend the view of practical rationality as a maximization activity. With our ends fixed by our biological nature and society's customs, rational activity becomes simply an instrumental endeavour to maximize the impressions we like and avoid those we dislike.

Notwithstanding its Humean intellectual pedigree a number of problems, some of which I have touched on already, have emerged for this outlook. A particularly compelling critique targeted at assumptions which lie at the very heart of rational choice theory has emerged from within the field of economics itself. This critique is represented in the work of Amartya Sen. Sen is a specimen of that most rare creature—a real economist—one who understands economics, at least in part, as a branch of moral theory. Economics, properly understood, must be concerned with questions such as: how aught individuals to organize themselves in order to meet their own needs and goals while performing the activities necessary for the maintenance, reproduction, and development of society? how should material benefits and burdens be distributed amongst individuals? what sorts of systems are necessary to achieve this distribution? and, how should these systems operate? The most fundamental of these are moral questions.

Mainstream economics presents itself as a branch of applied mathematics. It claims to address technical question through mathematical models which describe the economic reality. In actuality, mainstream economics is always thoroughly engaged in making normative pronouncements. A large part of what an economist does involves making pronouncements about what ought to be done. It is standard
for an economist to make claims such as: “We should raise interest rates. We must find a way to get people spending more. Business taxes ought to be lowered to attract international investment. We should allow a deficit this year to stimulate growth.” Unlike the mainstream of crypto-normativists masquerading as mathematicians, Sen is self-reflective about the nature of his discipline. He recognizes that, while economics is clearly concerned with some purely factual and technical questions, (e.g. what would happen if interest rates go up?) he recognizing the necessary connection to normative concerns (Would the results of raising interest rates at this time be good or bad?). Sen would no doubt be in accord (to foreshadow my comparison with critical theory) with Max Horkheimer’s claim that the “critique of economism…consists not in turning away from economic analysis but in engaging in it more fully.”

As a properly philosophical/normative economist Sen is critical of rational choice theory’s “widely used but narrowly formulaic views of rationality.” He identifies and critiques three standard characterizations of rational choice in mainstream economic theory as: “(1) internal consistency of choice; (2) self-interest maximization; and (3) maximization in general.” I will briefly consider some of the main problems with these assumptions.

The first and most obvious problem with internal consistency taken as a sufficient condition of rational choice concerns the idea of consistency. The problem is that one can be consistently moronic. A person who always chooses what he or she hates the most and values the least would be thoroughly consistent but clearly the antithesis of rational. Consistence, then, does not get rationality very far off the
ground. There is, however, a much deeper, more fundamental problem which
discounts internal consistency even as one in a series of necessary defining features
of practical rationality. This concerns the very notion that consistency is something
internal to choice.\textsuperscript{12} Sen shows that this idea is foundationally misconceived. The
error is to suppose that choices can be consistent with themselves, internally,
without any external referent. There is no way to decide if any given choices are
consistent with one another without considering what objective a person aims at
with their choice—which is to say, the motivation. This, however, introduces factors
external to the act of choice. As Sen says “what appears to be conditions of internal
consistency are typically the implications of external correspondence with some
standard.”\textsuperscript{13} So, in order to understand choice we need to consider the rationale
which lies behind choice. Choices aim at goals, strategies, preferences, values, and
so on. We can check to see if a series of choices consistently aims at some external
goal or conforms to some external standard, and insofar as they do a certain
coherence between choices will exist, however there is no way in which this is a
property internal to choice.\textsuperscript{14} Because the exclusive pursuit of self-interest
maximization would tend to result in a narrowly and consistently focused set of
choices, mainstream economics has often, quite mistakenly, understood the self-
interest maximization and internal consistency criteria as closely related. The self-
interest maximization view of practical rationality, whatever deficits it may have,
does not suffer the same conceptual incoherence as internal consistency; here there
is an external, albeit narrowly defined, objective, with reference to which patterns of
choice can be described as consistent or inconsistent.\textsuperscript{15} Nevertheless, this view is
also a grossly inadequate account of reason despite the fact that it is “effectively
dominant in contemporary economics” and is a presupposition of the central theorems of the field.¹⁶

The internal consistency view is one particularly acute example of how rational choice thinking in economics has incoherently attempted to analyze practical rationality in a neat, technical, and formulaic way completely cut off from any account whatsoever of values or interests. Like the internal consistency account, which effectively attempts to sever the rationality of action from any externality, the self-interest account would block rationality from any externality save one—a narrowly conceived account of personal well-being or self-interest. In the contemporary context, the justification, if there is one, for this one motivating value is often thought to derive in some way from our biological existence as subjects of a natural selection process. Human beings, like any naturally evolving biological organism strive for their individual survival, material security, bodily well-being, and ultimately reproduction. Our only goal, self-interest, is taken as given, fixed by our biological nature. Practical reason is purely instrumental to this one goal.

If our only rational end is maximizing our own self-interest and this is selected by evolutionary processes, or otherwise predetermined, then the idea that we can rationally select our own goals and actions is excluded from the outset. Ruled out is the possibility that there is a conception of the good life for a human being which can be discovered and pursued (Aristotle). Ruled out is the idea that in living with other human beings reason makes certain categorical demands on us
(Kant). Even Hume’s idea of other-regarding passions and Adam Smith’s idea of moral sentiments may be ruled out. Excluded from rational choice in general are any reasons, any values, any commitments, be they social, political, moral, or whatever else which go beyond personal welfare gains.

The fact that people seem to have other goals and commitments, including ones that prompt seemingly selfless and other-regarding behaviour, has not escaped the notice of rational choice theorists. Such cases, if they are to be counted as rational, are explained or reinterpreted as being in the ultimate or long term interest of the individual. Cooperative or self sacrificing choices can be counted as rational if some, often complex, chain of instrumental reasoning can be adduced to show that the direct cost to the self is likely to be offset by a larger payoff overall. As Sen notes, this gives rational choice theory an “almost forensic quality, focusing on the detection of hidden instrumentality.”

The possibility that apparently non-selfish behaviour can be connected to a person’s long term self-interest does not entail that this is the reason or motivation for the action. The factual claim that such a link exist in any given case does not negate the possibility of moral reasons for action. Likewise the fact that certain self-regarding goals prescribed by biology exist, does not negate the possibility that those goals can coexist or even be superseded by values and principle not selected by evolution but selected by human reason (e.g. people sometimes risk life and limb in the pursuit of objectives that do not make them better off simply because they are convinced that it is the right thing to do).
According to Sen "the insistence, which, alas, is rather common in parts of economics, that a person cannot reasonably value anything other than her own well-being does little justice to the reach of reason." The main problem, then, is that rational choice theory “has tended to choose, fairly arbitrarily, one very narrow interpretational story, rejecting other rival understandings of what can lie behind the regularities of choice,” and this has “exercised a particularly limiting influence on the explanatory reach” of this theory. Contrary to the formulaic systemization of rationality as selfish, instrumental maximization, our:

choices need not relentlessly follow our experiences of consumption or welfare, or simply translate perceived goals into action. We can ask what we want to do and how, and in that context also examine what we should want and how?

Sen’s main concern here is not to defend a full account of what goals and values practical reason commits us to beyond self-interest. Upon a self-reflective evaluation of our life and goals we may or may not be moved by moral considerations. His point, which I think should be well taken, is that rationality cannot actually require us to be selfish and mean by abstaining from reasoned self-scrutiny and eschewing commitments other than our own welfare. Whether or to what degree we accept such commitments must be decided in a process of rational scrutiny, not summarily ruled irrational by some formal methodological algorithm that decides the matter before that process has begun. In fact, taking the pursuit of self-interest as the necessary aim of our activity undermines the most distinctive and “profound capacity of the human self” the freedom to reason about what we should pursue. It is freedom of this sort which is neglected and undermined by the dominant
economic model which finds it adequate to leave us free to peruse self-interest through consumer choices on a “free” (i.e. unregulated) market.

Instead of trying to capture practical reason under a formalized set of rules, Sen argues that it should be understood quite generally as “the discipline of subjecting one’s choices—of actions as well as of objectives, values and priorities—to reasoned scrutiny.” Given this more inclusive understanding of practical reason, Sen’s objection to rational choice theory’s focus on maximization in general should come into relief. No maximizing behaviour can be wholly constitutive of practical reason, or for that matter human freedom. We could not call rational a person who consistently and ever so effectively acts to maximize something which is absolutely terrible. This is why it is so important for Sen that reasoned scrutiny applies to objectives (as it does for both Kant and Aristotle). Further, while maximizing activity may be an important part of instrumental rationality, full practical rationality, also requires the reasoned assessment and defence of goals and values. And this, I think important to make explicit, also seems to be a requirement of genuine human freedom.

Sen’s idea that reason is involved in understanding and assessing our various ends as well as in using these ends to make systematic choices about action, stands in opposition, not only to the particular overly-simplistic formulation of rational choice in terms of internal consistency, maximization, and self-interest, but more broadly to the “general presupposition that rationality can be made into something of a given formula, without any further procedural content.” I will have something more to say about Sen’s constructive alternative account later in this chapter. For
now, suffice it to say that while I agree that "it is important to reclaim for humanity the ground that has been taken from it by various arbitrarily narrow formulations of the demands of rationality" and indeed that Sen offers a more adequate account than the considered alternative, his account does remains a bit too general. Specifically, while I agree that part of the strength and appeal of Sen’s approach is its generality, and while I would oppose the charge that Sen’s account of rationality is empty, I believe it is possible, without lapsing into any kind of formulaic oversimplification, to flush out considerably the “procedural content” of rational scrutiny. As I will argue in Chapter Five, Jürgen Habermas goes some considerable distance in this respect.

Not only is there a point of contact here between Sen and contemporary Habermasian critical theory but also an important alignment between Sen and the critical theory tradition generally, including the first wave of the Frankfurt School—namely, both are leading critics of the dominant contemporary accounts of rationality—Sen from within the discipline of economics itself, and critical theory from quite radically outside anything it regarded as traditional theory. While the early rational choice theorists were busy devising formulae for calculating self-interested instrumental and strategic choices—so called “rational” choices—the Frankfurt Theorists were already countering these dominant accounts with powerful counter arguments while struggling towards an alternative conception of moral/practical reason.
1.3) The Frankfurt School Critique

Central figures of the Frankfurt school (e.g. Herbert Marcuse, Max Horkheimer, and Theodore Adorno) remained committed to key Marxian insights and continued to aim at an essentially Marxian goal. They remained, for instance, committed to a dialectical method, the critique of political economy, and to a political utopianism which sought a truly rational, emancipated, dis-alienated society. Critical theory aimed at releasing human potentiality by overcoming an obsolete mode of production. Here we see the central significance of Marx’s critique of political economy. As Horkheimer put it:

The critical theory of society is, in its totality, the unfolding of a single existential judgment...the theory says that the basic form of the historically given commodity economy...after a period of progress, development of human powers, and emancipation for the individual, after an enormous extension of human control over nature, it finally hinders further development and drives humanity into a new barbarism.27

This Marxism, however, was not orthodox. The critical theory tradition sought to update Marxian analysis, in part, by incorporating important insights from various other thinkers, such as Max Weber and George Lukács. Perhaps most centrally, they complicated Marxism by positing an autonomous cultural sphere. They studied the important function of cultural and ideological factors which were not seen as essentially unimportant epiphenomenal reflections of material, economic factors. This meant a departure from any attempt to comprehend civil society primarily in economic terms. As Simone Chambers puts it, civil society was now “seen as a system of ideas, values, ideologies and, yes, interests understood primarily in sociological and political terms.”28 Nevertheless, there was, in
Horkheimer's words, “fidelity amid all change” to “the idea of future society as a community of free men”\textsuperscript{29} a vision which could be defined in opposition to “a society in which material and ideological power operate to maintain privileges.”\textsuperscript{30} This is the defining utopian goal at which a critical theory aims.

Under capitalism, wherein cultural material is produced almost exclusively to be exchanged for money, culture has become business. Importantly, critical theory has analyzed the rise of this culture industry. Mass consumer culture, it argues, has created false consciousness and false needs. Capitalist reproduction and expansion requires constant growth. Advanced capitalism, having vast, technologically mediated, productive capacities has faced the historically unique problem of creating consumption to match this production. To remedy this, mass culture, delivered through an array of mass technological media, creates false needs—which are desires that have been manipulated rather than the result of free, rational deliberation. At the same time it creates a false consciousness, making economic relations appear natural, obscuring the fact that culture and our relation to it could be different and moreover that human being have an objective interest in a radical transformation of these conditions. In such a state, individuals can be allowed relative negative freedom while remaining docile and subservient to the interests of an élite class, which run counter to their own rational interests and real freedom.

The culture industry fosters a homogenized worldview based on self-seeking individualism, consumerism, and the identification of personal worth and identity with ones market choices and consumer lifestyle. In effect it actually helps level people down into creatures approximating more closely the “rational beings” (that
is, irrational beings) modeled by traditional economic theories. Creating this homogenized mass-culture centered on similar consumer aspirations and leisure pursuits also creates an apparent levelling of class differences which occlude the enduring class distinctions from view, and in doing so protects the interest of the dominant class. On the plus side however, even the power of the culture industry is unable to nullify human beings rational nature. Critically theory’s challenge to existing society and the traditional theory that supports it, as well as its articulation of an alternative, are carried out in the name of reason. Current society is irrational, while a free or emancipated society would be a rational one. Thus, Horkheimer can quite succinctly summarize the goal of critical theory as “the rational state of society.” Bleak as the situation may seem in a society held to be so deeply entrenched in an irrationality effectively masquerading as its opposite there is a cause for genuine hope that stems from human beings rational nature. As Horkheimer puts it “the thrust towards a rational society, which admittedly seems to exists today only in the realm of fantasy, is really innate in every man.”

The problem with existing free market capitalism is that it is chaotic and irrational. It is not simply that rational decisions are not made but that mechanisms by which collective decisions, rational or otherwise, can be taken on matters of great import to human life scarcely exist. As one commentator puts it, critical theory criticizes the general “tendency of cultural commodification to destroy an authentic realm of public discourse within which people can mount challenges to the

---

\(^{A}\) Arguably, some later works of the Frankfurt Schools thinkers became more radicalized in their critique of reason and more deeply pessimistic about the possibility of achieving such a rational state. I do not deal with these issues here.
dominant values and institutions.” In the absence of effective public democratic spheres, instead of rational planning, the sorts of projects that get pursued by society and the distribution of benefits and burdens of those projects gets decided by market forces, as if by inevitable natural forces. To ask a traditional economist “should we stop burning fossil fuels to counteract catastrophic changes to our biosphere?” would be like asking an astrophysicist “should we fly over to visit our neighbouring galaxy?”—just as the scientist would tell us the question is moot because intergalactic travel is physically unrealistic given our current understanding of the universe, an economist would explain that not burning fossil fuels would be economically “unrealistic” just as economists of the past had explained the economic impossibility of ending slavery. How much fossil fuel, if any, gets used is the result of market forces—a competitive interaction between self-interested individuals—it is not something that can be directly decided.

Critical theory would agree that previous social formations are primarily the result of various sorts of strategic interactions but that they need not be and indeed should not be. In Horkheimer’s words “The existence of society has either been founded directly on oppression or been the blind outcome of conflicting forces, but in any event not the result of spontaneity on the part of free individuals.” Rationality for the critical theorist means making society the result of free conscious decision making processes, which is diametrically opposed to the rational choice model which sees rationality as individualistic strategic self seeking.

The bourgeois type of economy, despite all the ingenuity of the competing individuals within it, is not governed by any plan; it is not consciously directed to a general goal; the life of society as a whole
proceeds from this economy only at the cost of excessive friction, in a stunted form, and almost, as it were, accidentally.\textsuperscript{35}

“Rationality” according to traditional economic models is quite compatible with the blind, accidental, and in fact destructive direction of society. The rules of capitalist economics ensure that what happens is, by and large, that which will maximize profit for investors regardless of what people think about it or what happens to be good for people. These narrowly instrumental conceptions of rationality are not arbitrary. They promote narrow instrumental ways of thinking and acting which lend support to the existing economic order, which in turn benefits currently dominant groups. Traditional theory serves a function within an existing hierarchical socio-political system. Theory in general (to simply seek an understanding of some subject domain), when unconstrained, is consider useless at best, and at worst quite dangerous insofar as it can lead to an understanding which challenges the existing order. “In society as it is, the power of thought has never controlled itself but has always functioned as a nonindependent moment in the work process, and the latter has its own orientation and tendency.”\textsuperscript{36} Theory is only tolerated to operate within narrow parameters; if, for example, it leads to an understanding which will A) contribute to military supremacy B) allow capitalists to make more money through more efficient control of natural and human resources C) legitimize the current arrangement, or some other function which serves given interests. This is evident today as a new vocationalism is hollowing out higher education and ravaging some of the last strongholds of autonomous theory in academia. It is unsurprising then that traditional theory would define rational action in such a way that it corresponds to action which safeguards and supports the
status quo. What a critical normative theory needs, if it is going to criticize society as it is, is to operate autonomously. Similar to Sen’s expanded focus on all valuable human functionings, as I hope to reveal below, what sets critical theory apart is that it “is not concerned only with goals already imposed by existing ways of life, but with men and all their potentiality.”

The idea of a just and free society realizable through rational planning is widely greeted with suspicion largely because critical theory “appears, to prevailing modes of thought, to be subjective and speculative, one-sided and useless.” The reason for this is that prevailing modes of thought are stuck in instrumental rationality. Once it is accepted that practical reasoning is about deciding means to pre-given subjective ends, then rationality become a technical and strategic endeavour for which negative freedom is fully adequate and in which discussions of virtues, values, or justice seem out of place. According to the critical theory perspective “the separation of value and research, knowledge and action, and other polarities, protects the savant from the tensions we have indicated and provides an assured framework for his activity.” The critical theorists and Amartys Sen are in agreement again, in that technical specialists cannot start out with any simple pre-given framework that will reduce rationality to a probability calculus and that practical rationality must be understood more broadly. In the absence of sensible alternatives, to fill out any such broader understanding will involve, I argue, breaking with “rational choice theories” and returning, in one way or another, to a conception of practical reason closer to both Kant and Aristotle.
2) The Radical Critiques

Quite apart from the critiques brought out in section one—those which challenge one’s understanding of what it means to chose rationally in the name of some alternate conception—stands a set of significantly more radical critiques which seek to challenge rationality tout court. Although his strategy was prefigured in earlier thinkers, such as Arthur Schopenhauer, the influential founder of this particular brand of critique can fairly be identified as Friedrich Nietzsche. I begin here by considering this Nietzschean alternative before moving on to what I consider one of its leading contemporary inheritors.

2.1) Nietzsche and the Other-of-Reason

More than twenty four centuries ago a pesky agitator—a gadfly, Plato called him—began making waves in Athens, upsetting orthodoxy and destabilizing long held ways of life in the community. The upheavals he instigated did not happen in the way which was at that time most usual; he held no office of influence, he was not leader of any political faction and he commanded no rebel force. In fact, he did not even write any treaties or manifestos. What Socrates did, above all else, was simply to go about asking people questions that challenged them to think for themselves about some of their own important beliefs. Quite often, in the course of the dialogue which ensued, what became clear to the interlocutors was that many commonplace beliefs could not be made sense of—could not stand up under close inspection. Famously, Socrates was put to death by the city he loved for causing people to scrutinize their own beliefs and consequently for undermining traditional ways of life. Fortunately, however, this did not happen before he was able to plant a seed
which would grow into the thriving Ancient Greek tradition of critical thought culminating in the likes of Aristotle. Socrates is a pivotal, indeed in the west a foundational, juncture in the continuing tradition of thought and practice which seeks to subject our ways of life to critical scrutiny—a broad tradition which I have set out to defend and in some modest way develop.

Nietzsche also recognizes this precise moment as a profound turning point, only for him, by sharp contrast to nearly the whole of Western philosophy, Socrates marks the beginning of a period of great decline in Western civilization. The centerpiece of Socrates supposed error is the "most bizarre" equation of reason, virtue and happiness. Instead, Nietzsche equates happiness with instinct. What's going on here is more than the straightforward opposition between passion and understanding. There is, for Nietzsche, no escaping instinct. The desire to escape instinct is merely another instinct. So, when Nietzsche equates happiness with instinct he is actually equating happiness with a particular sort of instinct—the Dionysian (the instinct associated with intense creativity, drunkenness, loss of identity, ecstasy, sexuality, becoming, originally unity and opposed to the Apollonian instinct, associated with discipline, clarity, rationality, sobriety, being and individuation). In analyzing Socrates problem, Nietzsche "seeks to comprehend what idiosyncrasy begot" this false equation. The answer, for Nietzsche, is the "wantonness and anarchy of his instinct"—an "instinct of slander, detraction and suspicion against life." It is this weak, life negating, ultimately nihilistic instinct which gives rise to the Socratic equation. That is why Socrates allegedly represents the beginning of the great turning of the tide away from a noble, healthy, Hellenic
spirit, towards dialectics and the "absurdly rational"—the weapon of the weak, sick and inferior.  

Socrates’ equation supposedly denies all impulses, and endeavours to give human life over to a new master—reasoned discourse. However, this endeavour is nothing more than a foolish day-dream on Nietzsche’s account. He argues that all morality based on reason, and all related judgments of value concerning human life, are illusions—groundless superstition. Reason has failed to show that there is a God or even a thing-in-itself that lies behind appearance. The belief in any such lie is well-suited only for the weak herd-animal, the ignorant, docile, and complacent masses. For Nietzsche, the value of life is inestimable. Life cannot be dissected—logically pulled apart and scrutinized such that we can analyze the wherefores and whys of its true value and meaning in order to devise maxims that will allow us to plot the morally correct course through life.

According to Nietzsche, like God, reason in general and moral reason in particular is a human invention. As social (Nietzsche would say “herd”) animals, human beings desire to live together; this creates the necessity for a peace pact. It is this desire to live in a “herd-fashion” requiring a “peace pact” which first drives us to establishing a set of fixed conventions and rules.\(^A\) This includes the invention of the idea of reason and the very idea of truth and knowledge which reason seeks to discover. One who uses designations in the recognized way is truthful, while one

\(^A\) Here, notice that Nietzsche may be contradicting claims he makes elsewhere to the effect that that states are formed, not through contracts or negotiations, but when conquered masses simply acquiesce in the face of what seems a terrifying nature-like force and are made subservience and shaped into a state by the violence of artistic masters with the power to command and the power to organise. See Chapter 17 in the second essay of *On The Genealogy of Morals.*
who abuses conventions is called a liar. The idea of truth as something that contains knowledge of “reality” or things as they are, something trans-cultural, trans-historical, or objective, only gains traction once we forget that it was an invented custom. Thus, truth is merely:

a sum of human relations, which have been enhanced, transposed, and embellished poetically and rhetorically, and which after long use seem firm, canonical, and obligatory to a people: truths are illusions about which one has forgotten what they are...to be truthful means using customary metaphors...to lie according to a fixed convention.

A further explanation of why it is that we come to believe in universal moral reasons can be found in Nietzsche’s account of a series of related core-mistakes, the so called “great errors,” on which moral philosophy and religion—the great detriments to ‘ascending’ life—are based. One such error, called the error of imaginary causes, is our psychological need to manufacture a motivation or reason to explain our feelings and actions. These imaginary causes are the product of fear and stand in the way of investigating the actual causes of things. Nietzsche argues that when we experience something entirely new, something strange or foreign, this is a very distressing and uncomfortable event for us. This compels us to look, not for the actual cause, but, a cause that will soothe our anxiety. In this way pleasure becomes the "criterion of truth." Here, Nietzsche is arguing for the inverse of the received view. On the received view, what we call motivation supposedly comes after what is first experienced. Not so for Nietzsche. On this account, we do something or we feel a certain way and we don’t know why. This clearly seems an unsettling way to go through life, so, we invent reasons/motivations to explain our states and behaviour, so that they make sense to us. All morality and religion
commit this error, which strikes Nietzsche as weak and pathetic. Instead, he claims, we should confront our fears by facing the unknown and unexpected—letting go of our imaginary explanation of moral reasons for actions.

It becomes evident here that Nietzsche does offer up an alternative to the Socratic equation and the supposedly life negating instinct from which it arose. What it comes down to for Nietzsche is this: pace Socrates and virtually all subsequent Western philosophy up to that point, the path to happiness was not the life of reason, in fact, there is no discoverable goal to life whatsoever, there is only life as it is, which one can either affirm or deny. The lies perpetuated by the (mis)use of reason, metaphysical lies about a noumenal realm or true world of being, or religious lies about a divine creator, divert us from the life that is in front of us. They lead us to distrust, question, deny and ultimately shrink away from life, leaving us complacent, unhealthy, degenerate (all too) human beings.

Nietzsche’s alternative is for us to strive towards a higher state of humanity by being like the tragic artist and developing or participating in a Dionysian spirit. The originally unity of the Dionysian experience allows us to A) experience the primordial becoming of the world and then B) allows us to say yes to it in its entirety. This alternative is a wholesale celebration of life and creativity. On this view there is a pre-conceptual, pre-linguistic, pre-discursive, organic or physiological realm of meaning or significance. On this level, without reason, without dialectics, we can be in contact with original unity. Choosing to say yes to this means embracing life in its diversity, it entails, as with the tragic art form,
celebrating every aspect of life including what is called terrible, it entails that "everything is deified regardless of whether it is [called] good or evil."\(^{48}\)

It is abundantly clear that in doing away with reason as a normative criterion Nietzsche does not want to commit himself to the position that every view or position is to be evaluated as equivalent in worth to every other view. Nietzsche certainly makes judgments, as in: "Every individual may be scrutinized to see whether he represents the ascending or descending line of life."\(^{49}\) Not even in a religious treatise is one likely to find prose more densely packed with normatively loaded descriptors; if there is anything made clear in Nietzsche’s writings it is that the views which are the subject of his scornful ridicule are stupid, miserable, and uninteresting, while his own ideas are brilliant, exciting, and interesting. His intention is not a "you are right from your side and I from mine" variant of relativism. It is not about right or wrong, it is about strong will. Only in this way can he attempt to maintain that accepting the brute contingency of our historical position and ultimately values, need not lead to nihilism.

As Herbert Schnädelbach puts it "whatever thinkers since Socrates tried to oppose to brute force—reasonable thinking, better argument, substantiated insight—was itself only a means to power as in the formulation ‘The will to truth is the will to power.’”\(^{50}\) All that remains then is will. Thus, Nietzsche’s normative standard can only be constructed from such material—from this other-of-reason. No reasoning can save us from nihilism only the affirmative will to say “yes” to life. The ideal becomes not goodness but what he calls “greatness”—to stand alone as an independent being “overrich in will” beyond the false categories of good or evil.\(^{51}\)
At this stage I want to flag some issues that will be considered in more detail below. Very broadly speaking, the main challenge facing Nietzsche and his disciples is two-fold. First is the issue of whether or not any such view can escape the straightforward charge of performative self-contradiction, which appears, \textit{prima-facie}, to be a knockdown argument that will prevent any totalized critique of reason from getting off the ground—for the employment of rational arguments (Which Nietzsche often offers) to refute reason is clearly self defeating. This sort of challenge is generally discussed in the contemporary context in connection with Habermas (which I will do in chapter five) but is at least as old as Aristotle who used it in his \textit{Metaphysics} to counter relativist conceptions of truth.\footnote{52} I will return to this issue again as I discuss more recent followers of Nietzsche below. Further, is the issue of whether the alternative vision is possible and desirable. This raises questions that are difficult for any Nietzschean to answer: is it possible to turn away from reason in the way prescribed? And, even if we could, why should we accept Nietzsche’s alternative standards of valuation? For example his valorization of pure creativity seems unable to account for the possibility that people can quite imaginatively create genuinely novel things that we find extremely bad. These questions will, likewise, be taken up in the following sub-section. Therein, I take up a discussion of one of Nietzsche’s most influential, clearest, and most persuasive contemporary defenders. The reason I hold off on a fuller account of some of the criticisms initiated above is because I aim to show how these criticisms can apply broadly, not just to Nietzsche but to a whole vein of thought following in that tradition.
2.2) Rorty: Philosophy as Advertisement

In both the negative or critical aspect and the positive or constructive aspect, Rorty's philosophy closely resembles Nietzsche's. On the negative side he adopts a totalized opposition to reason and therefore to Western philosophy up to the arrival of a heroic vanguard — including Nietzsche, Kierkegaard, Heidegger and other thinkers on whom Rorty approvingly bestows the title “ironist”—that began developing a new style of philosophy. On the positive side, ironists, in a sort of Nietzschean affirmation, accept and embrace the pure contingency of life as it is, devoid of truth, necessity, essences and ultimately of any foundation for our ends or values which transcend cultural and historical chance.

One problem, identified by Rorty, with any given non-ironist inquiry is that it will offer a set of conclusions or theoretical interpretations which are the result of some methodology. Methodologies themselves may be defensible by reference to some theoretical picture of the world; however, since this picture is itself a conclusion of an inquiry, this creates an inescapable circularity with the upshot that “nothing counts as justification unless by reference to what we already accept.”

For the ironist, in the final analysis, problems such as these will mean letting go of standard notions like truth, knowledge, reality and, of ultimate concern here, morality derivable from reason. Philosophy from Socrates onward is then seen as the misguided attempt to use such notions to escape from history.

Rorty follows through this line of thinking with a de-rationalized account of philosophical change, which, like everything else, is seen as the product of chance. Philosophy is irrelevant to how we deliberate and choose with respect to the
complex issues that arise in the public and political sphere. To think that one theory is abandoned because it is less well supported with reasoned argument, in favour of a competing theory that is supported by better reasoned arguments, is considered naive. Rorty’s radically historicist approach, by contrast, is to treat philosophical ideas as mere historical phenomena, with historical explanations, thereby revealing their contingency and dispelling their pretensions to inevitability. On this view, our values are not—cannot be—based on universal reasons, only on contingent factors such as our enculturation. This explains Rorty’s oppositional stance toward the supposedly desirable goal of overcoming ethnocentrism. For Rorty, we should not only tolerate but embrace ethnocentrism—to do otherwise would be to have no values at all (in Nietzschean terms, it would be nihilistic and life negating), for there is no purely rational position, transcending our particular form of life, from which we can derive any universal, categorical, rationally binding norms or values.

For Rorty, the already alluded to positive alternative to rationalist philosophy (broadly speaking) is the ironist approach. The ironist is someone who harbours doubts and uncertainty about their own ends and the particular language game in which these ends can be expressed, viz. their “final vocabulary,” and further, someone who does not hold these particularities to be either 1) derived from any transcendent or transcendental source or 2) truer, closer approximations of some reality than any rival alternative, and relatedly, an ironist is one who cannot help being impressed by the ends and final vocabularies of others. Nevertheless, ironically, despite all this, the ironist is someone who remains deeply and sincerely committed to his or her own ends and final vocabulary. In short, the ironist is the
person who “faces up to contingency” but manages to escape nihilism by willing oneself to living playfully at home in a world of pure contingency and chance.

Given its total opposition to the methodological instrument of philosophy—reason and an associated constellation of concepts (truth, reality, morality, etc.)—some thinkers allied with an ironist type approach are happy not to be referred to, or refer to themselves, as philosophers. Once reason goes, so with it, one would expect the whole of philosophy to follow. However, Rorty does not see the need for a terminological break of this kind because of his thoroughgoing anti-essentialism. Rorty does not think there is some key characteristic—reasoned reflection or whatever else—that is the essential feature of philosophy. We cannot ask, sensibly, “what is the true nature of philosophy?” Philosophy cannot be defined other than descriptively, by looking at what gets put in philosophy sections in bookshops and libraries, while philosophers would be defined by looking at who calls themselves philosophers and what people do in places called philosophy departments. More importantly this tactic makes sense strategically for Rorty because he does not want to simply leave the philosophers alone and go off to do some other kind of literature. We will be much better off, he maintains, if people stop doing that thing that has been called philosophy. What he wants is to change what we call philosophy into something else. The critical question is: how does he attempt to do this?

As an intelligent person, steeped in the philosophical canon and trained in the analytic methods of the discipline, Rorty is well adept at offering what old-fashion philosophers would recognize as rational arguments designed to establish the truth of some position. And, much like Nietzsche, this is often what he does.
Here is where the spectre of self-contradiction rears its head; for traditional philosophical techniques cannot be undermined through the force of those selfsame traditional techniques. There is one way out here. It is a way which Rorty explicitly takes and, although he may have been less self-conscious or explicitly clear on the issue, one might be able to interpret Nietzsche as taking the same route. In any case, this strategy, as I will show, proves equally fatal for this sort of approach.

As an ironist, Rorty sincerely holds to a set of ends, one of which is of course to replace old-style philosophy with something new, but he has no commitment to any particular method for achieving those ends—least of all the rational methods of the very tradition he seeks to supplant. Instead of supporting any methodology as “correct” Rorty takes an ad hoc approach, employing any technique that will have the causal effect he desires. This involves the use of varied strategies which include, for example, narrative, historical recontextualization, and yes, when he thinks it will work, rational argument. But when he makes use of argument in this way we cannot simply say “Ah, look, you do believe in rational arguments and everything which that entails?” because he can reply “Well, no, I just gave you an argument because I want you to accept my view, and an argument is what I thought you wanted to hear in this case; if I thought telling a story would work better I might have tried that.” As Alan Malachowski puts it “Rorty’s claims may have the appearance of substantive claims but they are intended to serve a different function.” While, from the perspective of the philosopher this may look like a misuse or abuse of argument, for Rorty, who has no reverence or commitment to any particular means, this is not seen as a problem.
Here, one must, at least, be sensitive enough appreciate how difficult it is to articulate a view this radical. Rorty wants to destabilize the whole practice of philosophy as a truth-seeking, reason-giving enterprise, not simply argue for some new position within that activity. As Rorty says, the “trouble with arguments against the use of a familiar and time-honoured vocabulary is that they are expected to be phrased in that vocabulary.”\(^{58}\) He is certainly willing to do precisely this when necessary, but his real aim is to get past the point where this is necessary. As Malachowski puts it in his sympathetic portrait of Rorty “if a particular argument looks like it will do the trick effectively... Rorty is happy to employ it. But his overall strategy is designed to circumvent any requirement for the kind of sustained, detailed lines of argument that characterize previous philosophical ventures.”\(^{59}\)

While he may have strategic reasons to co-opt traditional dialectic techniques, his goal is to change philosophy from the outside on what Malachowski calls an “extra-philosophical, performative level.”\(^{60}\) This is why Rorty’s outlook, despite some appearances to the contrary, is not simply another philosophy with a rationally supported theory about how things really are. Ultimately he is asking us to “set aside” the problems and methods of traditional philosophy rather than “argue against” them.\(^{61}\)

The traditional claims of philosophy including essentialist claims “are not things Rorty wants to ‘counter’: he simply wants us to be rid of them, to live our lives without giving them much thought.”\(^{62}\) Instead, Rorty wants us to think, to put it in his own words, of “human beings as generators of new descriptions rather than beings we hope to be able to describe accurately.”\(^{63}\) Rorty never means to describe
his alternative as better on account of its being more accurate, his goal is a view that is better according to different (non-rational) criteria—variously described as new, fresh, creative, interesting, less tradition bound and more useful, fruitful, or productive—which owe much to Nietzsche.

As an initial response I cannot but wonder how Rorty would assess his general strategy if I applied it in a different instance. Supposing I put it to Rorty "Your paper, book or whatever it may be, is totally useless nonsense—in short, just plain garbage." He might respond "that's an interesting tack John, how do you know that my paper is total garbage?" To which I could reply "Well, I do not want to get caught up in your question and ruminate over it, I want to get out from under your question, not by responding to what you present but by moving on to something else, that is, moving into a fresh and more interesting position where we come to view your paper as total garbage. Instead of wasting time responding to your detailed reasoning I simply want you to forget about them and for you to accept my exciting, new, and liberating invitation to think in a way completely different from the way that you present in your paper."

As a thought experiment, try imagining someone responding to your own arguments in this way. How would you react? I suspect it is hard for any of us to imagine a reaction that does not involve anger, incredulity, indignation or a sense of injustice having been done to us. Ironist thinkers in the Rorty camp sometimes express bemusement at the particularly heated negative response with which it is often greeted by other philosophers. It should not be surprising, however, that a response which is totally different in kind from other responses (reasoned
responses) should receive a reaction that is also different in kind. When someone responds to our reasons with counter-reasons we may not agree with them, we may find their response misguided and abhorrent, but we recognize that there is a medium—rational dialogue—recognized even by our opponent, through which we can offer a response and in which we can work to reach an accord. This is not a unique feature of philosophers or even occidental ways of life; virtually any adult considered sane at any time in any culture will greet the abject refusal to respond to reasons in kind, on a rational level, with some degree of frustration and hostility. Rorty, however, does not respond in this way since he, as Malachowski so forthrightly puts it, “does not wish to argue in detail against assumptions...he finds unfruitful.”

Even when he does give arguments he does so in a way which is, by admission, utterly insincere. We know that he has no intention of being open to the possibility of being moved through reason to accept any of our claims, that he will say anything to make his position seem more attractive, and that he does not even believe his position on account of the “reasons” he offers; for these “reasons” are merely enticements tailored to our sensibilities. Rorty believes that in the absence of objective standards “anything could be made to look good or bad, important or unimportant useful or useless, by being redescribed.” Thus, for him, philosophy is about dressing up the particular, contingent convictions he happens to have so as to

---

A Despite what he claims, I should be clear that I do not think it is possible, even for Rorty, not to be persuaded by reasons. Although not absolutely so, everyone believes some things because of reasoned arguments. Someone could make the belief that the words you are reading now are printed in indigo squid ink supremely attractive by agreeing to give you a billion dollars or grant your greatest wish if you were to believe it. Unfortunately for you in this scenario, belief, on some level, does not work this way; no external enticements are likely to counteract the fact that you have good reason to think these words are not printed in squid ink.
make them seem attractive. If some other method, say producing televisions commercials featuring Rorty, or a younger actor playing an idealized version of him, driving around in flashy cars surrounded by scantily clad women, proved a more effective method of attracting adherents than writing books and articles then presumable he would have to consider some such method.

So, somewhat ironically (though not in the way Rorty intends) this outlook ends up being quite similar in important respects to Hume and rational choice theory and diametrically opposed to the challenges posed to it, especially those posed by Sen and the critical theorists. For Rorty, as with the mainstream economic theories, our ends are rationally inscrutable contingencies about us. Also in line with the traditional economic theories, Rorty’s account sees human action—including philosophy itself—reduced to the instrumental and strategic. As discussed earlier in this chapter, the great problem with modern culture, as diagnosed by the critical theorist, was the irrationality of the culture industry which sought to bypass our rational capacities, playing off certain base instincts and certain irrational fears and insecurities in order to manufacture consumer desires and to titillate certain senses that will elicit diversionary and narcotizing effects that prevent understanding of both the nature of advanced capitalism and the objective interest of the masses in changing it. Rorty by stark contrast wholeheartedly embraces the tactics of “public relations” and “branding,” offering a new form of philosophy wherein the best way to establish a philosophical position is to make it look good. This form of “philosophy,” which openly surrenders rational decision
making to non-rational forms of coercion, is tailor made to fit the *zeitgeist* of a post-rationalist age of manipulative advertising persuasion.

To sum up so far, Rorty has two options. On one hand, instead of strategically mimicking a position, he could offer an actual substantive philosophical position. If we take him to be rationally defending a substantive philosophical position then it is riddled with contradiction. For example Rorty “argues” that many of the phenomena that *appear* to be necessary are *actually* historical and cultural contingencies. This, of course, presupposes things like a knowable appearance-reality distinction, which is one of the central edifices of traditional philosophy that Rorty want us to do without. Rorty is able to escape these problems to the extent that he opts to be thoroughly consistent with his fundamental anti-rationalism and drops the idea that he even has a “position” in anything remotely resembling what it has meant to hold a position in philosophy. However, he can only do this to the extent that he is able to totally and definitively exit all rationally mediated interaction operating with a notion of some degree of objectivity (what Habermas calls commutative action) in favour of openly strategic action in support of a frank ethnocentrism—and this does not seem like a possibility since reason is an unavoidable part of a conceptual scheme that is bound up with all human theoretical and practical activity (as we saw in Kant and will return to in connection with Habermas in chapter five).

Thus, there is good reason to maintain that Rorty’s self contradictions are a problem for him either way. That is to say, even if he avoids the contradictions being a problem in and of themselves, by retreating to strategic action, the very fact
that the contradictions continually and unavoidably crop up is itself a problem. When Rorty criticizes others his go-to strategy is to make appeal to seemingly objective standards. For example, in his critical review of Richard Wolin’s book *The Seduction of Unreason* he chastises Wolin for accepting the intellectual and cultural assumptions of modernity without doing what he ought to have done, which was provide us with reasons why we should accept them; rationally ungrounded ethnocentrism apparently being alright for Rorty but not his rivals. In general, he also chastises Wolin for talking too much about philosopher’s actions and not paying proper attention to “the arguments themselves”\(^66\)—again precisely the sort of thing we would expect form a traditional philosopher who believes in some form of reason that transcends cultural norms. In fact, Rorty can scarcely write more than a few pages of text without contradicting the outlook he explicitly urges us to adopt. This is a very telling fact. As Thomas McCarthy puts it, Rorty:

> has given us no persuasive account of how it would be to speak a language free from any notions of an independent reality about which we can make claims that are not just true here and now but *tout court*. He has not even given us an illustration of this, for all his attempts to do so are invariably rife with performative contradictions—and that is especially significant in a debate about unavoidable presuppositions.\(^67\)

He is gesturing here towards the suggestion, essentially a correct one, that there are such unavoidable presupposition, and that Rorty’s, untimely incoherent, urgings that we abandon what McCarthy calls “context transcending truth about an independent reality and any other ideas of universal validity implicated in received notions of reason and rationality”\(^68\) does more to demonstrate the truth of this fact than it does the contrary. Rorty uses the power of reason but denies the objectivity
without which it would be powerless to convince. This is problematic, as Aristotle put it “those who seek merely compulsion in argument seek what is imposable.”

These issues will resurface in chapter 5. For now I want to move on to a final issue arising from Rorty’s radical critique.

Once we exit rationally mediated interaction in favour of persuasion as a manifestation of will, what is there to prevent moving from narrative redescription to violent repression of dissenting voices? The answer, disturbingly, is “nothing necessarily.” Of course Rorty would never support such tactics because he is an ironist of the liberal variety, which for him means abhorrence of all cruelty. However, Rorty is not a liberal for any reason. Rorty just happens to be a liberal. On his account, the fact that I share many of his political convictions would not be a result of the fact we have been persuaded by similar sorts of reason but because he and I just happen to share a final vocabulary and a similar cultural background.

However, Ironists who do not just happen to abhor cruelty can be particularly scary things. One need not look far for an example.

For Nietzsche himself, violence is actually preferable to reason. Reason is for weak-willed “buffoons” like Socrates. “One chooses dialectic only when one has no other means” it is a “self defence for those who no longer have other weapons.” The strong give commands not reasons. Unsurprisingly then, Nietzsche openly despised liberalism, democracy and above all socialism. For him “the theory of ‘equal rights’ is an essential feature of decline. The cleavage between man and man, status and status...is characteristic of every strong age.” Thus, the proper form of society is a pyramid wherein the work of the masses of weak herd animals provides
the basis for the few great men. This is necessary, for the great and beautiful "can never be common property." One must always embrace the means to their ends; thus, if one wants the great and the beautiful and this requires slaves, one must embrace slavery.

In any case, whether an ironist’s concrete moral and political convictions seems pleasing or disturbing is more or less separate from the fact I want to stress here, namely, that abandoning the idea that human beings, as rational creatures, ought, categorically, to have some measure of freedom to exercise that reason, that capacity to legislate for themselves, entails that any means of getting other people to do what you want, be it coercion, overt violence, or whatever else, is not off the table. That said, although Rorty’s ironist standpoint is more palatable than Nietzsche’s it has its own disturbing elements, specifically its unflinching ethnocentrism which has faced critique from a wide range of perspectives, including from within a Nietzschean or ironist camp. For example, Farid Abdel Nour’s response to Rorty’s brand of liberalism forwards some strong lines of argumentation against his endorsement of ethnocentrism, however, given that Nour begins from the same philosophical starting point as Rorty, it is impossible for him to make these arguments gain real traction—at least in the way he supposes. Both Nour’s anti-ethnocentrism and Rorty’s anti-anti-ethnocentrism are consistent with their shared deeper philosophical commitments. Without appeal to philosophical resources alien to either thinker there is, in fact, no possibility for genuine philosophical disagreement between them—merely a difference of personal preference or disposition. Finally, I believe it will be prove illuminating to consider
this opposition in a little more detail as a case study in what happens when two irrationalists pretend to have a rational discussion. This will provide a glimpse of what discourse would digress to if we follow the irrationalist path.

Nour and Rorty are in full agreement that it is impossible to philosophically justify moral standards. Further, they both happened to subscribe to the values of liberalism, despite their agreement that these values cannot be universally justified through the obsolete enlightenment idea of reason or through any other objective standard. Lastly, they both think that without philosophical justification for their liberal values there is no viable alternative but to turn to the depths of who we are for moral guidance. The disagreement between them only crops up when it comes to the question of interpreting who we are.

For Rorty, "we" are a society with a commitment to certain enlightenment moral ideas, such as, the supreme Liberal values of human freedom and equality. Once we recognize that these normative standards are not universal or necessary but merely particular group customs, there can be no more questioning whether we are making universal rational judgments or whether we are being culturally biased—the epistemological fact of human finitude means there can be no neutral standard operating above cultural bias. According to Rorty, feeling guilty or agonizing about imposing our standards on those in a non-liberal context is based on the assumption that we could do otherwise—that there were some neutral way of facing the other. Once we recognize this is imposable we can stop feeling guilty and get on with the business of promoting liberal values, in other words the business of being who we are—which, after all, is the only thing anyone can do.
The core of Nour’s disagreement can be found in his charge that Rorty has “limited us to a hypostatized caricature of who we are.” Specifically, Nour argues that Rorty has simply failed to recognize that turning to the depths of who we are does not mean that we must succumb to the comforts of celebrating the familiar, owing to the fact that a deep commitment to other-regard is an essential part of who we really are. According to Nour, liberal self disgust felt over imposing alien standards on the other is an unavoidable consequence of this commitment to other-regard, and so, any attempt to erase or alleviate it simply fails to adequately recognize the full character of who we are.

Now, since the disagreement between these two thinkers hinges on the question of who we are, that is, what our society happens to believe or what customs it happens to endorse, the question seems best suited for a descriptive and empirical social-scientific explanation. If we were to answer the question through a more anthropological as opposed to philosophical analysis I think the answer would clearly emerge that “we” (Western culture) are not so uniform or easy to characterize as either thinker supposes—each offers only a favoured snapshot or “hypostatized caricature” (to use Nour’s own phrase) of who we are. As a matter of empirical fact we will find that many people will indeed follow Rorty and embrace ethnocentrism without guilt while other people, or perhaps even the same people in a different mood, will find that their commitment to other-regard will not allow for their guilt to be allayed.

So, while Nour is entirely correct to point out that Rorty has completely failed to appreciate the fact that a Liberal commitment to other-regard and the guilt that
goes with it, can “transcend philosophical justification”76 (that is, it can be seen as simply part of who we are just as much as any of the features about us that Rorty wishes to emphasize) he is equally guilty of pushing his favoured snapshot depiction of who we are. The reason why neither Nour nor Rorty are ready to admit to the facts that 1) the particulars of who we are is something much more partial, fragmentary, and complex than they maintain, and 2) the debate between them is primarily an empirical issue that should be left to anthropologists or sociologists, is not only because they reject notions underlying science, but largely because neither of them really means to talk about a contingent ethnographic fact of “who we are” but rather “who we should be,” and that is precisely what they can never talk about giving their commitment to the idea that normative standards admit of no properly philosophical justification.

When Rorty tries to defend his Liberal ethnocentrism saying it presents “the best hope for the species” or when Nour attacks it saying that not to engage the irreducible alienness of the other is “a very real wrong” they are not giving anthropological descriptions of norms or practices but are attempting to advance (crypto-)ethical arguments for why It would actually be better, for us and the for the other, to be or to behave in a particular way. If they wish to continue with this project they must completely reevaluate their deeper philosophical commitments. If, on the other hand they wish to drop this project and stay true to those commitments they have nothing left to discuss. Their philosophical “disagreement” appears as some form of debate mimicry, akin to the absurd spectacle of a tennis match with no ball. They make normative judgments and appear to be engaged in a
debate, but the appearance that they have philosophically engaging with one
another is pretense; for just as the ball-less tennis players have no medium in which
to engage one another, their positions and arguments do not make contact with one
another since they have both abandoned commitment to a medium of rational
interaction. Rationalism, once abandoned levees us to shout assertions at each
other: “I think we should be like this” or “I think we should be like that” even if we
dress up these assertions in fancy academic language that looks like argument.

2.3) Conclusion

“Immanence” has been used to describe a mode of critique grounded in an
examination of the tensions and possibilities within a given structure. The problem
here is that any internalist or historicist normative stance that relies solely on
immanent critique can never be radical, never escape the status quo, because there
is no gap between this reality and the norms or standards being applied. This
makes the sort of approaches we have been discussing in this section seem
undesirable relative to the alternative, that is, if an alternative is available.

“Transcendence” by contrast means having norms and values that can transcend the
given by appealing to culturally unrecognized norms which require change. The
danger here, as the radical critiques meticulously remind us, is an externalist stance
that attempts to critique from on high. Its apparent disconnection from the given
social reality seems to require a gods-eye-perspective or view-from-nowhere,
outside, independent of, even above the current historical situation. The issue here
is the uncertain foundation of such a transcendent values system.
Thinkers in the critical theory tradition, as well as others, as I will continue to argue in the next section, aim at a dialectic of immanence and transcendence, rejecting extremes grounded solely in one or the other and avoiding both problems by developing a theory with a foothold in the social world and the power to point beyond it. Easier said than done? Certainly so. The rearticulations of moral rationalism investigated in the remainder of this work are largely concerned with if and how this is possible. It seems to me that one way or another, at the broadest level, many of these approaches explore the promising possibility that a solution may lie in developing a moral theory with a powerful set of normative standards that would allow for a truly critical stance towards social reality without endorsing a determinant ethical system or full-blown conception of the good. However we address the ongoing challenge to make sense of morality and moral reasoning, the point I wanted to stress here is that abandoning reason is simply not a viable option. We are now in the midst, as Thomas McCarthy puts it, of “another round of critically rethinking our ideas of reason, truth, and objectivity—to which we still have no sensible alternatives.”

3) A Neo-Aristotelian Approach

In characterizing what I have called the contemporary dilemma of moral theory I hope to have made clear some of the reasons why we should reject both the dominant economic approach, which, like Hume, starts from desire on top of which is added a calculative instrumental reason, and the, ironically, similar approach of the totalized critiques of reason which seek normative guidance in some other-of-reason—pure desire or will. This, I argue necessitates a return to the sort of moral
rationalism we saw operating in Kant or Aristotle. This is not an easy task however. While the positive alternative of the radical critiques seem particularly flimsy, these approaches are at their best when on the offensive, and indeed some of their critical concerns must be added to the challenges facing the moral rationalism of Aristotle and Kant already considered in chapters one and two. Now, finally, we are in a position to consider how it might be possible to articulate a Kantian or Aristotelian inspired version of moral rationalism sensitive to the contemporary situation. Fortunately, a great deal of fruitful effort has been directed towards just this task by leading thinkers in normative theory today. I turn now to an evaluation of these accounts. In this section I will briefly consider a leading Neo-Aristotelian approach, while the next two chapters will be devoted to reconstructions of essentially Neo-Kantian strategies.

3.1) Contemporary Aristotelianism

To begin with, not all Neo-Aristotelianism constitutes a return to moral rationalism of any kind. In fact, a great deal of contemporary interest in Aristotle comes from sectors which borrow certain Aristotelian notions, such as virtue, but ultimately have more in common with Nietzsche and Rorty than they do with Aristotle himself. A case in point is Alasdair MacIntyre. Although he is concerned with virtues specified by rational inquiry, the “reason” appealed to in this context bears little resemblance to what Aristotle or other moral rationalists have in mind by that term. In stark contrast with previous moral rationalism which contrasted the authority of reason with the authority of tradition, MacIntyre develops a “conception of rational enquiry as embodied in a tradition.” For MacIntyre,
underlying the diverse and fundamentally incompatible range of judgments that exist on fundamental moral issues, such as justice, are incompatible traditions. “Reason” as something outside or above tradition cannot be appealed to as a methodology for adjudicating between these judgments. While Aristotle may have succeeded in showing that certain aspects of reason—i.e. rules of logic arising from the law of non-contradiction—are universal, any view of practical reason must go well beyond logic and draw on the settled convictions of some community.\textsuperscript{79} Thus, because reason is inescapably bound up with the perspective of a particular tradition, disputes about the nature of reason are just as intractable as the disputes over justice themselves.\textsuperscript{80}

Ultimately, on this account, the reason modern Western culture is so unsettled and seemingly irrational is that it draws from a mixed-up hodgepodge of incompatible traditions—a situation in which there can be no rationality. Rationality, on Macintyre’s account, only makes sense within some coherent tradition. Thus, in some important sense, the solution to irrationality must come down to more settled tradition. Here, reason is not about transcending the parochial in any meaningful sense; it is more like following the pre-constructed rules of a particular game. This ends up being more consistent with a Nietzsche or Rorty type approach. Certainly this is not the kind of reason that Aristotle required to make the sort of radical criticisms he brought against Athens, or the kind required to sustain the utopian challenge to advanced capitalism forwarded by critical theory. However, not all Neo-Aristotelianism takes this form. One powerful attempt to salvage the transcendent critical potential of Aristotle exists in the work
of Martha Nussbaum who develops an explicitly Aristotelian version of the capability approach—the groundbreaking innovation of Amartya Sen. Instead of a relative list of virtues which merely reflect local customs Nussbaum’s aim is a list of virtues which is objective in the sense of being justifiable by reference to reasons which have foundations in universal human features.\textsuperscript{81}

\section*{3.2) Nussbaum’s Aristotelian Moral Rationalism}

To understand Nussbaum’s approach it is best to begin by considering some pertinent details of Sen’s work. Sen’s capability ethics has now been taken up and developed in different directions by several other thinkers, (notably David Crocker and Martha Nussbaum) but their versions all hinge on the basic idea that moral progress should best be thought of as a process of enhancing certain valuable human functionings and expanding basic human capabilities. The Key question is then: what exactly does Sen have in mind by functionings and capabilities? Sen’s common objection to most previous theories stems from the fact that each theory, by focusing on one important space—whether income, primary goods, utilities, social participation, liberties, or some other set of freedoms or achievements—a theory neglects, in consequential ways, other important aspects of human well-being. For example, "Person 1 can have more utility than 2 and 3, while 2 has more income than 1 and 3, and 3 is free to do many things that 1 and 2 cannot."\textsuperscript{82} In other words, equality in one space is quite consistent with inequality in another space that matters. Sen’s purpose is to give a much richer account of equality by choosing a focal variable which takes into account the diversity of human beings and the wide
range of doings and beings which constitutes true human well-being and flourishing. The space Sen proposes is that of functionings and capabilities.

Functionings, for Sen, are the constituent elements of a person’s being, and are as such precisely what we need to look at if we are to assess accurately a person’s well-being. Simply put, functionings are any human beings and doings. The evaluative space of happiness for example is only one valuable human 'being.' However, Sen argues, one can be happy and still be seriously deprived in terms of many basic beings and doings. The main feature of a person’s well-being is their achievement in terms of their combined actual functionings, or their 'functioning vector' in Sen’s terminology.

A functioning vector is a bundle of actual well-being achievements. Capabilities, the possible functionings from which individuals choose, are valuable not only instrumentally, because they are necessary for actual well-being achievements (functionings), but they are also of intrinsic value, because they represent choice, and choosing is an important part of living. Capabilities then, are about freedom, that is, the real opportunities that people have to choose between different functioning combinations. The upshot is that Sen’s approach differs from most previous theories of social justice first, in that there are actual states that should count, besides just welfare, or goods, and second, in that it is not only actual states that should count but also the range of potential states that represent live choices for an individual.83

---

A Beings and doings simply refer to any quality or state of being that a person achieves.
Functionings and capabilities, however, are not separate focal variables. Functionings and capabilities are the same space. A particular functioning combination of an individual (a functioning vector) is a point in the functionings’ evaluative space, while a capabilities set, is a set or range of such points in the same space. Sen makes this clear with an analogy:

Just as the so-called 'budget set' in the commodity space represents a person's freedom to buy commodity bundles, the 'capability set' in the functioning space reflects the person's freedom to choose from possible livings.84

Development for Sen is about working for equality, and promoting well-being in the space of functionings, and the capacity to choose between possible functioning those which one has reason to value (i.e. capabilities).

At this point the obvious response is to reply that “this is all well and good to make functionings/capabilities the locus of value but that really has not said anything substantial yet.” This is similar to what I said before about Aristotle’s own golden mean, until we know how the mean is to be selected we do not have a normative theory that can offer any real guidance. As I showed in chapter one, Aristotle had an answer—*theoria*—for Sen, however, the question of real difficulty remains: *How do we know which beings and doings are the ones of value?* If we cannot simply justify why certain capabilities/functionings are good by appeal to some more fundamental standard such as welfare, then it seem we just have to accept them as good because someone says that they are. How can we explain why a capacity for friendship is a valuable human functioning but the capacity for slave mastership is not? To take another example, both Sen and Nussbaum assert that the opportunity for sexual satisfaction is a valuable human functioning. The problem is, there does not seem to
be anything here to stop someone from saying that being tortured or starved are valuable human functionings. I might agree that friendship and sexual satisfaction are valuable and slavery and torture are bad but that appears, on this account, to be just a fact about what I happen to prefer given my culture, my upbringing, and so on. If we do not have actual reasons or criteria for saying a thing is good, than anyone, it seems, can assert that anything is good, without the possibility of that being criticisable for reasons.

Sen has had a lot of difficulty addressing these questions. Sen himself has certainly recognized that there is an outstanding question regarding how to determine what valuable functioning’s are and that there are a lot of difficulties with existing attempts to answer this question. Sen’s response is simply that there “are no magic solutions to offer in dealing with these complex questions.” But, one would not be asking for sorcery to demand an explanation of the rational foundation for a theory that purportedly aims to rest on something other than a list of unsupported assertions. His standard reply to these sorts of criticisms is to say that it is better to be vaguely right than precisely wrong. Unfortunately however, Sen’s vagueness on this particular point, does little more than obscure the fact that he has a massive blind spot in his account which cannot be shrugged off easily. A

A rationalist foundation for choosing and ranking a list of functionings is articulated in Nussbaum’s version. Although he calls Nussbaum a “longstanding

---

A In a more recent book, The Idea of Justice, (see this chapter’s introduction) Sen offers an elaboration of his position which provides some insight into his reluctance to specify criteria. In short, Sen is broadly critical of the whole tradition which thinks it necessary to begin by identifying or specifying an ideal of “perfect justice” favouring what he calls a “comparative approach” – an alternative not analyzed here.
influence on my work” Sen had not recognized or developed its Aristotelian foundation himself saying:

The connection of this [capability] approach with Aristotelian ideas was pointed out to me by Martha Nussbaum, who has gone on to make pioneering contributions to this growing field of investigation and has strongly influenced the way the approach has developed.87

Nussbaum points to a number of close parallels in Sen and Aristotle. For both thinkers, the goal of politics, the political good, is to distribute the conditions necessary to live a good life, that is, to realize the human good. Sen specifically criticizes Rawls (the subject of the next chapter) for a sort of commodity fetishism—acting as though our goal were to evenly distribute some set of things with a value inherent in themselves. Sen claims, as Aristotle did, that goods are only valuable as tools, useful in the attainment of the capacity to function well. One cannot say what is or is not a tool before knowing the job. Therefore, one cannot know what a good is prior to taking some stand (even if a minimal one) on what a good life consists in. Again in accord with Aristotle, Sen is adamant that subjective desire is not a good criterion for determining virtue or valuable functioning since many people, through deprivation, miseducation and so on, have faulty or inadequate desires.88

Where Sen falters, according to Nussbaum, is on account of the fact that “the valuational procedure that is involved in capability selection seems to me, at least without further description, to be no more incorruptible than desire itself is.”89 The worry here is that shifting the focus from desire to capabilities achieves little if desire is the sole criteria for determining the value of a capability. Thus, Nussbaum maintains that:
Sen needs to be more radical by introducing an objective normative account of human functioning and by describing a procedure of objective evaluation by which functioning’s can be assessed for their contribution to the good human life.\textsuperscript{90} This is what she sets out to do.

Like the critical theorists, Nussbaum is self-conscious about the need to present a rational procedure that is both immanent, “non-detached” as she puts it, and transcendent, or as she says “objective.”\textsuperscript{91} On the one hand, our list of valuable functionings cannot be outside the conditions and experience of life, as though it could be some sort of value-free scientific fact about us. At the same time however we cannot simply accept, unquestioningly, every group or cultures account of good functioning, for to do so would be to slip back into the problematic desire and preference camp with Hume, the Utilitarians, rational choice theory and the non-rational approaches.\textsuperscript{92} This she argues can be achieved through a form of Aristotelian objectivism and Aristotelian particularism.

According to Nussbaum, Aristotle’s account of the virtues proceeds in stages. He starts by identifying a more or less fixed reference point, what she calls a “grounding experience.”\textsuperscript{93} This is a sphere of life in which all human beings necessarily have some dealings. Next he asks what it would be to chose or respond well in that sphere. Virtue, in a thin sense, can then be defined simply as the disposition to act appropriately in a given sphere.\textsuperscript{94} Likewise, the virtuous action can be called good or appropriate functioning within a particular sphere of life.\textsuperscript{95} Finally, the job of ethics is to fill in the thin account by specifying a thicker conception of what it is to function well in each domain.\textsuperscript{96}
These spheres are connected to universal features of the human experience; for example; our mortality and associated feelings like fear or aversion to death; the fact that we have a body, which entails that we experience the world through a particular set of senses and that we have certain bodily desires like hunger and thirst; the fact of our sociability or need for human affiliation; the experiences of humour and play; and various others. Nussbaum focuses considerable attention on the objective character of the grounding experiences which define the domain of the various virtues. While there is wide disagreement on the thick definitions of the virtues there should be near universal agreement on the thin definition. For example, with respect to truth telling, one might call the virtue “pure honesty,” always telling the truth regardless of circumstance, another might call the virtue “prudent honestly,” always telling the truth unless the greater good is served otherwise. Although there is a dispute here over the particulars, there is, at least, agreement that we are all confronted with the possibility of being honest or lying in various situations and that there is some proper way to conduct oneself with respect to this experience which we could call the virtue relative to this experience regardless of how we characterize it and the name we happen to give it. In short, we argue about the specification of virtues or good functionings, not their existence.

At the same time Nussbaum fully acknowledges, as Aristotle himself seems not to have done, that the human mind is an inescapable active interpreter shaped by its historical contingencies and that these grounding experiences are not, therefore, wholly uninterrupted givens providing an unmediated epistemic access to human nature which can lead in a clear and straightforward way to the
specification of virtue concepts. However, the absence of an uninterpreted reality to measure our moral judgments against need not spell the dissolution of moral reasoning. Although the grounding experiences are interpreted in quite different ways, their existence is universally shared, and this provides a “starting-point for cross cultural reflection” what Nussbaum calls a “nuclei of experience.”98 The specification of the virtues then becomes the, presumably ongoing, task of cross-cultural dialogue in which certain specifications will reveal themselves to be more in keeping with our “evidence” and “wishes” than alternative ones.99 In this way, Nussbaum supposes, a “foundation for ethics can remain inside human history and self-interpretation, and yet still claim to be a foundation.”100

There is a significant unanswered question here regarding how the virtue specifying discourse is to proceed, or, to put a finer point on it, how are we to decide what counts as a specification of a virtue? Nussbaum has an answer; the aim of our discourse on a candidate for a virtue, a supposedly good function, is to answer the question as to “whether that function is so important that a creature who lacked it would not be judged to be properly human at all”101 or, as she puts it simply, we ask “what is an essential part of any life?”102 Nussbaum also supplies an answer. Unsurprisingly it shares a great deal in common with Aristotle. For her, as it was shown to be with Aristotle in chapter one, there are many important components of the good life, but these “are distinctively human functionings only when they are done as parts of a life organized by practical reason and infused with reasons activity.”103 So, Nussbaum follows Aristotle first in asking after the defining essence of humanity and second in answering that question with “reason.” As she says,
“reason is what all the functionings have in common.” And it is on the basis of this understanding that she proceeds to fill in her more detailed list of valuable human functionings.

Nussbaum ultimately draws up a supposedly general and comprehensive list that is meant to tell us what is valuable for each and every human being. The basic idea, as I have said, is that universally valuable human functionings are defined as those beings and doings which contribute to a rich and fully-human life, which above all is a life organized by reason. The theory claims that there are certain features of special importance, without which a life could not be recognized as human. What Nussbaum is doing here is stipulating a specific normative definition of what it means to be a good human. But why should we believe humans should essentially be defined the way she says they should? Instead of blindly asserting what valuable functionings are, Nussbaum says that valuable functionings are what make us human and then goes on, essentially, to assert what it is to be human. The issue is complicated, but the underlying problem is not resolved.

Granted, the grounding experiences provide a fairly universal touchstone for debate but they certainly do not seem to take us very far beyond defining more precisely what it is that is being disagreed about. It might help us arrive at the idea that a human being that is functioning well generally needs to be able to sleep and breathe air and eat food and so on. This should not be trivialized, but it will hardly take us very far when it come to resolving most actual socially and politically relevant normative disputes over the specification of valuable capabilities. For example, we might all happily agree that corresponding to a particular set of bodily
and intellectual desires there is a universal sphere of human life called sexuality. But, this agreement seems to have little or no significance in the debate between, say, a secular liberal who calls the virtue a “healthy and active sex-life,” and a strict Catholic who regards any sexuality or sex-act not directly intended for procreation, as a temptation to be avoided and punished, and therefore defines the relevant virtue as “chastity.” This controversy would then be relevant to a political question such as: should contraceptive practices be taught in publicly funded schools?

To summaries, I think Nussbaum has followed Aristotle and Sen in correctly identifying the focal variable of ethics as functionings/capabilities. Nussbaum’s idea is that to figure out which functionings or capabilities are valuable, we need to ask, what are the essential features of a human life? However, even given a set of universal grounding experiences this question proves no less controversial or easier to answer. How then are we to answer? Nussbaum gets us part way there. For the most part she seem to avoid any appeal to a settled metaphysical biology which can settle ethical debates once and for all, recognizing that the ongoing job of ethical theory is to specify valuable human functioning through ongoing rational debate. In this way a conception of the virtues is meant to emerge from a reflection on the grounding experiences. The problem is the objective grounding experiences alone are not enough to get us to a set of universally binding norms. What gets us from the grounding experience to the specification of a rationally binding norm is the rational, reflective procedures of moral theory themselves. In addition to the objectivity of the grounding experiences, these procedures have objective elements, not fully analyzed by Nussbaum, which must form part of the normative core of any
moral theory with critical potential. I now turn my attention to certain broadly Neo-Kantian approaches which may, I believe, shed some light on the idea of rational procedures of this sort.
Notes: Chapter Three

5Sen, 29 & n64.
8Wilson & Hölldobler 13.
15Sen, *RF*, 22.
16Sen, *RF*, 22.
22Sen, *RF*, 46.
26Sen, *RF*, 51.
34Horkheimer, *TCT*, 200.
35Horkheimer, *TCT*, 203.
36Horkheimer, *TCT*, 212.
37Horkheimer, *TCT*, 245.
38Horkheimer, *TCT*, 218.
41Nietzsche, *TI*, 475.
42Nietzsche, *TI*, 475.
43Nietzsche, *TI*, 484.
46Nietzsche, OTL, 47.
47Nietzsche, TI, 497.
49Nietzsche, TI, 534.
52Aristotle, Metaphysics, 1011a 20-25.
56Rorty, CIS, xv.
58Rorty, CIS, 8.
59Malachowski, 43.
60Malachowski, 19-20.
61Rorty, PMN, 6.
62Malachowski, 19.
63Rorty, PMN, 378.
64Malachowski, 16-17.
65Rorty, CIS, 7.
68McCarthy, 97.
69Aristotle, Metaphysics, 1011a 15-16.
70Nietzsche, TI, 476.
71Nietzsche, TI, 476.
72Nietzsche, TI, 541.
73Nietzsche, TI, 511.
74Nietzsche, TI, 545.
76Nour, 213.
77McCarthy, 95.
79Maclntyre, 4.
80Maclntyre, 1.
84Sen, IR, 40.
85Amartya Sen, Commodities and Capabilities (Amsterdam: North-Holland, 1985), 32.
87Sen, If, n231.
Chapter Four

John Rawls and the Reinvigoration of Moral Rationalism
Introduction

Thinking—and of most particular concern here, that thinking which is directed towards social and political concerns—within the tradition of moral philosophy, as I have characterized it, began its descent into a phase of considerable decline in the period following Hegel (1770-1831) and his immediate predecessors. This descent continued to the point where Isaiah Berlin could make the bold pronouncement in his 1962 essay “Does Political Theory Still Exist?” that “no commanding work of political philosophy has appeared in the twentieth century.”

With the notable exception of continuing developments within the Marxian tradition it seems fair to say that a period of inactivity—relative to the pace of enlightenment era innovation—probably stretches back even beyond the turn of the century. However the “outstanding difference now...is that Berlin’s assertion is no longer true. It ceased to be so in 1971, when A Theory of Justice by John Rawls of Harvard was published in Cambridge Mass.”—to borrow an equally bold, and I think no less accurate statement by James Fishkin and Peter Laslett, with which Chandran Kukathus and Philip Pettit open their study of Rawls’ Theory of Justice. Before considering how Rawls changed the scene of moral philosophy it is worth going over some of the main causes for the state of the scene in which Rawls found it, in order to see what it was that Rawls was up against, as it were. While acknowledging that these causes are likely many and each of these multifaceted, here I shall simply enumerate three significant general factors.

---

^ In short, as rationalist in the broad sense.
1) The first factor is an issue emphasised by Kukathus and Pettit which concerns the division of labour amongst academic disciplines. Traditionally, they argue, political theory was concerned with what was politically desirable as well as what was politically feasible. This seems a sensible approach given that, if we are to find answers to fundamental questions about what is to be done, we need to work out feasible plans that take into account questions concerning both that which is possible—what we can hope to achieve—as well as questions concerning what is good or right—that is, what should we hope to achieve?

In an attempt to raise themselves to the status of real "science," economics and the social and political sciences have attempted to purge themselves of properly normative concerns and limited themselves to analysis of empirical facts. Feasibility, however, is only interesting in relation to desirability. To dramatize the point, the sciences could tell you if it is possible to extend health care benefits to a given national minority group or if it would be possible to round them up into camps and gas them, but it must remain wholly silent with respect to the question of which, if either, should be pursued.

Philosophy, on the other hand, has tended to settle into its own niche within a priori reasoning, supposing that it could work out a social and political theory without concerning itself with issues of feasibility; the upshot being that the 20th century was left with no single field that covered the traditional domain of political theory.3 It would be surprising, then, if the fracturing, if not dissolution, of political philosophy as a discipline did not have a deep impact on the development of actual political philosophies.
2) The second factor is a problem already discussed in some detail (chapter 3), namely, growing doubts about the very possibility of the normative component of political philosophy. If some version of moral scepticism is true, then normative political philosophy is not possible, and the fact that a considerable number of people within the academic discipline of philosophy, continental and analytic alike, harboured such reservations about the possibility of moral theory, meant that there was a dwindling stock of theorists for whom the task of political philosophy was a live possibility.

3) Finally, even if political philosophy is possible and seen to function as a unified area of study, there is a good explanation for its stalled development which is wholly internal to the content of political philosophy itself, and it is a problem which Rawls himself has emphasised to a significant degree; that is, the stalemate, within political philosophy, between utilitarianism and what Rawls calls rational intuitionism. This was not an external problem—a structural problem about how the discipline was organized, or a philosophical challenge about whether the discipline was possible—it was simply that the discipline faced a considerable intellectual hurdle in terms of its own content. In short, the field itself had reached a vexing impasse.

In contemporary theoretical physics there is a stalemate between physical descriptions of the world in terms of Relativity and Quantum Mechanics. Both seem compelling, so compelling that we cannot seem to do without either of them; yet they are fundamentally incompatible and cannot both be true. Currently, all physicists can do is appeal to these theories in an *ad hoc* way—using Relativity to
describe and predict very large things and events, and turn to Quantum Theory when dealing with very small things—knowing, or suspecting, all the while, that both are false and that a better physical description of the world is given by some, as yet unknown, theory (the so called “Unified Field Theory”). Theoretical physicists since Einstein have been trying to come up with such a theory. Their failure to do so is not primarily for lack of a branch of research devoted to the task or because anyone thought theoretical physics impossible, but largely because of the simple fact that the task is a difficult one. The problem Rawls faced, as I believe he saw it, was a remarkably similar one.

According to Rawls, utilitarianism had been the leading systematic account of morality for some time, having no real systematic alternatives. Utilitarianism’s fundamental moral principle and its basic underlying idea that right actions are ones that would maximise total good consequences understood in terms of human welfare is compelling, and it is hard to imagine doing without it entirely. Nevertheless it fails to accord with our considered moral judgments—throwing up what look like problems in many situations.

The alternative, according to Rawls, is rational intuitionism. Intuitionism in moral theory has meant either A) an epistemological theory about the nature of moral judgments, viz. that they are perceived directly or immediately by anyone with the necessary power of perception, or B) a moral theory with multiple fundamental moral rules with binding normative force independent of the consequences of following those rules. Rawls may intend both meanings; however,
in the present context of distinguishing rational intuitionism from utilitarianism the focus seems to be on the latter.

Rational intuitionism, on Rawls’ account, has not coalesced into a systematic alternative but remained a sort of deontological hodgepodge—a grab bag of values and principles appealed to in different circumstances. While rational intuitionism does not offer a satisfying alternative to utilitarianism, many intuitionist values and principles seem like ideas we cannot do without. Thus, a great deal of moral thinking defaults to “a variant of the principle of utility circumscribed and restricted by seemingly ad hoc intuitionistic constraints.”

In moral and political theory, as in physics, we have different and, at times, conflicting theories that seem to work (that is, accord with our considered convictions) in different circumstances, yet no one theory to make sense of it all. Rawls sets himself the task of creating, out of this messy landscape, a unified moral theory that fits together and prioritizes all our best or most-considered moral convictions into a single, neat, and coherent system that we can look to for guidance on particular moral/political issues.

Rawls set out to overcome the roadblocks facing normative political theory by 1) finding a way around moral scepticism, 2) articulating a new moral theory that could overcome the limitations of rational intuitionism and utilitarianism and 3) engaging in a theoretical exploration of the desirability and feasibility of putting his theory into practice. Regardless of how one assesses the degree to which Rawls may or may not have succeeded in these ambitions, his work marks a decisive turning point for political theory. In one sense at least it is a re-turning point, marking a
return to a bold constructive attempt to offer a thoroughly worked out and genuinely innovative normative theory to govern our political relations—namely, his theory of justice as fairness. Since Rawls, there has been an explosive proliferation of fruitful new discussions—a genuine reinvigoration of normative political theory.

One further fact about Rawls’ reinvigoration that I want to call attention to, a fact which is the main focus of this chapter, is that Rawls not only revived political philosophy by considering feasibility in conjunction with normative questions, or that he offered a new and systematic answer to those question that seemed to offer advantages over leading available theories, but that he also contributed to a reinvigoration of a particular strategy for answering those questions. Unsurprisingly this is the traditional philosophical strategy of basing normative principles and judgments on practical reason. Rawls thinks “any workable political conception...must endorse rationality as a basic principle of political and social organization.” Of course, Rawls thinks he has done so in a new way that avoids the criticisms facing previous versions of this strategy. At the same time, his theory can be seen as an attempt to reformulate a version of a broadly Kantian form of constructivism.

The aim of this chapter is to move in on Rawls’ conception of practical reason in his later work by asking: what precisely is practical reason for Rawls? how does it function within his theory?, how, more specifically, does it avoid the perceived problems with previous such attempts? and, ultimately, is Rawls successful in developing a theory of justice based on practical reason? I Argue A) that Rawls saw
himself as a kind of moral rationalist, B) that his conception of practical reason and its function as a basis for his moral/political principles is original and compelling (to a point) and, as such, marks a decisive moment in the development of moral rationalism, but C) that this strategy exhibits certain defects that are deeply problematic for his overall theory. I will argue that the substantive normative content which Rawls attempts to derive from practical reason (specifically that part of practical reason he calls the reasonable) is, in the final analysis, an uncritical appeal to status quo norms and values which would forestall the possibility of rational progress and undermines the moral rationalist project which Rawls set out to develop.

1) Rawls’ Moral Rationalism: A Thumbnail Sketch

Rawls connects with the broad philosophical tradition which I have been calling moral rationalism in that he sets out to defend the existence of an objective order of moral values, principles, and duties, not founded on any external authority—a deity, sacred text, sovereign, canon of law etc.—but, rather, arising in some way from human nature, and accessible to every person with normal human powers of practical reason. He further connects with this tradition in that no external (divine or sovereign) sanction is strictly necessary to compel obedience. Practical reason itself is seen as capable of motivating action in accord with moral duty.

A key qualification, the full significance of which will become clear later, is that Rawls sees certain limits on the sort of values derivable from practical reason. In short, Rawls attempts a derivation of political values only. These values are
represented by his proposed political conception of justice. The first thing to be clear on is what this limit entails. The political is not defined, as in certain popular usage, in contradistinction to moral—as, for example, that which can be achieved through power struggles and negotiations rather than that which ought to be done. Rawls is clear that “political rights and duties are moral rights and duties”\(^\text{11}\) and “the domain of the political and a political conception of justice are normative and moral ideas.”\(^\text{12}\)

What then does this limit consist in? The relevant distinction is between a political value, which concerns the right, and what Rawls calls an “ethical value,” which concerns the good.\(^\text{13}\) The good is a comprehensive idea that includes everything that is valuable or worthwhile in life. The right is a more restricted idea that concerns what we owe to each other as citizens. The right would be one component of various particular wider conceptions of the good. Thus, ethics, with its fuller conception of the good, is normative and presumably encompasses the widest part of morality, whereas the political comprises a narrower subset of moral considerations which exclude ethical value. When Rawls speaks of the moral-political or justice he is concerned with that which we owe to other citizens and it is this part of moral value, that is, political value, which is derivable from our publicly shared practical reason alone.

The basic reason for this limitation is that conceptions of the good require justifications which reference some complete, or at least wide ranging, religious or philosophical understanding of the world—what Rawls calls a “comprehensive doctrine”\(^\text{14}\)—which a citizen cannot reasonably expect to justify to all other
reasonable citizens in a pluralist society, while, on the other hand, the political value of justice needs no metaphysical foundation, since the principles and conceptions of practical reason alone are sufficient for their justification.\textsuperscript{15} Rawls makes this intent clear when he says “a political conception... sees the public principles of justice as founded on the principles and conceptions of practical reason”\textsuperscript{16} or “the correct model of practical reason as a whole will give the correct principles of justice.”\textsuperscript{17} This derivation yields a moral-political conception of justice which is “freestanding” in the sense that it contains its own moral ideal and does not need to be derived from any particular comprehensive doctrine.\textsuperscript{18} However, this conception is also modular in the sense that, although the right is derivable from a public procedure that does not need to make reference to any particular comprehensive view, every reasonable comprehensive view will be able to include the right as a component or “module”\textsuperscript{19} of its comprehensive view of the good life. Particular comprehensive doctrines may, and very often will, have other, and from their perspective deeper, religious or philosophical reasons to support the political conception and what it upholds as right under their own conception of the good.

Rawls’ political conception of justice as a whole has two parts: A) substantive principles of justice, and B) “guidelines of inquiry: principles of reasoning and rules of evidence in light of which citizens are to decide whether substantive principles properly apply and to identify laws and policies that best satisfy them.”\textsuperscript{20} As we just saw, Rawls claims the principles of justice (A) derive from the principles and concepts of practical reason. However, Rawls intends for the complete conception of justice (A & B) to rest on practical reason seeing as the “guidelines of inquiry of
public reason as well as its principles of legitimacy [B] have the same basis as the substantive principles of justice [A].”\(^{21}\)

If one were to apply a label to Rawls’ particular view of practical reason—insofar as it forms a basis for moral-political value—and discuss it under a single heading, this would be “public reason”—where “public reason” is employed in a rather technical sense. For Rawls, as I hope to make clear, public reason stands for something quite specific and indeed complex. By Rawls’ own account, the main distinguishing feature of his liberal view—his political conception of justice—is his specific account of the practical reason which grounds his theory; that is, in essence, public reason. As Rawls puts it, “not all liberal views would accept the idea of public reason as I have expressed it. Those that would...we may call political liberalism.”\(^{22}\)

It is unsurprising, then, that Onora O’Neill claims that the “most fundamental parts of Political Liberalism are the discussions of public reason, and the most central question that can be raised about the work is about the adequacy of the proposed conception of public reason.”\(^{23}\)

The ultimate aim of this chapter, then, is to work towards answering the question “how is Rawls’ political conception, including both substantive principles and rules for their application, meant to be derived from public reason?” and to consider whether, or to what degree, this derivation is successful. I’ll begin with a brief, preliminary statement of Rawls’ strategy and proceed by trying to flesh this out.

In the tradition of Kant, Rawls see himself as a constructivist. Rawls’ principles of justice are the outcome of a procedural device of representation. This
procedure is designed to model practical reason. According to Rawls, if the procedure successfully “embodies all the relevant requirements of practical reason” then it can be employed to construct fair principles of justice.

The distinction between what Rawls calls “the reasonable” and “the rational” is a fundamental one for Rawls and he will place a great deal of emphasis on it. For Rawls, “principles of practical reason” include both “reasonable principles” and “rational principles.” Since Rawls wants to model practical reason and since both the reasonable and the rational fall under the rubric of practical reason, the aim of his constructivist procedure must be to express “all the relevant criteria of reasonableness and rationality that apply to principles and standards of political justice.” Specifically, since the reasonable is prior to the rational for Rawls, the procedure will model rational individuals subject to the constraints of the reasonable. When reasoning is carried out subject to the constraints of the reasonable this is called public reason in Rawls’ specific sense. Thus, in modeling rational individuals subject to reasonable constraints, the procedure is expressing public reason. Similarly, when it comes time for real citizens to apply the substantive principles of justice they must do so by exercising public reason themselves.

According to this basic strategy, Rawls attempts to do what philosophers have been trying to do since at least the ancient Greeks; that is, employ reason to answer questions about how we ought to live. Although he thinks we must limit ourselves to what he has called “fundamental political questions”—“those that involve constitutional essentials and matters of basic justice”—he maintains
public reason is up to the task of objectively grounding general moral principles and particular moral judgments and that this political conception can form the basis of a stable constitutional democracy. In what follows I consider this strategy in detail, beginning with a direct consideration of Rawls’ conception of practical reason.

2) Practical Reason: An Overview

According to Rawls: “Practical reason has two aspects: principles of practical reason and judgment, on the one hand, and persons, natural or corporate, whose conduct is informed by those principles, on the other.”29 In order to understand Rawls’ conception of practical reason one must consider its two aspects.

1) First I consider principles of practical reason. These would certainly incorporate principles connected to what Rawls calls “general capacities of reasoning”30 such as logic, inference, and judgment, for “all ways of reasoning...must acknowledge certain common elements: the concept of judgment, principles of inference, and rules of evidence, otherwise they would not be ways of reasoning but perhaps rhetoric or means of persuasion.”31 But, like Kant, Rawls understands practical reason as primarily concerned with the production of things in accordance with an idea of those things, as opposed to understanding things as they are, which is the domain of theoretical reason.32 Thus, we can presume practical reason will include principles which go beyond the mere general capacities of reasoning.

At this stage of the argument Rawls has little to say directly regarding what these are. This seems surprising, at least until one grasps the significance of the fact that, on Rawls’ account, and in ostensible contradistinction to Kant, “there is no such thing as the point of view of practical reason as such.”33 Rawls argues that, logic,
inference, and judgment are expressed only in the thinking, inferring, and judging of rational/reasonable beings. Instead of focusing on a pure form of the principles of practical reason Rawls thinks we must consider them in connection with their actual context, a context which is specified by what he calls ideas or concepts of practical reason. After considering these concepts of practical reason I will ultimately be in a position to proceed further with an examination of practical reason as a whole.

As far as the justification of the principles of practical reason is concerned, Rawls is fairly clear. Rawls points to Kant as the historical origin of the idea that reason is “self-originating” and “self-authenticating”—a view which he endorses when he claims “the principles of practical reason originate, if we insist on saying they originate anywhere, in our moral consciousness as informed by practical reason.” As a feature of our consciousness, our rational nature, they are simply basic or fundamental, and it would not make sense to look elsewhere for their justification. Rawls recapitulates this Kantian point when he says “we cannot ground these principles and canons on something outside reason. Its concepts of judgment and inference, and the rest, are irreducible. With these concepts explanation comes to an end.”

2) Now I turn to the ideas or concepts of practical reason, which is to say, a conception that specifies who or what it is that is supposed to carry out the reasoning. Rawls bases his moral-political theory on practical reason only insofar as we understand practical reason to encompass its two aspects—not simply principles of practical reason on their own. Thus:
the principles and ideals of the political conception of justice are based on principles of practical reason in union with conceptions of society and person, themselves conceptions of practical reason.\textsuperscript{37}

Ideas about society and persons are properly considered ideas of practical reason because they are necessary to characterize the agents who engage in reasoning processes. The conceptions give substance to practical reason by specifying the framework in which principles are applied. As Rawls puts it, “without conceptions of society and person, the principles of practical reason would have no point, use, or application.”\textsuperscript{38} Principles of practical reason, then, must be considered in the context of agents who reason.

Given this understanding, Rawls thinks we must begin by asking what individuals and societies must be like if they are to engage in practical reason.\textsuperscript{39} The answer we give to this question will represent fundamental ideas—that is, ideas that are basic for an understanding of practical reason. This means that the conceptions of society and person are equi-primordial with the principles of practical reason themselves. Like the principles, the conceptions of practical reason have no deeper justification. Exactly where these ideas come from is an issue I return to in the last section of this chapter. At this stage, however, I shall ask: what precisely are these fundamental ideas?

For Rawls, the ideas of practical reason can be thought of as connected to his single “fundamental organizing idea”\textsuperscript{40} of society. Specifically this is the idea of “society as a fair system of social cooperation between free and equal persons.”\textsuperscript{41} This conception of society is not meant to be more fundamental than the other fundamental ideas; it is not epistemologically or metaphysically prior to them in any
sense. This conception of society can be called the single most basic idea simply in the sense that, conceptually, it is the broadest idea “within which the other basic ideas are systematically connected.” 42 In a sense, then, the other ideas can be thought of as components of this idea of society. This overarching concept is not, however, a simple idea. It consists of various separate and irreducible component ideas which are not derivable from any one master idea.

Rawls characterizes the starting point of his normative project by stating “We start with the fundamental idea of society as a fair system of cooperation between reasonable and rational citizens regarded as free and equal.” 43 A Within this statement of the idea of society we see various other ideas: the ideas of persons as citizens of some form of political association, persons as free, persons as equal, persons as reasonable, and persons as rational. Appealing to this specific understanding of society is only meaningful after these component ideas are fleshed out. Given these facts, in appealing to the idea of society, Rawls acknowledges that he is not appealing to one idea but what he has referred to as a “family of fundamental ideas.” 44

The idea of society as a fair system of cooperation, at least at the broadest level of generality, has two “companion fundamental ideas.” 45 One is the more specific characterization of how a society can actually become a fair system of cooperation—what Rawls calls a “well-ordered society.” 46 This would be a fair and stable society as regulated by a political conception of justice. 47 To understand this idea requires a fuller characterization of Rawls’ political theory of justice as a whole;

---

^Henceforth, by Rawls idea of “society,” I mean precisely this.
thus, I leave it aside at this stage. The second, and, I think, more basic idea is that of the persons who make up society. This, for Rawls, is the fundamental idea of “citizens” which his version of liberalism, unsurprisingly, characterizes as “free and equal persons.”

Rawls is clear that his fundamental idea of the person is intended to be a political and moral conception, not a natural or social science concept. Being a person is not defined, for example, by genetic makeup; a person in this moral-political sense is understood to specify anyone capable of being a citizen. A citizen is “someone who...can play a role in social life, and hence exercise and respect its various rights and duties.” As such, citizens make up the “basic units of thought, deliberation and responsibility.”

It should now be asked if there is anything that underwrites the freedom, equality, and citizenship of individuals. Is there anything in virtue of which someone is a citizen? Is there anything in virtue of which someone is free? Is there anything in virtue of which someone is equal to other citizens? Certainly, for Rawls, there is. This was already alluded to above in the idea that, as morally responsible agents, citizens necessarily think and deliberate, and the claim that citizens who are seen as free and equal are both reasonable and rational. The moral-political status of free and equal citizenship is underwritten by the capacity of practical reason. According to Rawls, citizens are seen as free and equal in virtue of three capacities. One such capacity is, quite generally, the “powers of reason” and powers of “judgment thought and inference connected with these powers.” Presumably this covers the general intellectual skills associated with both practical and theoretical
reason. The remaining are specific capacities relevant to practical reason which Rawls calls our “two moral powers.” For Rawls, persons are regarded as free in virtue of possessing these powers to a sufficient degree, just as “having these powers to the requisite minimum degree to be fully cooperating members of society makes persons equal.”

The “two moral powers,” as I hope to make plain in a moment, are not something required in addition to reason but simply constituent powers that make up parts of practical reason understood as a capacity. In the broadest sense, Rawls claims practical reason can be defined as either an activity or a capacity. As an activity he conceives practical reason as the way an individual or society orders its ends and makes decisions. Any group’s ability to do this is rooted in reason as a capacity of individuals. He defines this capacity as an “intellectual and moral power” of persons. Given this understanding of practical reason, we can understand the two moral powers as aspects of the capacity of practical reason. The two respective powers relate to the two distinct aspects of practical reason mentioned above—the reasonable and the rational. All these concepts are members of the family of basic ideas connected under the idea of society. These ideas and their relationship to one another are the topic of the next section.

3) Moral Powers, the Reasonable, and the Rational

For Rawls, in order to fill the role of citizen, an individual must possess certain “intellectual and moral powers” including the two moral powers mentioned, but not yet named or explicated, above. These moral powers are 1) a “capacity for a conception of the good” and 2) a “capacity for a sense of justice.”
Briefly, a capacity for a conception of the good is one’s “capacity to form, to revise, and rationally to pursue a conception of one’s rational advantage or good.” In addition, Rawls assumes that at any given time, although it is revisable, an individual will have some “determinant conception of the good” or “scheme of final ends” as an outcome of the normal exercise of this capacity.

A “capacity for a sense of political justice” is defined as “the capacity to understand, to apply, and to act from the public conception of justice.” The core idea behind connecting with a public conception in this way is to act in accord with the right—what we owe to others as free and equal citizens. Essentially, what we owe other citizens, on this account, is their freedom not to be compelled by any external force they could not reasonably be expected to endorse themselves. Having a sense of justice “expresses a willingness...to act in relation to others on terms that they also can publicly endorse.” A fuller account of what each of these powers entails and consists in requires an explanation of the corresponding fundamental ideas of the reasonable and the rational. I will now consider the reasonable and the rational in connection with their distinctive moral powers.

The capacity for a conception of the good is directly connected with the aspect of practical reason called the rational. The rational applies to agents “with the powers of judgment and deliberation in seeking ends and interests peculiarly its own.” The rational covers two things. First, because it applies to the choice of one’s means to their perceived good, it includes standard instrumental means-ends reasoning. But Rawls is explicit that “rational agents are not limited to means-ends reasoning.” Thus, because the rational also applies to how one’s “end and
interests are adopted and affirmed, as well as how they are given priority, it is connected with our actual final ends or good.

On Rawls’ understanding, without the rational, individuals would have no final ends to pursue. This is why Rawls can use “the good” synonymously with the term “rational advantage.” Because the rational is connected with a complete or overall account of what is valuable for an individual, it is directly connected with what can be called ethics and is not limited to the moral-political concern with justice. Given this understanding of the rational I believe its connection to the first moral power, the capacity for a conception of the good, begins to come into focus. This power, as a capacity for a conception of one’s rational advantage could simply be described as the capacity to exercise the rational—it is the rational as a capacity. In the same way, the second moral power, the capacity for a sense of justice, connects with the reasonable. The reasonable is a bit more complicated, but, in general, it is expressive of a “basic desire to be able to justify our actions to others on grounds they could not reasonably reject.” Even a fully rational agent, on this view, could lack a sense of justice or morality and fail to recognize the independent validity of the claims of others. Reasonable persons, by contrast, are moved by a desire for social relations in which they, as free and equal, can co-operate on publicly acceptable terms. They are willing to propose fair standards and abide by them given some assurance that others will do likewise. As such the reasonable is connected with a moral disposition to act justly. At this stage it becomes evident

\(^A\) For Rawls, the idea of one’s own rational advantage is not an egotistic idea. One’s rational advantage will generally include goods that are in benefit of other people. A conception of the good approaches being psychopathic to the extent that it only includes goods in benefit of oneself.
that the capacity for a sense of justice can, in effect, be understood as the capacity to be reasonable.

To understand this capacity to have a sense of justice, and Rawls’ view of justice itself, it is necessary to dig a bit deeper into the idea of the reasonable. More specifically then, the reasonable, consists of two elements: 1) a “willingness to propose and honour fair terms of cooperation,” and 2) a “willingness to recognise the burdens of judgment and to accept their consequences.” The first element should be relatively clear given the general account of the reasonable. By terms that are “fair” Rawls essentially means terms that can reasonably be expected to be endorsed, not simply from the point of view of one’s own comprehensive ethical doctrine, but by all reasonable citizens. The second element of the reasonable speaks to what it means for something to be publicly endorsed or endorsable in this way.

“Burdens of judgment” is a technical term which refers to the causes of reasonable disagreement. Even amongst people who possess a “common human reason”—which is to say, comparable powers of thought, the ability to draw inferences, weigh evidence, balance conflicting considerations etc.—and the two moral powers, there will still exist moral disagreements due to the fact that it is difficult to make sound judgments on such matters. For example, people could disagree over the evaluation of certain empirical facts that bear on a particular moral issue, or on the weight to give certain facts. Often there are different kinds of normative considerations on different sides of an issue, making comparisons
difficult. Agreement is also hampered by the fact that the way one assesses an issue is shaped, to some degree, by their own particular life experience.\(^7^4\)

Recognizing the burdens of judgment means accepting that there are intractable reasons why fully rational individuals can have deep disagreements about the nature of the good—\textit{viz.} what is of meaning and value in life. Accepting the consequences of the burdens of judgment means accepting, first of all, what Rawls calls “the fact of reasonable pluralism”—the fact that moral disagreement is an inevitable result of practical reason operating under conditions of freedom.\(^7^5\)

Further, accepting the consequences of the burdens of judgments means accepting that the fact of reasonable pluralism limits the range of what can reasonably be justified to reasonable others. Since a reasonable individual seeks fair terms which other reasonable individuals can accept, and since other reasonable individuals have incompatible (from the other’s perspective untrue but not unreasonable) comprehensive doctrines, a reasonable person cannot make their own comprehensive doctrine—what they see as the whole truth—the basis for moral-political principles and judgments. The basic idea, as Rawls puts it, is that:

\begin{quote}
in recognizing others’ comprehensive views as reasonable, citizens also recognize that, in the absence of a public basis of establishing the truth of their beliefs, to insist on their comprehensive view must be seen by others as their insisting on their own beliefs. If we do so insist, others in self-defence can oppose us as using upon them unreasonable force.\(^7^6\)
\end{quote}

Given that the fact of reasonable pluralism is an outcome of free or unconstrained reason, a political order based on a particular comprehensive doctrine can only be maintained by violently or coercively placing severe restrictions on individual freedoms. He calls this the “fact of oppression.”\(^7^7\) Thus,
Rawls claims measures such as the Spanish Inquisition were no accident; if a social and political order based on Christianity was to be maintained, heretics could not be tolerated. For a democratic regime that supports individual freedom, however, such methods are clearly inappropriate. Thus, it is inappropriate, given Rawls' stated fundamental idea of society, to expect any comprehensive doctrine to be accepted by all members of a political society as a prerequisite for justifying and adopting moral-political principles that apply to that society.

In recognizing limits on what can reasonably be justified to other citizens one endorses a form of toleration—toleration of other comprehensive moral views—and a publicity requirement for moral-political reasoning. Thus, it is the aspect of practical reason called the reasonable, and our corresponding capacity to be reasonable or to have a sense of justice which makes practical reason public. I will come back to this idea in the next section but first, I want to briefly recap before considering one outstanding concern.

Rawls' political conception of justice—a moral theory restricted to fundamental political questions—is derived from the principles and ideas of practical reason. The ideas of practical reason are intended to characterise those who reason by specifying what they must be like in order to fill the role of a practical reasoning citizen. The organizing idea of practical reason is the idea of society as a fair system of cooperation. The individuals who make up said society are free and equal in virtue of having the capacity of practical reason, which, besides the general powers of reason, includes the capacity for a conception of the good (the rational as a capacity) and a sense of justice (the reasonable as a capacity). Thus, for
a person to be in accord with practical reason (and to be fully autonomous in Rawls’
political sense as I will address below) means being reasonable and rational. The
fundamental idea of society connects to the fundamental idea of the person, whose
status as a free and equal citizen is explained in terms of certain moral and
intellectual powers which constitute their ability to use practical reason.

Thus far I have considered Rawls’ account of the idea of the person as a free
and equal citizen in virtue of their capacity for reason. Therefore, in Rawls’
organizing fundamental idea of society as a fair system of cooperation between
reasonable and rational citizens regarded as free and equal, what is meant by
“reasonable and rational citizens regarded as free and equal” should now be
comprehensible. What is meant by “fair system of cooperation,” however, remains
in need of some further clarification.

The idea of “cooperation,” which is also a constituent of Rawls’ basic idea of
society, has two main elements: 1) the idea of each participant’s “rational advantage
or good” which they hope to pursue through cooperation and 2) the idea of “fair
terms” to govern that system of cooperation. These elements derive from aspects
of practical reason already discussed—the reasonable and the rational. Conceptions
of the good or rational advantage derive from the part of practical reason called the
rational. Without some final end there would be nothing to achieve via cooperation
and therefore no meaning to “cooperation.” The idea of fair terms for the system of
cooperation derives from the part of practical reason called the reasonable. Without
the reasonable the social system would be unfair and undeserving of the title
“cooperation.” In this way the reasonable and the rational work together to specify what it means for society to be a fair system of cooperation.80

For Rawls, the reasonable and the rational are themselves independent basic ideas and elements within the basic idea of society.81 In addition, since the reasonable and the rational connect with their own distinctive moral powers, each moral power is also associated with an element of the idea of cooperation.82 The capacity for a conception of the good is related to one’s rational advantage or good, while the capacity for a sense of justice relates to fair or reasonable terms for cooperation.83

Having now rounded out a preliminary account of Rawls’ view of practical reason, as I see it, with a basic sketch of Rawls’ fundamental ideas of practical reasons—the idea of society and its various component and companion notions—I now turn my attention to the question: how exactly does Rawls derive a substantive theory of justice from practical reason?

4) Constructing a Theory of Justice

Rawls’ theory of justice qua social contract theory has, to Rawls’ own frustration, frequently been misunderstood. Although Rawls has reinvigorated the idea, at the time he first articulated his theory, the contractarian approach had been widely abandoned by most contemporary political theorists—and not without good reason. The traditional social contract theories of Hobbes, Rousseau, Locke and Kant sought to explain the legitimacy of state power by appealing to the idea that individuals were bound to the state by a contract of some sort, either between subjects and ruler or between citizens themselves. Hence, it is this agreement which
holds a political society together and gives the state’s laws a force which can be considered morally binding in some way.

Fruitful though this tradition may have been in many respects, it is ultimately difficult to see how any such theory could get off the ground on account of one fairly simple but powerful criticism which stems from the fact that there is no social contract. The idea of a political society founded on an actual contract is historically inaccurate; people do not enter a society by signing a contract, thereby exiting some fabled state of nature. If on the other hand the contract is merely hypothetical it is morally insignificant and therefore no contract at all. It does not make sense to say anyone is bound to a non-actual contract to which they did not agree. A hypothetical contract, then, cannot be relied on to do this kind of legitimising work.

Rawls’ theory is a social contract theory in a strikingly different sort of way, and is not subject to many of the criticisms (notably the aforementioned) faced by previous incarnations of the contractarian tradition. Rawls is not offering an argument which legitimises state power or authority by appealing to the idea that citizens are party to any contract. In point of fact, Rawls’ social contract theory has much more in common with Kant’s moral theory, particularly the categorical imperative, than it does with the political theory of any of the traditional contractarians. That is because the Rawlsian social contract plays an “evaluative” rather than a “legitimising” role (to borrow terms from Kukathus and Pettit). Rawls’ social contract is not real but imaginary. As such it is not a contract that real people are party to or bound by, rather, like the categorical imperative, it is a device
we can employ in order to test something’s desirability/acceptability, in Rawls’ case that of basic social arrangements.

For Rawls, if we want to know what principles should govern our basic social structure we should ask what principles would be chosen by certain specifically defined fictional contractors in a fictional situation defined by particular fictional features. This is called the “original position.” But what justifies the appeal to this fiction? As I argued already, Rawls thinks justice derives from practical reason. What justifies the principles chosen in a fictional social contract set in a fictional situation is the fact that the situation is carefully designed to model all relevant features of practical reason. The original position itself is “simply a device of representation.” As Michael J. Sandel puts it “underlying the device...is a moral argument that can be presented independent of the thought experiment.” At this juncture it seems natural to ask; why introduce a fictional device rather than appeal to practical reason directly? If it is practical reason that justifies the theory of justice then its principles could be appealed to directly rather than employing a fictional situation that models them. In that case Will Kymlicka would be right when he claims “the contract device adds little to Rawls’ theory.”

Kymlicka’s statement is accurate in an important sense; practical reason, not the social contract, is the core of Rawls’ argument for a theory of justice. But, as Kymlicka acknowledges, the contract is a useful heuristic device that vividly and precisely renders all relevant features of practical reason in a single, coherent, and organised perspective which can be used as a test for determining which principles are in accord with practical reason as Rawls understands it. Rawls’ strategy, then,
will be to present his principles of justice as the outcome of a procedural test that clearly models practical reason.\(^A\)

A moral evaluative test can be either “definitional”—wherein something is good or right because it passes the test—or “heuristic”—wherein something passes the test because it is good or right (to take another set of terms from Kukathus and Pettit).\(^89\) While the test Kant applies to maxims—the categorical imperative—directly defines the moral law, Rawls’ test is better understood as a heuristic device which explicates and systematises our reasonable and rational considered convictions by modeling the reasonable and rational features of practical rationality. As discussed in chapter two, because practical reason is purely formal for Kant, it can be expressed in a sort of logical consistence test, but Rawls’ understanding of practical reason is not such that it can be defined by any unidimensional logical test.

Now I go on to ask: 1) what is Rawls’ contract device? 2) how does it embody practical reason? and 3) what substantive principles does it support? The primary concern here is with the second question and although the other questions are touched on in the process of answering it I will say little about them directly. To give context to the discussion then, I begin with a few brief comments on questions 1 and 3, the basic details of which are well known and the intricacies of which are the subject of a vast literature.

\(^A\) Given this understanding, I think it becomes easy to see precisely why Michael J. Sandel can claim that Rawls fictional contract is not “a pale form of an actual contract and so a morally weaker thing” but rather a “pure form of an actual contract, and so a morally more powerful thing.” (Michael J. Sandel, *Justice: What’s The Right Thing To Do?* (New York: Farrar, Stuart & Giroux, 2009) p. 151) It is taken to be more powerful than any actual contract because it perfectly embodies all of what we hold to be principles of practical reason, generating fundamental moral-political principles which are perfectly fair.
Rawls’ original position is a hypothetical situation wherein rational contractors representing particular individuals within society choose the principles to govern the basic structure of that society. The situation has two main structural features: 1) the contractors are situated equally to one another, and 2) the contractors are situated behind a “veil of ignorance” such that they do not know any specifics about the individual whose interests they are contracting on behalf of, including their specific conception of the good, their position in society or even their particular natural talents and abilities.

Although contractors cannot seek to advance the particular idea of the good which belongs to the citizen they represent, because they do not know what that conception is, they can seek to secure for that citizen certain all-purpose means which are necessary for the exercise of our moral powers and the pursuit of any determinant scheme of final ends; Rawls calls these primary goods.

In such a situation Rawls argues the contractors would select his two principles of justice: 1) A principle of liberty which guarantees, for all citizens equally, a broad set of fundamental individual freedoms, and 2) a two part principle which requires A) that all social and economic benefits be distributed equally unless allowing inequality benefits the least well off—the difference principle—and B) that where social and economic inequality is allowed it must attach to positions which all citizens have a far and equal opportunity of achieving—

---

^“Each person has an equal claim to a fully adequate scheme of equal basic rights and liberties which is compatible with the same scheme for all; and in this scheme the equal political liberties, and only those liberties are to be guaranteed their fair value.” (Rawls, PL, p. 5.)
the equal opportunity principle. The principles would then be arranged into a coherent system by placing them in a priority order with principle-1 (liberty) taking precedence over principle-2 (social and economic benefits) while sub-principle-B (opportunity for social and economic benefits) trumps sub principle-A (social and economic benefits themselves).

Again, these principles are justified because they would be selected in the original position and Rawls thinks he has come up with a description of the original position that accurately models practical reason. In the original position, the rational is modeled directly by making the contracting parties rational. They would be rational, in the first place, in the Humean/rational-choice-theory sense of instrumentally rational. As well, contractors are presumed to represent an individual who is rational, in the sense of possessing a “capacity to form, follow and revise their individual doctrines of the good,” and at any given time in possession of some particular conception of that good (though what this consists in is

---

A “Social and economic inequalities are to satisfy two conditions: first, they are to be attached to positions and offices open to all under conditions of fair equality of opportunity; and second, they are to be to the greatest benefit of the least advantaged members of society.” (Rawls, PL. p. xx.)

B This scheme of principles is thought to resolve the intuitionism vs. utilitarianism deadlock. The leading concern with the utilitarian principle is that in certain circumstances it requires extreme and even ultimate sacrifices or violations of an individual or small group when this is in benefit of the greater overall utility. Accepting such unequal treatment is not rational, on Rawls’ account. Inflicting it or accepting it violates the idea of treating people as free and equal. For Rawls, individuals as free and equal must be 1) protected by certain absolute and inviolable rights and liberties which would act as limits on the way individual can be expected to sacrifice for the benefit of others and 2) guaranteed a share in the benefits and burdens of society according to a scheme which can be considered fair or just, and all this prior to any calculation of utility. In this way consequentialist thinking is circumscribed by a set of deontological principles. According to Rawls, this corresponds to the bulk of normal and commonplace moral thinking and the most considered moral convictions of ordinary people as well as many philosophers. What distinguishes the theory of justice is that, instead of an ad hoc jumble, Rawls claims to have made these intuitions into a coherent and sensible system based in an account of practical reason.
unknown). In this way, citizens’ “capacity for a conception of their good...is modeled within the procedure by the rationality of the parties.”

The contractors themselves are not reasonable—they have no sense of justice. Their sole concern is with calculating the best way to advance the final ends of the party they represent by protecting their interests—which in the original position means securing access to primary goods (liberties, opportunities, wealth etc.). In the original position the reasonable and the associated sense of justice is represented indirectly by the context or conditions under which the purely rational contractors make their selections. As Rawls writes:

Citizens’ capacity for a sense of Justice is modeled within the procedure itself by such features as the reasonable condition of symmetry (or equality) in which their representatives are situated as well as by the limits on information expressed by the veil of ignorance.

In the first place, because they are situated equally, no one is in a position to use force to get others to consent to a system contrary to their interests. The fact that they are “symmetrically situated” models the equality of citizens which is part of the fundamental idea of society. Secondly, because of the veil of ignorance, the act of deciding what is best for any one citizen is equivalent—has the same outcome—as deciding what is best for all. Because of the structural constraints on knowledge in the original position, the rational contractors must take account of everyone’s basic good as if it were one’s own. In this way, even though the contractors are not themselves concerned with subjecting their rational interests to the constraints of the reasonable by recognising the burdens of judgment and

---

^[For more on this see: Susan Moller Okin, “Reason and Feeling in Thinking About Justice” *Ethics* 99, no. 2 (1989).]
proposing fair standards all could reasonably accept, they are forced to do so by the conditions of the original position.A

Rawls’ theory of justice, then, is the outcome of rational choice under controlled conditions representing the reasonable—which is to say the rational within the constraints of the reasonable. Notice, here, that the original position not only models the reasonable and the rational but the priority of the reasonable over the rational. In the original position rational deliberations are subject “absolutely” to reasonable conditions “the modeling of which makes the original position fair.”

The tendency towards equal distribution of primary goods represented in the proposed principles of justice is essentially a function of the fact that it would be irrational to agree to a less equal share while it would be unreasonable to expect more than an equal share. Not accepting less than ones share is connected to the advancement of one’s rational advantage. The idea of not expecting more than ones’ own share—of respecting fair limits on our rational pursuit of the good as we see it—is connected directly to the reasonable. Limits that we owe to others to impose on our reason—the rational pursuit of our final ends as we understand them—are what we would call reasonable limits, on Rawls’ scheme. These limits confine reason to what can be made public, what can be justified to others who do not, and could not reasonably be expected to share our comprehensive ethical view. The rational is not public; the reasonable is. Constraining the rational by the reasonable

---

A Public political principles of justice selected in the original position are guaranteed not to be coloured by any particular comprehensive doctrine because of the fact that which comprehensive doctrines a given citizen is to actually endorse is unknown. Thus, it is ensured that the principles of justice are genuinely public—acceptable to any reasonable citizen.
is what makes practical reason, public reason. As Rawls puts it “it is by the reasonable that we enter the public world.”

Reasonable conditions limit what can be put forward as good reasons. The structural conditions of the original position enforce these reasonable conditions by ruling out the possibility of illegitimate non-public biases. But what reason is there to justify the privileging of the reasonable over the rational within practical reason? Perhaps the best way to think of it is the following: the rational is what justifies individual beliefs while the reasonable is what justifies appealing to a belief in support of the exercise of political power. To make reason public by subjecting it to the limits of the reasonable does not involve a challenge or denial of the truth of those reasons or the goodness of those values on some epistemologically higher plane, but a challenge or denial of the fact that those features can be a legitimate basis for exercising coercive force over other fully reasonable and rational individuals who do not share those judgments and convictions.

That the reasonable is prior to the rational in this way is the reason why the right takes priority over the good in Rawls’ theory. This division and, moreover, hierarchy within Rawls’ conception of practical reason is the reason his moral theory takes the form of a theory of justice rather than an ethic. No ethical theory can form the basis of a workable political theory under the conditions of reasonable pluralism, which are simply a correlate of freedom.

This leads into the issue of the close connection or, I think it could be said, identification of reason and freedom that exists in the moral theory of Rawls as it does in Kant’s moral theory (albeit in a different way). For Rawls, “freedom at the
deepest level calls upon the freedom of reason, both theoretical and practical, as expressed in what we say and do.”

How the parties deliberate in the original position models, as we have seen, the rational and in doing so it also models what Rawls calls “rational autonomy.” Insofar as we freely exercise our moral power to form and pursue a conception of the good we are rationally autonomous, but rational autonomy is only part of autonomy because it represents only one part of practical rationality. We are “fully autonomous,” for Rawls, only when we are also reasonable.

Full autonomy would correspond to acting in accord with practical rationality as a whole including both the reasonable and the rational; it is rational autonomy made public by operating within the constraints of the reasonable. Full autonomy is modeled in the original position only via the addition of structural aspects that model the reasonable (how the parties are situated with respect to one another and the limits on information). As Rawls puts it, “full autonomy is modeled by the reasonable conditions imposed on the parties as rationally autonomous.”

Rawls is careful to point out here that full autonomy is a moral-political value not an ethical value as it was for Kant. He neither affirms nor denies that being autonomous, in any sense, is the highest good of human life, but that full autonomy as a political value can be accepted as a modular component of any conception of the good which shares in the fundamental ideas and conceptions of practical reason.

The thrust of this section is to establish that Rawls’ theory of justice is the “result of a procedure of construction in which rational persons,...subject to reasonable

A Sometimes, in Rawls, this is also referred to as “political autonomy.”
conditions, adopt the principles to regulate the basic structure of society.” To further clarify, it is now required that I specify more precisely what is meant here by “construction.” What Rawls proposes as fair terms of cooperation are not established by a transcendent or external authority of any kind, nor are they established on the basis of any discoverable natural law. His political liberalism neither affirms nor denies the possibility of such foundations. What he does affirm is that fair terms of cooperation can be constructed by an undertaking amongst citizens themselves—that an order of political values is the outcome of the activity of practical reason itself.

“Constructed” does not mean made up out of nowhere. That moral-political principles and values are constructed means they are the outcome of a process that embodies practical reason. To be clear what is constructed, according to Rawls’ political constructivism, is the content of the political conception of justice—the principles of justice selected in the original position. What is not constructed is the original position; this, as Rawls says, is simply “laid out.” The social contract procedure and the original position in particular are the pre-existing material from which the principles are constructed, and, in turn, the starting points of the procedure are the principles and conceptions of practical reason which are taken as basic. Thus “the political constructivist regards a judgment as correct because it issues from the reasonable and rational procedure of construction.” By modeling the reasonable and the rational the original position is a way of identifying public moral-political principles. If the procedure expresses the principles and conceptions of practical reason it will yield the right principles.
5) Principles of Justice: Application

Once derived, principles of justice, must of course be interpreted and applied in society by actual flesh and blood people. That is why “it is essential that a liberal political conception include, besides its principles of justice, guidelines of inquiry that specify ways of reasoning.”114 In general, the way of reasoning Rawls proposes is public reason. In order to understand how public reason operates and why it is meant to operate the way that it does in the lives of citizens, I begin by highlighting the key fact that public reason is a wholly relative term for Rawls.

There is no private reason for Rawls.115 Reason is about justifying things to others. Practical reason is a way of deciding what is to be done by some group whereby individuals propose reasons to other members of that group. A particular way of reasoning is always public with respect to some group and non-public from the point of view of those outside the group. The way of reasoning of any given group depends on the self-understanding of that association, or as Rawls writes, “different procedures and methods are appropriate to different conceptions of themselves held by individuals and corporate bodies.”116 Ways of reasoning that would be public from the perspective of members of, say, a particular church community would not be public reason for other members of the same political society.

So, if there is no absolute difference in kind between public and non-public reason what do those terms mean for Rawls? and, what does it mean to claim citizens have a duty to use public reason? or, to put a point on it, public reason is public relative to what? What Rawls classifies under the heading “non-public
reason” are the many different reasons that belong to various groups within the background culture of a civil society, and are non-public from the perspective of other members of the “public political culture.”\textsuperscript{117} Public reason as the “reason of citizens” belongs to the public political culture. Within a political society, then, there is only one public reason which is appropriate to the broad self understanding of that society.\textsuperscript{118}

As citizens we have what Rawls refers to as a “duty of civility”\textsuperscript{119} to make our reason public by subjecting it to the constraints of the reasonable when fundamental political questions are at issue. In other words unlike the purely rational contractors of the social contract artifice, real citizens are duty bound to respect moral-political values by exercising their moral-political capacity for a sense of justice and acting within the limits of the reasonable. For Rawls, “the limits of public reason are the limits we honour when we honour an ideal: the ideal of democratic citizens trying to conduct their political affairs on terms supported by public values we might reasonably expect others to endorse.”\textsuperscript{120}

The need to make reason public does not apply either to non-political values or to personal or non-public associational reflections on political values which take place in the background culture.\textsuperscript{121} Here one is at liberty to judge and act on the basis of what one regards as the whole truth as one sees it according to their purely rational capacity for a conception of the good. The need to appeal to reasons that are public reasons is limited to political advocacy in public forums, political campaigns, voting, and, most clearly, to any government agents or representatives acting in official capacities or in official forums.\textsuperscript{122} When we do take part in the
public life of society in these spheres by forming, advocating, or acting on the basis of positions regarding fundamental political questions “we should sincerely think that our view of the matter is based on political values everyone can reasonably be expected to endorse.”

This commits Rawls to a rejection of a popularly held understanding of democracy wherein public institutions are regarded as fora for advancing our most deeply held convictions about the good. Roughly, the basic idea, for this understanding of democracy, is that people vote their personal preference and the majority rules. Such a majority-rule democracy entails a failure to respect the duty of civility. If we vote our preference in the knowledge that those preferences may be contrary to what other reasonable citizens could support, we have turned our backs on the ideal of governing our collective affairs on the basis of reasons all can accept, and ultimately on the very idea of society as a fair system of cooperation amongst free and equal citizens.

Acting reasonably, acting from a sense of justice, respecting the bonds of civic friendship by fulfilling our duty of civility; these involve adherence to what Rawls labels “the liberal principle of legitimacy” — the core of which is simply the notion that political power can only be exercised legitimately by a government whose authority derives from the fact that the values and principles on the basis of which it operates are acceptable to all reasonable and rational citizens.

Thus far we have seen how public reason gives us principles of justice and the basis for their application. In the next section we consider why Rawls thinks this moral-political theory could actually form the basis of a stable social order.
6) A Two Stage Theory

Rawls' theory is unfolded in two basic stages corresponding, more or less, with the previously discussed aims of addressing issues of desirability and feasibility respectively. The former can be discussed under the heading of a freestanding conception of justice, while the latter falls under the heading of an “overlapping consensus”\(^\text{126}\) of reasonable comprehensive doctrines which converge on said freestanding conception.

Stage one is to develop a freestanding conception of justice complete with “its own intrinsic (moral) political ideal.”\(^\text{127}\) Specifically, this stage consists in an endeavour to show how rational agents subject to reasonable restrictions, as modeled in the original position, would select certain principles to regulate the basic structure of society. The bulk of this chapter has been devoted to this stage.

It is important for Rawls that he begins by developing a moral theory that is presented as independent or “freestanding” with respect to comprehensive doctrines so as to arrive at what he regards as the right kind of consensus at stage two of his project. Rawls' political conception of justice does not set out be fair to the various conceptions of the good that happen to exist in the background culture by striking a balance or compromise between them. Rather, his theory of justice is meant to honour and respect those rational beings with the capacity to form a conception of the good—free and equal persons—by treating them fairly.\(^\text{128}\)

Working only from fundamental ideas and conceptions of practical reason Rawls presents a freestanding conception, with the hope, though there is no guarantee, that it can gain the assent of citizens.\(^\text{129}\) He does not, however, want to design a
conception of justice with, from the outset, an eye towards whether or not actual comprehensive doctrines would agree to it. For him “the question of stability does not arise until the principles of justice are already provisionally selected.”

Rawls’ hope is that his freestanding conception can be the focus of an overlapping consensus. It would seem a futile and misguided enterprise to place our main hope for any lasting peace, freedom, equality, and prosperity in a scheme which is next to impossible to succeed. Thus, Rawls needs to justify his hope. Towards this end, at the second stage of his project, Rawls’ focus shifts to the question: “How is it possible that there may exist over time a stable and just society of free and equal citizens profoundly divided by reasonable though incompatible…moral doctrines?”

A society based on a single comprehensive doctrine could only be maintained by force, coercion, and heavy-handed sanctions which rein in citizen’s autonomy. Such a society inherently tends towards instability. Besides this, such a society is already undermined by the simple fact of existing pluralism in contemporary Western democracies and, moreover, the entire emerging global cultural marketplace. The alternative of allowing individuals to think and act freely, according to their own reason, their own understandings of the good, would require an ethically neutral state, most likely some form of democracy.

As Rawls points out, a stable democratic regime must be supported by a significant majority. If it is to freely gain such support the fundamental principles on the basis of which it operates must be sustained by the comprehensive doctrines of its citizens. Thus, Rawls thinks his principles of justice can be stable only if they
can be the focus of an overlapping consensus. He argues they can. Stage two’s principle conjecture is that because the principles are developed from principles of humanity’s “common practical reason” and from ideas of society and person which are also held in common, we can expect them to be the focus of an overlapping consensus.

The consensus Rawls has in mind runs much deeper than a mere *modus vivendi*—a compromise, balance of power, or tentative and uneasy way of living together. It is up to the adherents of each doctrine to figure out how justice fits within their comprehensive view, but it should be supported as a part of that view. In order to provide real stability, the overlapping consensus Rawls envisions is one which is “moral in its object and motivation.” That means the principles of justice can be incorporated as an element of the various comprehensive doctrines of diverse citizens who, though they are separated by their complete visions of the good life, are nonetheless morally united by their consensus and bound together by the relationship of civic friendship and the corresponding duty of civility.

7) An Advancement of Practical Reason Beyond Aristotle & Kant?

Rawls thinks his theory of practical reason as public reason—which could be stated alternately as the rational subject to the constraints of the reasonable—represents an advancement over the conceptions of practical reason proposed by previous theories; of particular concern here are the forceful and influential paradigms of thought represented in Aristotle and Kant’s respective conceptions of practical reason. His theory of practical reason is considered an advancement in a rather particular sense, however.
Rawls does not want to claim that the comprehensive moral doctrines of Aristotle or Kant, views based on specific accounts of practical reason, are outright false and ought to be replaced by his theory of justice. On the contrary, he would want someone with an essentially Aristotelian or Kantian ethical doctrine to be able to sign on to the overlapping consensus which supports the political conception of justice. The problem with these other conceptions is not that they are incorrect (Rawls neither affirms nor denies their central claims). The problem is that they, like any comprehensive doctrine, are inappropriate as a basis for a public social and political systems because they cannot reasonably/legitimately be expected to be adopted by other reasonable citizens.

By contrast, Rawls’ conception of practical reason is supposed to be capable of supporting values and principles that can be publicly justified to other reasonable and rational citizens when it is restricted to the realm of the moral-political. It is in this sense that Rawls claims his moral rationalism represents a development in moral theory. With this in mind I consider some of the ways Rawls’ moral rationalism contrasts with these landmark predecessors.

Recall from chapter one that, for Aristotle, there is a single idea of the good. Human beings have a final end within an ethical order of reality. Via a rational perceptive capacity we can know our place in the natural order. On the basis of such knowledge of the independent moral order, and our place in it, we are able to make true ethical judgments. On this conception, a just political arrangement would then be seen as one that promotes our good.
As such, Aristotle is a prime example of the sorts of views Rawls discusses under the heading “rational intuitionism.” Rawls identifies four closely related ways that his own approach to moral theory, which he calls “political constructivism,” differs from rational intuition. 1) Unlike Aristotle’s “good,” Rawls’ “right” is constructed by a rational procedure, not discovered in an independent ethical order. 2) Aristotle’s ethics begins and ends with theoretical reason. For him practical reason is essentially the way we orient our lives towards the good, which is known to be theoretical reason, in its object, by theoretical reason, as the instrument of its discovery. Because he is not looking to discover what is good in nature but to create what is right in accordance with reasonable and rational principles Rawls makes a decisive shift of emphasis from theoretical to what he understands as practical reason.

3) While Aristotle’s ethics commits him to a particular epistemological theory of truth, which Rawls would characterize as a sort of traditional correspondence theory, wherein true ethical statements correlate somehow to facts about the ethical order of the world, Rawls’ political constructivism does not employ the notion of truth at all. Rawls says that he relies on the notion of the reasonable which is clearly not the same as truth. 4) However, Rawls does acknowledge that he needs to rely on a rather complex notion of the individual and society and this of course will be necessary for him to get to the notion of the reasonable which he employs in place of truth. This constitutes a fourth contrast with Aristotle, who

\[\text{\textsuperscript{A}}\text{The focus here is on the first sense of rational intuitionism. See the introduction to this chapter.}\]

\[\text{\textsuperscript{B}}\text{(See Chapter One page 20)}\]
requires only a rather sparse conception of the individual as someone capable of knowing/perceiving and capable of desiring in accordance with knowledge.\textsuperscript{138}

Insofar as Aristotle goes on to make the political claim that a just society must be one guided by his proposed conception of the good life, Rawls must, at this point, beg to differ. In spite of this, with respect to the aforementioned differences, Rawls’ general strategy of avoidance allows him to skirt conflict simply because his theory is concerned only with political justice, not a complete account of the whole of morality (including ethics). Aristotle, in giving a complete account of the good life is engaged in a different sort of project which Rawls is not competing with. Rawls wants to say: maybe there is an independent order of moral value that we can know and make true claims about, maybe there is not; maybe a life of contemplation is the best kind of life, maybe it is not—regardless, the point is, even if you take the affirmative stance, there are reasonable members of society who do not share those conclusions, and to use state power to repress not-unreasonable views is unjust because it is contrary the fundamental idea of citizens as free and equal and society as a fair system of cooperation between them. Any view of the good life which maintains that it is appropriate to do so becomes, at that point, unreasonable.

Next, I consider some contrasts between Rawls’ political constructivism and a view that is considerably closer to it—the so called “moral constructivism”\textsuperscript{A} of Immanuel Kant.\textsuperscript{139} In contrast to Rawls’ political liberalism, the moral (and, moreover, ethical) version of liberal theory that Kant supports takes the form of a comprehensive doctrine in the sense that it seeks to give a full account of value.\textsuperscript{140}

\textsuperscript{A}“Moral constructivism” is a misleading term in this context because Rawls’ political constructivism is also moral.
What Rawlsian constructivism constructs are moral-political principles and values. What Kantian constructivism constructs is value—full stop. Rawls is trying to find a public basis of justification for questions of political justice rather than articulate a complete worldview to guide the good life (Kant/Aristotle).

For Kant, in contrast with Aristotle, “the independent order of values does not constitute itself but is constituted by the activity...of practical (human) reason itself.” While Rawls works with the same Kantian idea that reason is self-authenticating, he says it is a separate question, indeed I believe he would maintain a metaphysical question, as to whether or not practical reason constitutes the entire order of value for human life. On this metaphysical question political liberalism seeks to remain silent.

While reason/autonomy is the centrepiece of either philosophy, this has a rather different significance for their respective projects. For Kant, autonomy or acting in accord with practical reason constitutes the only intrinsic good while Rawls’ defence of full autonomy as a political value avoids any such “deeper meaning of autonomy.” In remaining silent on a possible deeper significance of autonomy, Rawls also thinks he can avoid Kant’s need to appeal to a metaphysical (transcendental idealist) conception of the person and thing-in-itself which is necessary to constitute the realm of value—the kingdom of ends. The extent to which Rawls succeeds in this regard, together with several further contrasts with Kant’s philosophy are explored in the next section.
8) Practical Reason: A Closer Look

For Rawls, as I have tried to show, “the Reasonable and the Rational are unified within one scheme of practical reasoning” and “the way the Reasonable frames the Rational in the original position represents a feature of the unity of practical reason.”¹⁴³ In spite of this unity, however, The Rawlsian notion of practical reason is one which is fundamentally divided, and moreover, divided, in a sense, against itself—for the reasonable and the rational 1) are separate and irreducible and 2) exist within a scheme of practical reason “which establishes the strict priority of the Reasonable with respect to the Rational.”¹⁴⁴ While Rawls acknowledges that the “constraints imposed on the parties in the original position are indeed external to the parties as rational agents,” he wants to maintain that because “these constraints express the reasonable” they are not “external,” and therefore heteronomous, but part of practical reason and therefore expressive of full autonomy.¹⁴⁵ He claims full autonomy is never something which is externally imposed or required by any authority, even the other citizens who make up a cooperating public.¹⁴⁶

Far from being incidental, it is critical for Rawls’ defence of full autonomy, as the central moral-political value of justice as fairness, that he is able to maintain this. If Rawls is unable to maintain that both the rational and the reasonable are internal to the individual and, moreover, unified within every individual who possesses practical reason (at least broadly defined), then his appeal to the idea of autonomy loses all sense, and, moreover, it seems the promise to defend moral principles on the basis of practical reason will go unfulfilled—all on account of the fact that what
was supposed to be “practical reason” would have become external and heteronomous. To gauge the overall success of Rawls' version of moral rationalism (political constructivism) it is necessary to delve somewhat deeper into the adequacy of his notion of practical reason as public reason.

Rawls is quite clear about certain negative specifications of his moral rationalism. He is decidedly not continuing any form of what he sees as a particular sort of traditional “enlightenment project” of trying to find a comprehensive secular doctrine founded on the rational.\(^{147}\) He is not articulating a form of moral rationalism that defends a “secular doctrine founded on reason and viewed as suitable for the modern age now that the religious authority of the Christian age is said to be no longer dominant.”\(^{148}\) Here Rawls is distancing himself from enlightenment liberalism and other forms of moral rationalism in two key ways; first, in that he does not seek to do away with or even compete with comprehensive doctrines, religious or otherwise—in short he does not claim public reason is the fount of all value—and second, in that he is not trying to ground his political values in reason as understood by previous moral rationalists, such as Kant, or even by other contemporary Kantians, such as Habermas.

Many other moral rationalists (Kant can be taken as a paradigmatic case) have tried to derive moral principles purely from something similar to what Rawls calls the rational or some even more restricted idea of rationality, while Rawls is clear that he has “no thought of deriving those principles from the concept of rationality as the sole normative concept.”\(^{149}\) Rawls' appeal is to the reasonable and the rational. For Kant, practical rationality and true freedom is a matter of following
formal logical rules, but, as discussed in chapter two, this idea faced serious criticism by Hegel and others who doubted the possibility of deriving substantive moral principles from logic. Rawls is largely in accord with these doubts, making the broad claim that all “serious attempts to derive the reasonable from the rational do not succeed, and so far as they appear to succeed, they rely at some point on conditions expressing the reasonable itself.”

Kant’s ethic contains substantive commitments to a form of reciprocity and the idea of human beings as free and equal ends-in-themselves, which somehow sneak their way into his supposedly formal/logical categorical imperative.\(^\text{A}\) Here, Rawls likely has Habermas in mind as well. Habermas attempts to ground moral validity in a human, universal, communicative capacity for reason, which presupposes rules for discourse that embody an idea of reciprocity which, on Rawls’ view, cannot be grounded in rationality itself, and must, therefore, come from someplace else—which he thinks has to be the reasonable. In fact, the main innovation in Rawls’ account of public reason, as he himself understands it, is the central role of the reasonable and its duty of civility.\(^\text{151}\) To put it succinctly, what Rawls thinks he has added to moral rationalism is the reasonable.

Rawls’ appeal to the rational is relatively straightforward. For one thing, it is in line with what is normally understood by reason or rationality. It is a cognitive, epistemological concept which is connected to the idea of truth or validity. Because there seems to be no space apart from reason to look for justification of principles of reason (rules of inference etc.) it is also comprehensible why the rational might be

\(^\text{A}\)See discussion of transubstantiation in chapter two.
regarded as a self-authenticating starting point. However, the same cannot be said about the reasonable, and the real normative work in Rawls' theory is being done, not by the rational, but the reasonable.

We employ “the rational” in exercising our capacity for a conception of the good to arrive at what we take to be true or valid ethical values. We then go on to cooperate with other citizens in order to further our own vision of the good. However, the only values which can be justified to all members of a democratic political community are those which are said to make this cooperation fair, that is, moral-political values, not ethical values. Although cooperating citizens are taken to be rational and to have conceptions of the good, the political values, in virtue of which their cooperation is said to be fair, do not come from the rational but follow directly from the reasonable. That is why Rawls says his political conception’s sole standard of correctness is reasonableness; and, of course, the reasonable’s ideal is a form of reciprocity.

Rawls thinks the moral sceptic cannot be answered by deriving this idea of the reasonable from the rational. Clearly, for him, this is not seen as a problem thanks ultimately to the fact that the overall scheme of practical reason includes the ideas of practical reason. These ideas are the ultimate source of Rawls’ moral-political value. Their importance, therefore, can scarcely be overestimated. On this conception, then, “being reasonable is not an epistemological idea (though it has epistemological elements). Rather it is a part of a political ideal of democratic citizenship.” We only get to the reasonable, with its ideal of reciprocity, because of the conceptions of reason—that is, because we already think of society as a fair
system of cooperation between free and equal citizens. Thus the heaviest normative lifting is done by the conceptions of practical reason—organized around the fundamental idea of society. This is why Rawls says “the basic idea of society as a fair system of cooperation...is developed into a conception of political justice.”

Somewhat surprisingly and disconcertingly, Rawls tells us these conceptions are merely taken from the existing ideas and institution of culture. Rawls characterises the starting point of his theory by admitting that he begins “by looking to the public culture itself as the shared fund of implicitly recognised basic ideas and principle.” Again, in his own words: “Society’s main institutions, and their accepted forms of interpretation, are seen as a fund of implicitly shared ideas and principles.” Elsewhere he writes of political liberalism: “The basis of this view lies in fundamental ideas of the public political culture.” So it turns out that we are meant to “think of citizens as free and equal persons” simply because “we start within the tradition of democratic thought,” and we only get to the idea of practical reason as public reason because of the pre-existing notion of a “shared political life.” Instead of modeling a rational procedure that yields valid results, the original position simply “models what we regard—here and now— as fair conditions.” Instead of modeling reason, it actually “models what...we regard as acceptable restrictions on reason.”

This is all disconcerting for several reasons which I shall attempt to make plain. Rawls begins by holding out the unequivocal promise that he will justify his moral-political principles, independent of any external authority, on the basis of our own practical reason—the result being a freestanding political conception of justice.
He also tells us that the rule of practical reason is justified because it comes from us—it expresses our full autonomy. His conception of practical reason, and therefore autonomy, are robust notions. His practical reason is not merely formal. Substantive notions, centrally the ideal of reciprocity, enter practical reason via the reasonable and ultimately the conceptions of practical reason. That seems well and good until it is learnt that these ideas are simply withdrawn from the fund of existing ideas, principles, institutions, and interpretations of current society.

As I suggested above there is good reason for taking the reasonable as a fundamental starting point, but if Rawls expects a specific complex notion like the reasonable to be taken as a fundamental starting point, surely one would need a reason better than one which says something to the effect that: we start from these ideas because they are what a lot of people happen to think around here at the moment.

What Rawls calls autonomy requires accepting ideas simply taken from the existing public culture. By any sensible account this cannot be autonomy but heteronomy. I do not mean to suggest that to be rational and autonomous means one must completely transcend context and escape an existence embedded in material and social structures. Being socially constructed, to some degree, does not equate with being un-free and irrational; but, for Rawls, the very ideas which define full autonomy are lifted, uncritically, straight from the public cultural fund, which is thereby treated as authoritative and final. Surely this cannot be called autonomy in any meaningful sense. Autonomy means thinking for oneself, it is fundamentally opposed to unreflectively accepting what is given.
For the more traditional moral rationalists Rawls criticises—again I will take Kant as an example here—a shared point of view is possible because of our common human reason; the moral law, given by the categorical imperative, is the same for us all. Rawls does well to avoid the traps associated with either metaphysical appeals to supersensible realms of things-in-themselves or with attempts at reducing practical reason to logic. But, in order for him to find a shared point of view, he makes an uncritical (and, I suspect, unnecessary) appeal to the external authority of a shared tradition.

Justification, for Kant, as well as for Kantians, such as Rawls or Habermas, is justification to some audience. For Kant, as discussed previously, and for Habermas, as I will be discussing in the next chapter, the audience encompasses all rational agents. Thus, for Kant and Habermas, practical reason is something which is meant to apply to all rational beings. It also means that reciprocity is specific to the human species (at minimum). For Rawls, by contrast, public reason is only meaningful given an existing political community—a bounded society with a common political identity.

Rawls’ reciprocity only applies to a specific insider group since public reason, as I showed above, is public, not absolutely, but relative to a pre-existing community with a certain understanding of itself. Instead of the scripture and clergy, which embody the particular values of a church community, Rawls might appeal to a constitution and founding figures who embody the ideas of a national political culture. The appeal, however, is still to external, heteronomous authority not self-
legislating, autonomous reason. So, in avoiding the Kantian snares, Rawls appears to have slipped into an equally fatal trap of his own.

Besides being incongruous with his stated value of autonomy this reliance on cultural tradition is inherently problematic. It takes what was presented as a form of moral rationalism and turns it into a version of communitarianism. It reduces what was supposed to be a critical, rational perspective to a form of uncritical traditionalism. Starting within a contained tradition we run the risk that this cultural fund of ideas might be bankrupt. Accepted ideas can turn out to be irrational, even monstrous, upon critical scrutiny. To appeal to such ideas purely on the authority of their popularity, even when they seem like good ideas, is morally dangerous.

To be clear, the problem is not with reciprocity, but with the idea that it only comes from culture. In the next chapter I will consider how Habermas, without appealing to a particular cultural tradition or returning to Kantian formalism or metaphysics, runs an argument which attempts to show that the basic idea of reciprocity is inherent in the universal human capacity for reason itself. On Rawls’ scheme, reciprocity is certainly part of what he calls public reason, but the reasonable—the public-making aspect of public reason—only applies to a particular group of people who see themselves as cooperating members of a political community and happen to have adopted a certain set of ideas about what society should be, what individuals are like, and how they should be treated. As Rawls writes, it is “shared political conceptions” which serve as the “basis of public
reason,” but these are shared only because they happen to be implicit in the public political culture.\textsuperscript{164}

These problems appear even more acute when we take into consideration not only the principles themselves but also the second aspect of Rawls' theory of justice—the rules for their application, which could be summed up as applying a (reasonable) publicity constraint on the purely rational, and moreover, all non-public forms of reasoning. Publicity means appealing only to “commonly shared beliefs confirmed by methods of inquiry and ways of reasoning generally accepted”\textsuperscript{165} which is intended to encompass “common sense”\textsuperscript{166} and well-established, widely-held “procedures and conclusions of science”\textsuperscript{167} and to entail that

principles are to be rejected that might work quite well provided they were not publicly acknowledged, or provided the general facts upon which they are founded are not commonly known or believed.\textsuperscript{168}

This formulation of what it means to reason publicly has some deeply problematic consequences. For one, it can bar one from appealing to well established facts and requires one to accept the manifestly irrational and absurd in certain circumstances. For example, according to a recent Pew Research Center poll\textsuperscript{169} 18% of Americans believe that their president is a member of the Muslim faith, which is hardly surprising given the fact that a similar percentage of the population are consistently found to believe all manner of blatant falsehoods; a case in point being a Gallup poll\textsuperscript{170} which found that 18% of Americans believe that the Sun revolves around the Earth. What is more surprising is that the Pew poll found that another full 43% of the population said they did not know what religion their
president belonged to. A poll by Newsweek\textsuperscript{171} also found that roughly half of Americans held the belief that Christianity historically predates Judaism. This means that if I were to talk about Christianity in the public forum I could not talk about the development of Christianity out of the Jewish tradition or about how all U.S. presidents have been Christian without appealing to controversial facts which are not generally accepted or believed.

I suspect Rawls may be able to adjust his formulation to make some allowances for appeals to well established truths in the face of widespread ignorance by stressing that public reason should accord with generally accepted "considered convictions"\textsuperscript{172}—that is, what people would believe after full consideration in light of all relevant information etc. However, this would not help in cases where even people's well considered convictions are irrational. If public reason means appealing only to already accepted beliefs then it is not possible to launch critiques of existing ideas, principles, institutions, and social arrangements in public fora—and without that possibility it is difficult to see how rational progress or development is possible.

I also grant that the conservative standards Rawls is appealing to only pertain to establishing and applying the principles of justice and must not therefore be taken to infect every practice or form of discourse in actual societies. So, in Science, for example, Aristotle would be free to publically express his arguments for a round earth, Copernicus his arguments for a heliocentric solar system, or a modern day equivalent to express some new description of the galaxy, despite the fact that they are all forwarding descriptions that are not publicly acknowledged
and founded upon general facts which are not commonly known or believed. This however only serves to highlight the problem that the cognitive, reason-based developmental processes germane to other spheres of human life are denied to the moral-political sphere. How are we to explain the development of a society based on slavery to one that is not if it were not for a hand full of people who were willing to stand up and say “slavery is wrong and here is why” even when that position is likely to be met widely with scorn, ridicule or worse?

For these reasons, Rawls' public reason appears to have frighteningly dogmatic and traditionalist consequences which forestall the possibility of continued human development through radical critiques of, and utopian breaks with, the currently dominant tradition. I know Rawls does not want to accept this consequence. In a footnote he says “no specification of these criteria [principles of correct practical reasoning] can be declared to be and treated as final; any rendering of them is always open to be checked by critical reflection;” but, how are they supposed to be checked when we are duty bound to restrict our public debate to what is currently accepted? Rawls gives no account, indeed there is no room in his theory for an account of how this “critical reflection” is supposed to be carried out. On the contrary, the picture of the well ordered society that Rawls paints looks more like a liberal end of history than a world in which critique and advance to more rational forms of society is possible. The ideal of rational progress or development requires the ability to critique established consensus as reason dictates. Thomas McCarthy expresses similar concerns:

to suppose that the stock of shared political ideas and convictions is in some way given, there to be found and worked up, or that it could
somehow be fixed by the theorist, is to hypostatize or freeze ongoing processes of public political communication whose outcome cannot be settled in advance by political theory.\textsuperscript{173}

Also, as O’Neill notes, Rawls’ public reason is an insider’s reason which excludes all outsides, not only foreigners but also “citizens who stand back from the way things are, and ask whether they should be that way.”\textsuperscript{174} From the perspective of such outsides what insiders call reasonable can seem anything but. Again, Rawls intention is for criteria of correct reasoning to be provisional, but it is not possible for this provisionality to be explained if there is no account of how they can be rationally challenged.

On Rawls’ scheme the content of public reason is set out in advance by the political values and principles which are realized by imposing the limitations of the reasonable. This leads to a particular understanding of what it means to participate in public reason and \textit{ipso facto} what it means to be a citizen of a democratic society. Rawls puts it nicely when he writes: “Public reason sees the office of citizenship with its duty of civility as analogous to the duty of judgeship.”\textsuperscript{175} The job of a citizen, like that of a judge, is to uphold and interpret their inherited principles of justice—to act less like rationally autonomous co-creators in a collective process of will-formation and more like guardians of the authority of tradition. On this view, the practice of practical reason in public arenas is not an opportunity to challenge, debate, and forge new ideas on the basis of what seems most rational, so much as it is an activity whereby we apply existing principles according to existing publicly instituted guidelines for their application. This is why Rainer Forst aptly describes
Rawls’ political discourse as “principle-interpreting” rather than “principle generating.”

Political liberalism takes fundamental ethical questions off the political table. However, if people believe they hold a position for some good reason, and are committed to the idea that they ought to believe whatever has the best reason in its favour, that is, if people are rational (in a broad and fairly ordinary sense), then there is, in principle, reason to think people might be able to narrow their disagreements, even on fundamental ethical issues; therefore public debate may be worthwhile.

One apparent way around these difficulties that remains open to Rawls and which, at times, he seems to actually lean towards, might be to give up on the idea of a freestanding political conception grounded in practical reason and say simply that it is a view which is presented independent of any comprehensive doctrine. Such a view is no longer freestanding; it is “free” from any specific comprehensive doctrine but it is not “standing” on anything. Rawls could then rely on the fact that his political conception is a module that can be plugged in to various comprehensive doctrines which would then support this conception on the basis of its own reasons.

This would entail giving up on the idea of the reasonable as something basic and irreducible to the rational. Ultimately, it would entail the only basis for accepting rational principles is that they are supported by some overall account of the good life. So, although the accounts may be, not one, but many, the only ground for the reasonable is still to derive it from the rational. The original strength of the position Rawls proposed was that it could say to every rational citizen: whatever
else you happen to believe about the good life, there are good reasons why you should accept these principles of justice as part of it. But, if there are no independent, internal political values in justice as fairness which are presented on the basis of good reasons everyone should accept, then literally all one can do is throw out the principles and hope for a lucky convergence of comprehensive doctrines.

If Rawls is not trying to find rational universal principles but principles appropriate for a particular kind of modern Western democracy such as the United States which already happens to have in its public political culture certain explicit, or at least implicit, notions of society and person, then he seem to have abandoned desirability—skipped stage one and gone straight to stage two. Without the desirability argument, however, Rawls is not doing philosophy any longer. Trying to find a description of something that a particular group of people with a fixed set of ideas can agree to is to practice the art of diplomacy or *realpolitik,* not political philosophy, which is what Rawls set out to do. A number of other commentators from a variety of perspectives have recognized, in one way or another, the fact that, as Omid A. Payrow Shabani puts it, Rawlsian liberalism “reduces the distance between the “aught” of the critical ideal and the “is” of the real in favour of the latter.”  

---

A Amit Ron has suggested that one might understand Rawls as offering a Frankfurt-School style immanent criticism demanding that existing institutions stand up to the values that they promise. While it may be helpful to understand Rawls in this way I argue it is important to see how it is both necessary and possible to go beyond this. Amit Ron, “Rawls as a Critical Theorist,” *Philosophy and Social Criticism* 32, no 2 (2006), 175.

allowing for what Seyla Benhabib calls “fulfillment” without “transfiguration”—that is, by closing itself to the possibility of any qualitative utopian break and confining itself to the realization of implicit potentials of the present.

This leads me to a separate point; the division within practical reason, which places the moral-political core of the theory within the reasonable and relegates ethical value to the proverbial black box of a publicly inscrutable conception of the good, lands Rawls somewhat closer to Hume. Although, for Rawls, our conception of the good is derived from the rational, instead of being basic facts about what we happen to like or dislike independent of any rational process, as is the case with Hume, on either account, our final ends are occluded from rational public debate and scrutiny. Similar to the way reason, for Hume, is about finding effective means to pre-given ends, public reason, for Rawls, is about finding a workable way to get along with citizens with final ends pre-given by their (not unreasonable) comprehensive doctrines. By placing the rational component of practical reason inside a non-public black box and making the reasonable the basis for his political values, Rawls cuts his own moral theory off from the cognitive part of practical reason which could give it rationally binding normative force.\(^\text{A}\) Moreover, on this conception the cognitive component of practical reason is constrained by the merely traditional.

Rational persons are appropriately responsive to the weight of reasons. Reasonable persons “desire for its own sake a social world in which they, as free and equal, can cooperate with others on terms all can accept.”

People may have this desire on the basis of non-public reasons or because they have been enculturated to have it, but whether or not people happen to have this desire is not itself a cognitive issue. The fact that people are expected to have this desire is purely on account of the fact that it can be formed on the basis of ideas latent in the culture.

Rawls wants to articulate a theory of justice which accords with our considered judgments, but in order to answer the question “whose considered judgments?” Rawls makes recourse to a presupposed, pre-existing political community, and without fully realizing it ends up with a moral theory very similar to Macintyre’s. He wants citizens who do not agree about ethical values to be able to agree that they owe certain basic things to each other as fellow citizens of a political community. Here, however, a community with a particular shared understanding needs to be presupposed. The upshot is that the rational, which is a cognitive concept concerning what we hold to be true, is limited absolutely by the contingent, particular, socio-culturally specific notion of what is reasonable according to the mainstream of the nation we happen to be citizens of. This hardly seems to accord with anything one would want to call reason—practical or otherwise.

I will conclude with a couple of final points regarding Rawl’s moral rationalism as a form of proceduralism and social contract theory. Part of the

---

problem with earlier versions of a social contract theories which claim that a binding contract is somehow implicit in society, is that it is incompatible with the ideal of autonomy and of binding norms being rationally acceptable to those subject to them. The individuals who are supposed to be party to such a contract would not have the opportunity to agree or disagree or to rationally revise, discuss, consider or deliberate upon the contract. If the contract is supposed to be binding it cannot be binding in virtue of the rational autonomy of individuals but by the heteronomous external force of existing norms and social structure which enforce compliance with the, so called “contract.” It is a “contract” one is born into, not one they rationally accept. It was explained above how Rawls avoids this straightforward objection, owing to the fact that his contract is a device for modeling relevant features of our practical reason rather than a “contract” implicit in the society we are born into. However, once it is clear that the ideas which define Rawls’ contract situation are ideas implicit in culture, it begins to look as if he has merely replicated the same problems at a higher level of abstraction.

Further, as a proceduralist theory, Rawls’ version of the social contract is informed by the basic thought that a fair procedure will yield a fair outcome. Notice here that the notion of “fair” is a presupposition not an outcome of the procedure. While the “guidelines and procedures of public reason are seen as selected in the original position” the reasonable—the public making aspect of public reason itself—is not the outcome of the original position or any other sort of rational procedure, rather it is what defines the original position in the first place. Political values are thus built into, rather than the outcome of, the procedure.
These values act as constraints on the purely rational. A major portion of the
disagreement within the broadly Kantian tradition of moral rationalism will concern
the nature and origin of constraints on the rational (in Rawls’ sense). Rational
choice, for Rawls, is limited by constraints which embody political values that are
ultimately latent in culture. A I have just suggested why this seems problematic. For
Kant, these limitations are an expression of our true rational nature as members of a
Kingdom of ends. Chapter Two considered some ways this seemed problematic.
Next I consider the moral theory of Habermas—another neo-Kantian who suggests
that these constraints are latent in the communicative nature of human rationality.
Like Rawls he is a proceduralist. On his version, valid principles cannot be the
outcome of a hypothetical procedure which models some ideal of practical reason
but must be the outcome of actual, rational, communicative procedures. While
rooted in certain enlightenment notions of reason, the centrality of the dialogical
dimension of this account of reason also harkens, in some ways, back the Athenian
origins of western moral rationalism.

A For a related argument that Rawls is unable to give good reasons why people should exercise this
kind of restraint on nonpublic reason, and that the ideal of public reason ought to place more modest
restrictions on citizens, see Micah Lott, “Restraints on Reason and Reasons for Restraint: A Problem
Notes: Chapter Four

3Kukathus & Pettit, 2-3.
5Rawls, PL, xiv.
7Rawls, PL, xv.
8Rawls, PL, 177.
9Rawls, PL, xxvi-xxvii.
10Rawls, PL, lx.
11Rawls, PL, xlii.
12Rawls, PL, xxxvi.
13Rawls, PL, 77.
14Rawls, PL, 12.
15Rawls, PL, 126-127.
16Rawls, PL, 97.
17Rawls, PL, 96.
18Rawls, PL, xlii.
19Rawls, PL, 12.
20Rawls, PL, 224.
21Rawls, PL, 225.
22Rawls, PL, 253.
24Rawls, PL, 90.
26Rawls, PL, 103.
27Rawls, PL, 240.
28Rawls, PL, 240.
29Rawls, PL, 107-108.
30Rawls, PL, 93.
31Rawls, PL, 116.
32Rawls, PL, 93.
33Rawls, PL, 116.
34Rawls, PL, 107-108.
35Rawls, PL, 100.
36Rawls, PL, n121.
37Rawls, PL, xx.
38Rawls, PL, 108.
39Rawls, PL, 108.
40Rawls, PL, 9.
41Rawls, PL, 9.
42Rawls, PL, 15.
43Rawls, PL, 103.
44Rawls, PL, 89.
45Rawls, PL, 14.
4Rawls, PL, 14.  
5Rawls, PL, 14.  
6Rawls, PL, 18.  
7Rawls, PL, 14.  
8Rawls, PL, 18.  
9Rawls, PL, 18.  
10Rawls, PL, 19.  
11Rawls, PL, 34.  
12Rawls, PL, 19.  
13Rawls, PL, 212-213.  
14Rawls, PL, xlv.  
15Rawls, PL, 19.  
16Rawls, PL, 19. See also: Rawls, PL, xlv.  
17Rawls, PL, 19.  
18Rawls, PL, xlv.  
19Rawls, PL, 19.  
20Rawls, PL, 19.  
21Rawls, PL, 51.  
22Rawls, PL, 50.  
23Rawls, PL, 50.  
24Rawls, PL, 50.  
25Rawls, PL, 49.  
26Rawls, PL, 52.  
27Rawls, PL, 49.  
28Rawls, PL, 49.  
29Rawls, PL, 55.  
30Rawls, PL, 56-57.  
31Rawls, PL, 37.  
32Rawls, PL, 247.  
33Rawls, PL, 37.  
34Rawls, PL, 37.  
35Rawls, PL, 34.  
36Rawls, PL, 52.  
37Rawls, PL, 51.  
38Rawls, PL, 52.  
39Rawls, PL, 34.  
40Will Kymlicka, Contemporary Political Philosophy (Toronto: Oxford University Press, 2002), 61.  
41Kukathus & Pettit, 27.  
42Rawls, PL, 25.  
44Will Kymlicka, 69.  
46Rawls, PL, xlv.  
47Rawls, PL, 104.  
48Rawls, PL, 104.  
49Rawls, PL, 104.  
50Rawls, PL, 79.  
51Rawls, PL, n25.  
52Rawls, PL, n25.  
53Kukathus & Pettit, 47.  
54Rawls, PL, 114.
99 Rawls, PL, p. 25.
100 Rawls, PL, n51.
101 Rawls, PL, n223.
102 Rawls, PL, 77.
103 Rawls, PL, 77.
104 Rawls, PL, 78-79.
105 Rawls, PL, 77.
106 Rawls, PL, xx.
107 Rawls, PL, 99.
108 Rawls, PL, 103.
109 Rawls, PL, 103.
110 Rawls, PL, 104.
111 Rawls, PL, 96.
112 Rawls, PL, xlix.
113 Rawls, PL, xx.
114 Rawls, PL, 223.
115 Rawls, PL, n220.
116 Rawls, PL, 220-221.
117 Rawls, PL, 220.
118 Rawls, PL, 220.
119 Rawls, PL, 219 & 248.
120 Rawls, PL, 253.
121 Rawls, PL, 215 & 246.
122 Rawls, PL, 252.
123 Rawls, PL, 241.
124 Rawls, PL, 219.
125 Rawls, PL, 217.
126 Rawls, PL, 10.
127 Rawls, PL, xlv.
128 Rawls, PL, 40.
129 Rawls, PL, 40.
130 Rawls, PL, 64.
131 Rawls, PL, 40.
132 Rawls, PL, xviii.
133 Rawls, PL, 38.
134 Rawls, PL, 64.
135 Rawls, PL, 90.
136 Rawls, PL, xli.
137 Rawls, PL, 91.
138 Rawls, PL, 91-94.
139 Rawls, PL, 89.
140 Rawls, PL, 99.
141 Rawls, PL, 99.
142 Rawls, PL, 99.
144 Rawls, KC, 319.
145 Rawls, KC, 319.
146 Rawls, PL, 98.
147 Rawls, PL, xviii.
148 Rawls, PL, xxviii.
149 Rawls, PL, 53.
150 Rawls, PL, 53.
151 Rawls, PL, 253.
152 Rawls, PL, 127.
153 Rawls, PL, 54.
154 Rawls, PL, 50-51.
155 Rawls, PL, 62.
156 Rawls, PL, 34.
157 Rawls, PL, 8.
158 Rawls, PL, 14.
159 Rawls, PL, 97.
160 Rawls, PL, 18-19.
161 Rawls, PL, 98.
162 Rawls, PL, 25.
163 Rawls, PL, 26.
164 Rawls, PL, 48.
165 Rawls, PL, 66.
166 Rawls, PL, 67.
167 Rawls, PL, 67.
168 Rawls, PL, 69.

172 Rawls, PL, 8.
174 O’Neill, 421.
175 Rawls, PL, liii.
179 Rawls, PL, 50.
180 Rawls, PL, 62.
Chapter Five

A New Rationalist Foundation
For Critical Social Theory:
Habermas’ Reassertion of Moral Rationalism
The highest potentialities of mankind lie in the rational union of free individuals.

—Herbert Marcuse

Introduction

Perhaps the most significant contemporary reassertion of a normative social and political theory grounded squarely on reason is that of Jürgen Habermas. This chapter will consider his attempt to develop the idea of a communicative form of rationality that can form the basis of a normatively critical theory of society. Habermas’ theoretical outlook is complex; its articulation, dense; its themes, wide ranging. His work connects to, and assumes a background in, a broad spectrum of empirical and theoretical disciplines ranging over the full spectrum of social sciences including anthropology, sociology, psychology, and political science, as well as linguistics, and of course, an array of traditional philosophical topics analyzed by thinkers on either side of the great analytic/continental fault line which has effectively fractured post-enlightenment philosophy into two relatively isolated disciplines. The centrepiece of Habermas’ whole approach and its various theoretical programs is his moral theory (discourse ethics) and the heart of this moral theory is its rationalist core. It is this contemporary moral rationalism that is most germane to the present investigation. When considering this rationalism, however, a challenge presents itself owing to the fact that, because of its central significance for Habermas’ project, and because of the degree to which it is integrated with each sphere of his analysis, its foundations, details, significance and implications are often scattered throughout his various interlocking theoretical programs and his discussions of numerous topics. Therefore, messy as it may be, it
will be necessary to plot a course that ranges over many dimensions of his work in order to tease out the details of Habermas’ defence of moral reasoning. Obviously, in a work such as this, it would not be possible to summarize all the major strands of Habermas’ work; nevertheless, in this section I will strive to shed some light on how his conception of reason functions within his overall theoretical framework and how it confronts and develops some of the major ideas about practical reasoning discussed in the earlier chapters of this thesis.¹

According to the analysis of first generation Critical Theory, typified by Max Horkheimer and Theodore Adorno, the enlightenment had established instrumental rationality as the fount of true knowledge. As the instrument which allows for the technological mastery and control of nature, the natural sciences came to occupy a privileged position. Scientific understanding of the operations of the physical world could be exploited to bend nature to serve our particular ends. In this sense science/technology was understood as an apparatus of means/ends reasoning. Under this new scientific outlook, human activity could more efficiently and effectively shape the world to its purposes. Influenced by Max Weber, they understood the implementation and institutionalization of this reason, called the “rationalization”¹ of society, to have resulted in industrialization in production, and bureaucratization in governmental and administrative control. According to this analysis, the main current of enlightenment thinking claimed that instrumental

¹ Despite a deep, fundamental continuity, Habermas’ thought does change over time, but there is not time here to analyze the details of this development. The Habermas I present here is essentially the Habermas of the major statements of his mature theory, circa the 1980’s/early 1990’s (e.g. The Theory of Communicative Action and Moral Consciousness and Communicative Action) but may reflect earlier or later insights to the extent that they are, in my understanding, compatible with this articulation.
reason is linked with human freedom since it liberates us from the uncontrolled forces of nature which frustrate our goals.

But, as the history of the 20th century shows, the results of this reason have not been all sunshine and rainbows. The early critical theorists expressed an exceedingly pessimistic view of reason of this kind. Indeed, Horkheimer claims that modern reason is an

Instrument oriented to expediency, cold and sober... When even the dictators of today appeal to reason, they mean that they possess the most tanks. They were rational enough to build them; others should be rational enough to yield to them.²

But, at the same time, insofar as they thought their theories could be critical of current social forms and patterns of thought, they understood themselves to be employing that selfsame, traditional philosophical instrument—human reason in the broadest sense. This aporia is explained in Adorno and Horkheimer’s own words:

The dilemma that faced us in our work proved to be...the self destruction of the Enlightenment. We are wholly convinced...that social freedom is inseparable from enlightened thought. Nevertheless we believe that we have just as clearly recognized that the notion of this very way of thinking, no less than the actual historical forms—the social institutions—with which it is interwoven, already contain the seed of the reversal apparent today.³

For the critical theorists, then, enlightenment reason proved both necessary and impossible.

I think it is fair to say that Habermas’ central concern is with overcoming this roadblock which the early critical theorists hit against. His question, in other words, is: how can we hold on to the baby of enlightenment rationalization (humanitarian ideals and the possibility of legitimate critique) while discarding the murky
enlightenment bathwater (bureaucratic-technocracy, alienation, loss of freedom, loss of meaning, etc. based on a reduction of human beings to strategic calculators)? In short, the solution Habermas proposes ultimately lies in what he sees as a sufficiently sophisticated understanding of reason. Specifically, through a detailed analysis of its various functions, Habermas arrives at a highly differentiated conception of reason. The most basic and most important distinction is the one Habermas draws between *instrumental rationality* and the hitherto neglected *communicative rationality*.

1) Linguistic and Social Theoretical Foundation

1.1) Enlightenment Contradictions

While Habermas broadly endorsed the Frankfurt School critique of Enlightenment reason, he claims their analysis was lopsided—neglecting certain positive outcomes of enlightenment reason. He is explicit in his recognition of what could be called the “Janus face of possibilities”\(^4\) inherent in enlightenment reason. Probably, for Habermas, the most important positive upshot of enlightenment reason was the emergence of a reasoning public which could function as a check on both markets and the power and authority of the state.

With the growing penetration of enlightenment reason into the daily life of the 18\(^{th}\) century European citizenry, one can begin to discern the formation of what Habermas calls the “public sphere.”\(^5\) The public sphere is the space in which public opinion about matters of common concern can be formed through open debate. Such a sphere would include institutions which facilitate face to face communications such as clubs and learned societies, literary salons, coffee houses
and pubs in which the political issues of the day were being discussed, as well as various media used to facilitate such debate, which in the 18th century meant books, personal correspondence through letters, and especially the many small journals and newspapers which began to flourish at that time. This public sphere represents the space between the private institutions of civil society, such as the family or church, and the formal apparatus of the state. It was here that private interest and perspectives could compete in a rational discursive contest where, ideally at least, no one was forced into compliance through promise of reward or threat of punishment nor through any manner of coercion or manipulation; instead, no participant is required or expected to be moved by any force other than that of the better argument—that is to say, the force of reason. The ideas that win out in such debates feed into a developing public opinion about matters of common concern. In other words, the public works towards decisions about what is in the common good. A healthy modern society does not operate on the basis of any law or authority that relies ultimately on force or manipulation to compel obedience. Instead, communicative power is translated into administrative power—the immanent and therefore unforced or autonomous force of reason becomes the intersubjective basis for binding moral norms and the legitimist laws which reinforce and protect them.

Specifically, laws and other government policies and procedures can now be held up to a public scrutiny in which certain questions can now be put—such as, who benefits from existing laws? or more precisely, does a given practice or norm represent or safeguard a genuinely public interest (the interest of those taking part
in the discussion) or does it instead benefit some particular group-interest? If it does not represent a genuinely “generalizable interest,” then from the perspective of civil society it is illegitimate and the practice should be abandoned or the norm should be purged from the order of those held to be legitimate. By “generalizable interest” Habermas means a need that can survive discursive testing. He puts the mater succinctly in _Legitimating Crisis_:

Insofar as norms express ‘generalizable interests,’ they are based on rational consensus...Insofar as norms do not regulate generalizable interests they are based on force [Gewalt]; in the latter context we use the term normative power [Macht].

If governments, or those who wield power or exercise authority, can be made responsive to the discursive will-formation of a reasoning public sphere, then domination and exploitation can be corrected or curtailed and the exercise of authority can be limited by excluding all save its legitimate use in service of the public good. According to Habermas, while it is true that at first this public sphere was practically restricted to an educated leisure class of men, nevertheless, it was through the public sphere that the (then radical) normative idea of the common good started to gain traction.

For Habermas, the public sphere became the lifeblood of modern democracy and ultimately, greater freedom. While both Habermas and Adorno share a certain fundamental agreement with the Kantian notion of autonomy, having individual autonomy—the capacity to use reason and think for oneself (mündigkeit)—was essentially negative for Adorno—in the sense that emancipation can mean nothing other than resisting the dehumanizing capitalist order—whereas for Habermas
there was a positive alternative of creating conditions to foster autonomy. Specifically, the alternative is to build and strengthen 1) a healthy well-functioning public sphere, under which reason guides public will-formation, and 2) democratic institutions that could function to make government and the economy responsive to the will of the people—effectively limiting the exercise of authority to the role of servicing only that which is directed towards the common good. The clear aim here is to foster and protect these rational areas of life.\textsuperscript{10}

1.2) Lifeworld

Civil society is part (the institutional expression\textsuperscript{11}) of a more basic category which Habermas calls the “\textit{lifeworld}”\textsuperscript{12} (a phrase appropriated from Edmund Husserl and influenced to some extent by Wittgenstein’s analysis of “forms of life”). The lifeworld can be contrasted with the “\textit{system}”\textsuperscript{13} which, together, make-up the two basic spheres of social reality. For Habermas, the lifeworld is essentially the communicatively established background knowledge and shared understandings which form the context of social life. It includes all informal and non-marketized aspects of our daily life—encompassing both the voluntary organizations, non-party political associations and media outlets of the public sphere as well as private institutions such as the family.

The lifeworld is the cultural horizon against which action takes place. In one sense it is a stock of resources we draw upon when we attempt to coordinate our action.\textsuperscript{14} Any attempt to coordinate our action—that is, I think it fair to say, any attempt to do or say anything in a social context—will necessarily make certain assumptions about what is in the world, how the world works, what is valuable,
what is fair, etc. Suppose I ask you something as simple as, “would you like to come eat dinner with me at my house?” The question seems to assume: 1) that a house exists which belongs to me and that I have food there, 2) that people eat a meal called dinner in the cultural context in which the utterance is made, 3) that given our relationship and our respective status it is appropriate for me to make such an invitation, etc.

On its own, however, this explanation can be misleading. The lifeworld is not merely a static stock of cultural resources owing to the fact that our ability to act and communicate allow us, on the one hand, to sustain and renew existing elements of the lifeworld or, on the other hand, either to challenge and modify these resources, or create entirely new ones. Since this very process is part of its functioning, the lifeworld itself is, in fact, an exceedingly fluid and dynamic entity. Insofar as our social horizon goes unchallenged it functions as a relatively stable platform for social interaction and this functioning remains more or less hidden.\footnote{15}

But, when background assumptions are questioned, when implicit aspects of the lifeworld become explicitly thematized, disturbances will erupt in interaction. When this happens the lifeworld has a mechanism for repairing the disturbance. This mechanism is reason, specifically communicative reason, and the action based on this reason—that is, what Habermas calls “communicative action.”\footnote{16} And, it is on account of this regulating role of communicative reason that the lifeworld is called “symbolically structured.”\footnote{17}
1.3) Communication and Social Order

Given the fact that “every society has to face the basic problem of coordinating action” Habermas, like countless political philosophers before him, begins with the fundamental question, how is it that human beings are able to coordinate their individual activities with one another? that is, how is social order possible? If human beings are merely instrumentally-rational beings strategizing how to maximize their own benefit by interacting with others in essentially the same manner as with mere things, it is difficult to see how there can be peaceful, stable and productive integration of individuals’ actions. One influential answer to this problem, developed by Hobbes, was that instrumentally rational individuals would, for the sake of peace and security, collectively agree to submit to the rule of a sovereign—a powerful authority with the ability to back its rules with credible threats. A The difficulty, or at least one major difficulty, with this sort of social contract solution to the problem of social order has been called the free-rider problem. B The idea here is that whenever the sovereigns back is turned the rational individual would break any law that he or she could get away with if it is to his or her own advantage. And, of course, the sovereign cannot be watching and prepared to punish everyone at all times. Therefore, social stability could not be based primarily on threats and coercion. 19

---

A Another influential answer is Adam Smith’s idea that, essentially, free markets naturally coordinated the activities of self-interested actors for mutual benefit and public good. I believe this solution is already undercut, at least to some degree, on the basis of arguments forwarded against a closely allied instrumental accounts of practical reason (chapter 2 Section 1.3 and Chapter 3 Section 1)

B See chapter 4 section 4 for another conceptual problem with this approach.
Moreover, if we look at real individuals and real societies most people do not do things primarily because of the threat of sanctions but largely because they have internalized certain shared norms and a shared understanding of the world. The basic idea here is that although social interaction may be filled with bribes, threats, trickery and strategic action of all kinds, any lasting mass-acceptance of a norm must depend, at least in part, on reasons—in other words, the norm must seem justifiable to those it is meant to apply to. Put simply, you will not be able to coerce acceptance of all of the people all of the time. Since social order “remains unstable so long as the moral moment of conscience and obligation—that is to say, the orientation of action to binding values—is missing” it can never be based merely on a “collective instrumentalism.”

Influenced, to some degree, by the alternative represented in the work of sociologists Emile Durkheim and Talcott Parsons, Habermas endorses the idea that social stability is the result of a process of socialization into a shared consciousness. However, the question remains: how is it that human beings come to this shared understanding of the world? Habermas answers that we come to a shared understanding through communicative action and ultimately communicative reason. Habermas claims that all human actions are coordinated primarily through language, which entails a commitment to justify the validity claims raised in our communications with reasons. Through language use it becomes possible that actions are guided and social cohesion is achieved based on a mutual recognition of good reasons.
1.4) System

Before directly considering communicative reason, let me turn for a moment to the second aspect of social reality identified by Habermas—namely, the system. In relatively small and simple societies it is possible for individuals to interact using our ordinary everyday social competencies, particularly our ability to communicate meaningfully, which relies, most fundamentally, on capacities to formulate and be responsive to reasons, recognize this competency in others, and as a result treat others as partners with whom we must work if we are to reach a shared understanding. In a society integrated on this basis it is possible to negotiate a shared understanding of how things are and what it is that we should do. This would result in a social organization that runs according to rules which everyone could recognize and understand the reasons for. Habermas calls this sort of social cohesion, the sort germane to the lifeworld, “social integration.” But, as societies become larger and more complex our communicative competency becomes overburdened, requiring, in addition to social integration, the introduction or expansion of a different type of organization; Habermas calls this “system integration.”

When it becomes necessary for thousands of people to coordinate their activity it is not possible to meet, let alone negotiate every detail of our association, with each one of them. Also, when the projects undertaken by a society become complex, requiring the coming together of many geographically disparate people, resources, and technologies, coordination cannot be achieved on the basis of our ordinary communicative competence alone.
For example, suppose I want to have a party. I can sit down and negotiate the details with my friends. I suggest Friday night works best. Jared thinks Saturday is best because he has plans on Friday, but Emily convinces him to change his plans because she will bring Wendy, who Jared especially wants to see. We ask Rodrigo to bring some of his great snacks but he says “why should I, since I brought the food last time?” Eventually Drew volunteers to bring the food. Suzanne will only come if she can bring David but David does not get along with Chris, so we agree not to invite Chris this time. In the end, everyone is satisfied, all the issues are resolved, and the party goes ahead. If further problems or disagreements crop up during the party these too stand a fair chance of being resolved through further discussion.

But, suppose that I’m just too tired for any parties and so, instead, I decide to entertain myself at home on Friday night. I go to the electronics store to buy a new television. The salesperson does not ask why I want a television, what I am going to do with it, how I got the money for it, etc. So long as I behave within the range of expected behaviour I do not need to explain or justify my choices in any way. Likewise, I do not have to know why the salesperson is selling televisions, where he is from, why he chose this job or what he thinks of my decision to buy a television. Taking the price as my cue, I give the salesperson the money; as expected, the salesperson gives me the TV—end of story. Our interaction is simple and predictable. This simplification is not what defines systems integration but it is an importance consequence of it, and one which accounts for why this form of coordination becomes necessary. In this scenario, the reason behaviour is so predictably, and indeed smoothly, coordinated, without the need for a complicated
negotiation of a shared understanding through communicative action, is due to the fact that our interaction is mediated through a type of mechanism which Habermas calls “language-independent”\textsuperscript{23} or \textit{non-symbolic steering media}—in this case the medium is money—and this is the fundamental characteristic demarcating the systemic domain.\textsuperscript{24}

It is crucial to note, however, that not only is my activity smoothly coordinated with the salesperson but with many thousands of other people, including those who mined and processed the materials, transported those materials, designed and built the various components, designed and produced the machines that built those components, organized the labour and managed the corporations which undertook these projects, and so on almost \textit{ad infinitum}. If then, I wanted to have a party on Friday night, this could be enabled primarily through the social integration of the lifeworld, but there is no way I could rally those same communicative resources in order to watch the Friday night feature on The Movie Network. There is no way I could go around the world and negotiate the participation of all the experts necessary to build the television, not to mention the thousand of artists, technicians, and administrators required to produce the programming, or those scientists and engineers who conceived and built the rockets to launch into orbit the satellites that beam the signal into my home. All this requires the coordinated work of hundreds of thousands of specialists whom I have never met and could never meet.

In a society, in any system, “exchange between system and environment and the exchange among functionally specific units within a system...have to take place
by way of some medium or other." In a lifeworld rational linguistic communication can fill that role, but, functioning as a “relief mechanism” certain societal sub-systems have given that role over to a different sort of media.

Habermas analyzes two particularly important non-symbolic steering media, money, which is the regulating medium of the economic sub-system and power, which is particularly instrumental in regulating the political sub-system. Like money, power serves to simplify our interaction by slotting it into pre-established tracks. Imagine a government or corporate bureaucracy has an objective. The General, Minister, CEO, or whoever it may be, delegates tasks to subordinate department heads, who in turn give orders to lower level managers, and so on down the line. Subordinates need not, and indeed cannot question the overall objectives of the institution or their assigned roll in carrying these out. The authority to command attaches to the various offices of the institution’s structure. The answer to, “Why should I do X?” is simply “Because I am a general and you are a private.” It does not matter if you do not want to do what your manager tells you or if you disagree with them; you are simply expected to do your job, play your assigned roll and allow the system to run smoothly. Social hierarchies of this kind are a very efficient way of organizing to accomplish complex tasks.

In conceiving society as a system, Habermas has been influence by Talcott Parsons, Niklaus Luhmann, and others who have contributed to an aptly named “systems theory” which was in turn inspired by looking, very early, at biological systems, and later at cybernetic systems. First off, it is worth asking, what is meant by “system?” A system is simply a self-sustaining rule-governed structure of
components. A system is distinguished from its environment in that it is made up of a finite number of elements and, in that, only a relatively narrow range of definite interactions between those elements are possible. On this understanding, which Habermas seems to share with systems theorists, many sorts of things qualify as systems, from certain institutions to organisms and machines.

For example, a bicycle functions as a system. It has a limited number of parts which interact with each other in regular ways that constitute the rules for its operation. If we disassembled the bicycle into its commonest parts and threw it into a trash heap it is no longer a system. Its parts are now free to interact with each other in new and potentially infinite ways (e.g. the handlebar with the gears, the seat with the tires) and in new and irregular ways with elements previously external to the system (e.g. the chain with a frying pan, the spokes with a banana peel). If a system is to maintain a boundary between itself and the chaotic external environment, that is, if it is to survive at all, it must perform certain functions—adapt to its surroundings and maintain its internal organization. Eventually a bicycle would break down, rust out, dissolve into its environment and cease to be a system if it not maintained. A healthy maintenance schedule is part of the bicycle system, without which its various sub-systems would deteriorate until they erupted as a system-wide crisis threatening the functionality of the whole bicycle. A neglected bicycle may not remain a system for long, but even a poorly maintained

---

*A Language itself could also count as a system for most straightforward systems theorists but since Habermas distinguishes systems from the meaningful, linguistically mediated structures of the lifeworld his account would not sit comfortably with the idea of language, or any social interactions which are the manifestation of our general capacity to act communicatively, as a system. I will return shortly to the most important ways in which Habermas differs from systems theorists.*
bicycle can still function as a system until broken parts lead to catastrophic functional disruptions at the system level.

Besides accomplishing these necessary tasks, as a system becomes larger and more complex it must cope by undergoing increasing "system differentiation"—that is, it must develop more specialized sub-systems. As the system develops and continues to perform its functions, it maintains its stability and cohesiveness because its component parts and subsystems are integrated and organized by the rules governing the system as a whole.

Sociological systems theorists have suggested that society can be best understood when viewed from this perspective—when viewed as a system. Habermas agrees that insofar as society exhibits system-integration it is indeed a system. However, he diverges from systems theorists such as Luhmann, claiming that, insofar as it is socially integrated, society also forms a lifeworld. The system integration of society, achieved through the meditation of non-symbolic media, is a good thing insofar as it is necessary for large complex societies, but only to the extent that it is subservient to the communicatively-rational aims and interests of the lifeworld. This important caveat is what sets Habermas apart from the systems theory approaches which influenced this dimension of his thinking and it is a significant difference which bears directly on the central insights of his entire view. Therefore, I will try to specify it more clearly below (sections 1.5-1.7).

1.5) Social order: System or Lifeworld?

Recall that for Hobbes society can only be integrated by the coercive force of a central authority—the sovereign. Once this alternative becomes problematic the
best available alternative seems to be that society is bound together by reason. This alternative can be subdivided into two basic forms—society as a machine (instrumental reason), or society as a discourse community (communicative reason). Underlying these forms of integration are the pivotal conceptions of reason alluded to here, and they will be considered in detail shortly.

Habermas argues that ultimately society must be bound together as a discourse community and that systems integration must always be parasitic (not necessarily in an insidious way) on the social integration of society as a lifeworld. However, when systems gain a degree of autonomy and begin to encroach on the communicatively established integration of the lifeworld, deep problems result. Habermas calls this phenomenon the colonization of the lifeworld by the system and it will be considered in more detail in section 1.7.3 At present suffice it to characterise the problem crudely. Insofar as we exhibit systems integration our relationships are governed by rules that regulate the system. People, like the elements within any system, respond in a more or less automatic and predictable way to stimuli. Like gears in a machine we do not relate based on meaning. Systems are not organized through meaningful communication wherein we confront others as communicatively competent individuals like ourselves. Instead of genuine subject/subject (ego/alter) relationships we relate to others as we would any other physical component of the system or our environment, that is, in terms of a subject/object relation. Unchecked, this instrumentalism inherent in the very nature of systems can be dangerous—it can push out and take over the space for meaningful social interaction.
There is overlap between what gets covered on this system/lifeworld account and what gets covered on a Rawlsian-style liberal account of the major social institutions which make up the basic structure of society. In *A Theory of Justice* Rawls understands an institution as a public system of rules. Many commentators on Rawls have not focused on the fact that the intended application of his principles of Justice is to institutions. Karen Wendling—who has (along with Paul Viminitz) focused on and developed a Rawlsian analysis of institutions and institutional justice/injustice—offers the following elaborated definition which I see as consistent with the Rawlsian view: “an institution is a public system of rules that creates regularities in social behaviour, specifies behaviour in recurrent situations and is either self-policed or policed by an external authority” adding, “however and by whomever institutional rules are created, though, their function is the same: to guide the behaviour of the members of the institution.”

The liberal definition of institutions encompasses both social systems, such as government and the economy, as well as aspects of the lifeworld such as language itself. Wendling’s elaborated definition makes this fact more explicit. Institutions which rely on “external authority” can be thought of as those governed primarily by non-symbolic steering media while institutions that are “self-policed” would fall more in-line with institutions governed by communicative interaction. While both types or aspects of institutions function to guide behaviour there is a fundamental difference between, on the one hand, institutions that are guided by rules that we have a hand in shaping and can understand the reason for and assent to or dissent from on the basis of free, rational, autonomous understanding and, on the other
hand, institutions which are externally or heteronomously imposed through force uncoupled from rational oversight.

There is wide agreement, at minimum between radical/left-leaning liberals and liberal-inspired Marxian socialists, that institutions are a primary site of injustice and inequality and it is therefore a paramount role of normative theory to justify these institutions or criticise them when they institutionalise injustice. What the lifeworld/system analysis has to add to a liberal account of institutions is an account of how we can criticise irrationality and injustice in institutions: namely, by making non-symbolically integrated systems ultimately accountable or responsive to the reasoned guidance of lifeworld control. This claim is elaborated in the following subsection.

1.6) Communicative vs. Strategic Action

In order to appreciate more precisely how systems can be useful and in what ways they are dangerous I must first pause to consider the sort of action that is associated with either sphere. There are two fundamental types of action according to Habermas—communicative action (associated with the lifeworld) and instrumental action (associated with systems). When we act communicatively we understand other’s behaviours as actions which are done for some reason. In coordinating our behaviour through communicative action we must consider the reasons behind other’s actions, adduce reasons to justify our own actions and negotiate an understanding of our common situation. The goal of communicative action is understanding.\(^{38}\)
Unwilling or unable to engage in lengthy and often futile discussions aimed at justifying their decisions, many parents often present that familiar, thinly disguised pseudo-justification—“just because.” However, if it is difficult for adults to remember being on the receiving end of the “just because,” take another mundane but hopefully familiar example: often times, after carefully explaining a company’s mistake to a stream of corporate representatives, usually over the phone, we are met with a canned non sequitur such as “I’m sorry sir/ma’am but that’s just our policy.” There seems to be a special sense of injustice, a special indignity and anger reserved in the human heart for situations in which we are confronted with people’s outright refusal to offer any justification for seemingly unjust policies or decisions. When concerned parties question decisions, they expect their concerns to be taken seriously and they expect explanations. This expectation of rational justification is at the heart of communicative action and it is why, in all action oriented toward understanding, rational explanations are due.

By contrast, when we act instrumentally we directly manipulate the world for our own purposes. The goal is not mutual understanding but success in terms of our own aims. Habermas argues this type of action has an important role to play in human life. It has been of fundamental importance to the natural sciences and has allowed us to gain increased mastery over the forces of nature. However, systems can encourage us to act towards other subjects in a way that is similar to instrumental action. This social form of purposive action is called “strategic action” and it can become deeply problematic. Before I can get to that problem
some clarification is in order. An explanation of Figure-1 should help to clear up some terminology and the relationship between the concepts employed here.

What Habermas calls *purposive*, or sometimes *teleological*, action is, quite generally, any action directed by “the rationality of means to ends.” It is action that aims at the successful realisation of a given goal through causal intervention in the world. “Teleological action can be judged under the aspect of effectiveness. The rules of action embody technically and strategically usable knowledge.” Habermas puts it more succinctly by saying “the model of purposive-rational action takes as its point of departure the view that the actor is primarily oriented towards attaining an end.” The end or “success” in the context of purposive action “is defined as the occurrence in the world of a desired state, which can, in a given situation, be casually effected by goal-oriented action or omission.”

Fig. 1

Any action oriented towards success can be either social or non-social. Non-social purposive action is called instrumental action. Social, purposive action, on the other hand, is called strategic action. In other words purposive/teleological

---

A Of course, Habermas is aware that in other philosophical contexts (which indecently includes earlier parts of my present investigation) as well as common English speech, “instrumental” and “purposive” are often taken to be more or less synonymous, and that he is therefore simply stipulating technical meanings for purposes local to his theoretical framework.
action is a category that includes both instrumental and strategic action.\textsuperscript{48}

According to this scheme, if we were to speak of strategic action as instrumentalizing or objectifying we would be speaking only metaphorically or analogically. Strictly speaking no social action is instrumental.

We call an action oriented towards success instrumental when we consider it under the aspect of following technical rules of action and assess the degree of efficiency of an intervention into a complex of circumstances and events. We call an action oriented to success strategic when we consider it under the aspect of following rules of rational choice and assess the degree of efficiency of its influencing the decision of a rational counterpart in action. Instrumental actions can be connected with social interactions; strategic actions are themselves social actions.\textsuperscript{49}

Actions taken towards others—social actions—are of two fundamental types. Social action can either be strategic or what Habermas calls communicative action.\textsuperscript{50}

Communicative action can...be distinguished from strategic action in the following respect: the successful coordination of action does not rely on the purposive rationality of the respective individual plans of action but rather on the rationally motivating power of feats of reaching understanding, that is, on a rationality that manifests itself in the conditions for a rationally motivated agreement.\textsuperscript{51}

At a broader level all action in general can be divided into two basic and irreducible categories or what Habermas calls "elementary types of action"\textsuperscript{52}—purposive action and communicative action. Communicative action is qualitatively different from all instrumental action (the specifically non-social version of purposive activity\textsuperscript{53}) as well as purposive forms of social interaction (strategic action). The key question is: what sets communicative action apart from all other action?
It cannot be simply that it is social or that it is mediated by language because these things are true of strategic action as well. In fact, the grounds for the distinction are further complicated by the fact, explicitly recognised by Habermas, that, in some sense, all action is purposive. He claims that “all actions, linguistic or non-linguistic ones, can be conceived as goal oriented activity.”\(^{54}\) On the face of it this may seem like a strange concession. How is this compatible with his insistence that communicative and strategic action are incompatible and mutually exclusive,\(^ {55}\) and the claim that “acts of reaching understanding...cannot themselves be reduced to teleological actions”?\(^ {56}\) The basis of the distinction lies in the fact that there:

is a peculiar rationality, inherent not in language as such but in the communicative use of linguistic expressions, that can be reduced neither to the epistemic rationality of knowledge (as classical truth-conditional semantics supposes) nor to the purposive-rationality of action (as intentionalist semantics assumes).\(^ {57}\)

Habermas explains that “like all action, communicative action is purposive. But here, the teleology of the individual action plan is interrupted by the action-coordinating mechanism of reaching understanding.”\(^ {58}\) Understanding the nature of this “interruption” is the key to this fundamental distinction. Habermas crystallises the point in a different passage:

in communicative action, the structure of the use of language oriented towards reaching understanding is superimposed on the fundamental teleological structure of action and subjects the actors to precisely such constraints as compel them to adopt ...an attitude that is more laden with presuppositions than is the objectivating attitude of the strategic actor.\(^ {59}\)

What distinguishes communicative action then is the fact that action is subject to certain constraints inherent in the structure of language. Action subject to these constraints is communicative action—it is no longer merely
purposive/teleological. To use language in order to reach an understanding is fundamentally different and incompatible with acting merely in order to casually affect a desired change. Strategic action, though it makes use of language, does so in a way that is purely “parasitic” on the “original mode” of language use which is an orientation toward reaching a rationally motivated understanding. All this requires a great deal of argument and elaboration, but at this stage my primary interests is in laying out some basic terminology, distinctions and relationships.

The main point I want to return to here is that when we act strategically we treat action (which is motivated by particular reasons) as mere behaviour. We simply consider how others respond, what is the expected behavioural output to given input stimuli. We can then adjust our action to produce the desired response from others. Action sheds meaning and functions merely as stimuli. When I say “one television please” It does not matter if the clerk understands me as long as I get what I want. A television vending machine may do just as well. Although not the clearest example of strategic action this is a particularly salient example. Let me explain.

In the case of “Manifestly strategic action” validity claims are replaced with power claims. If I am holding a Gun and say “give me your wallet” I am using language to make a threat; I am pointing to a means of influence external to language. What makes the claim acceptable is not the reasons for you to accept its normative validity but the sanction conditions attached to it. If I were to attempt to rob you of your wallet through trickery, by saying, for example, “The authorities are coming to my house and if I cannot produce some valid papers and show them that I
have some money I will be deported. Please lend me your wallet for fifteen minutes.” this would be a case of “Latently strategic action.” Here I am only pretending to use language as a means of reaching understanding. Latently strategic action can only succeed so long as it can maintain this pretence and exploit the natural attitude of language users to have an orientation towards reaching understanding. As soon as we realise that someone is simply saying things to bring about their desired state their action cannot succeed.

Unlike in these clearer examples, in the interaction with the electronics store clerk it may not be the case that anyone directly involved is trying to threaten or coerce the other, nevertheless the interaction may exhibit a purposiveness that is symptomatic of the prevalence of unchecked overarching social systems that can tend to funnel our activities into narrow, prescribed channels of purposive activity that shrink the space available for genuine communicative action. Systems rely on common recognized cues that organize action and link us together by easy to manage rules without the need for communicative action. One never has a wide range of options since action is always more or less automatically determined by the rules of the system. Consider friendship. Friendship remains a relationship primarily established on the basis of communicative action. When I have an interaction with a friend we are both free to act in ways that makes sense to each of us. We are under no (or minimal) obligation of acting a certain way because it is the only option according to the rules of some system, whether it be political, economic or otherwise. Our actions are primarily guided by the vastly richer and more complex rules of the lifeworld. If I suggest a description of a situation or prescribe a
course of action, my friend is under no obligation to accept it if it does not make sense to him or her. I cannot force a friend to accept my proposal; I can only present my reasons for them. My friend will only accept my proposal when or if I manage to persuade him or her with my arguments; when my proposal comes to makes sense for them as it does for me; when my reasons become theirs.

Now consider a formal teacher student relationship. Suppose I forward a description of a situation or prescribe a course of action to one of my students in a university course. Our relationship as it exists within a formal academic institution is mediated by a non-symbolic steering medium, namely, power. I may try to run my classroom according to the ideal of a free and open dialogue wherein each of us will endeavour to understand the other and defend our claims only with reasons. However, at the end of the day, I am the one who marks the exam. As a function of the institutional position I hold, I have the power to decide precisely how correct or incorrect a student’s interpretation of the material is and assign a precise quantifiable value to reflect my judgment. I have the power to give them a C+ and keep them out of law school or whatever it may happen to be. I know I have this power, the student knows I have this power, and all our interactions will be tainted by this fact. If my friend thinks my interpretation is stupid he is free to tell me so, my student, understanding my power and notwithstanding my intentions to foster openness, would never be likely to feel quite so free.

Notice that it is not necessary for me to “abuse” my power in order for this dynamic to exist. Personally, I may not throw my power around. I may be the most reasonable person you can imagine—always responding with reasons, never
blocking discussion with a “because I say so.” But institutionally, systematically—
power is attached to me by virtue of my position in a power structure regardless of
how I act personally. Despite the fact that a healthy teacher/student relationship is
a paradigmatic example of a relationship that we would expect to be firmly planted
in the fertile communicatively-rational soil of the lifeworld, it is important to see
how, even here, social systems can have a distorting and dangerous effect if not
controlled.

1.7) Colonization of the Lifeworld

The problem with systems integration comes when the various social sub-
systems, each of which operates according to its own rules, in other words, each of
which has developed its own internal logic, begins to develop a life of its own, a life
independent of, and indeed indifferent to, human needs. When this happens
fundamental aspects of social reality confront human beings as alien forces even
though it is us who creates and sustains them. They become like forces of nature.63

When a differential in electric charge builds up between the earth and a
cloud this charge is equalized in the form of an electrical discharge—lightning.
Although it makes a big difference to you, the lightning is indifferent as to whether it
strikes a rock, a tree, or you. Lightning does not care about your interests, it simply
behaves according to the physical laws of electrical conductivity. Once social sub-
systems are decoupled from the communicatively directable aims of the lifeworld
they too begin to follow their own laws, indifferent to human interests. To take
what may be the most important example, the capitalist markets that form the
contemporary global economic system operate according to the principle of
increasing capital—turning money into more money for investors. Evident here is how, as Simone Chambers puts it, “economic rationality has a tendency to overwhelm and crowd out other rationalities.” Unchecked and unregulated, this market imperative leads to more and more intensified production and consumption which, it now seems clear, is driving us towards a depletion of resources and degradation of the environment which will have catastrophic consequences for human civilization.

If, despite sensible precautions, someone is exploded by a lightning bolt we might reasonably say “we do not control the forces of nature and although this was a tragedy it was not one that could reasonably have been avoided.” But if the social assistance system is set up so that you can only pick up your check in person yet you cannot get out of bed because you are sick from cancer treatment, it would be strange if we had to say “that is the way the system works so you will just have to starve.” Likewise, if the entire planet were to be virtually ruined for human habitation by the economic sub-system it would be strange if we had to say “well, that’s a tragedy but it was unavoidable because we do not control the forces of the free market.” It would be strange because social systems, unlike natural forces, are things that are created and sustained by human activity, supposedly because they help us organize and coordinate ourselves to meet common objectives. Once established, however, social sub-systems tend to exhibit a kind of inertia, operating according to their own quasi-natural principles. Like the bolt of lightning they come to operate according to laws that are independent of communicatively established rational guidance.
For Habermas the only legitimate power derives ultimately from the public will, which is formed through the rational communicative processes of a healthy lifeworld. The greatest threat to this democratic potential is the colonization of the lifeworld—the takeover of action coordination from normatively regulated discourse by the illegitimate use of non-symbolic steering media such as money and power. Once we let go of, or significantly restrict, major spheres of social life in which we might act communicatively by handing over action to the simplified imperatives of systems which only allow us to act strategically it becomes increasingly difficult to subject the various sub-systems to rational/critical scrutiny and debate. Colonization of the lifeworld can be said to have taken place in proportion to the degree to which systems have gone significantly off the rails of communicative reason and genuine public interest and are coasting freely and uncontrolled according to their own logic—typically a one-sided logic of competitive individualism. When any system, instead of serving the human interests for which it was intended, starts to function more or less independently as though it were an end in itself, when, for instance, the profit maximization of the economic system is indifferent to human needs and becomes like a natural force which goes unchecked by any public reasoning about what is in the general interest, the lifeworld has become colonized by that system.

Once colonized the lifeworld is stunted and impoverished. Systems steer individuals towards ends not related to understanding or consensus, imposing external (in Kant’s language heteronomous) constraints on action. This tends to produce serious problems for the individuals living under such conditions. When
humans primarily act purposively, responding automatically to the demands of systems, there is a seriously diminished role for human freedom, creativity, and responsibility. In the absence of the rich and meaningful interaction of a healthy lifeworld, human action, indeed human life itself loses its meaning. If we do things just because it is functional from the perspective of a system, we become like automatons, our very life activity becomes a cog in social machine. At some level, however, things need to have meaning for us. We need to have the experience of doing things, ultimately, for reasons that make sense to us.

Of course, for Habermas, no matter how eroded the lifeworld becomes it cannot be erased because human beings are essentially communicatively rational beings. It goes against our nature to be pure machines, thus society can never be pure system. If systems themselves are to survive at all they must always be parasitic on some semblance of a meaningful lifeworld. Rampant colonization ultimately tends to make society malfunction even as a system by eroding its base—the lifeworld. This creates what Habermas calls social pathologies, which can eventually become civilization-threatening crises. These pathologies include a felt decrease in any sense of shared meaning/understanding (anomie), feelings of helplessness and lack of belonging (alienation), an unwillingness to take responsibility for individual and collective actions and for current social structures and conditions (demoralization), and eventually even the erosion of social bonds (disintegration) and the breakdown of order (social instability). This is where Habermas departs radically from straight systems theory.
1.8) Beyond Instrumental Reason

So far, I have discussed Habermas' two fundamental social spheres (lifeworld/system) and how they are related to two fundamental types of human action (communicative and instrumental/strategic), but only alluded to the fact that these types of action are underpinned by two even more fundamental categories (instrumental rationality and communicative rationality). Instrumental action—action intended to manipulate the physical world by intervening in casual relationships between things in order to bring about one's desired results—is governed by, and the outcome of instrumental reason, the reasoning that analyzes causal relationships and determines the most efficient, effective and appropriate means to given ends. The basic details of instrumental reason have been outlined in previous chapters. Since Habermas' understanding does not differ on these basics it is not necessary to say much about it here apart from contrasting it with communicative reason.

It is important, however, to emphasize the point that Habermas does not reject this type of reason or propose communicative reason as an alternative to it. Unlike earlier critical theorists, such as Herbert Marcuse, who were somewhat sceptical of the instrumental attitude towards the world—full stop—Habermas thinks that in order for us to get along in the world and for the species to reproduce itself materially it is necessary to understand and instrumentally intervene in the causal relationships of the world. In this way Habermas, like Rawls, takes on board insights issuing from the Humean side of enlightenment reason. The refinement of instrumental endeavours like science/technology and institutional organization, can
allows us to more effectively realize our interests and satisfy our desires, which is a good thing. The problem comes only when we reduce all rationality to instrumental rationality and when we illegitimately take an instrumental attitude towards other persons (strategic action/reason) by seeking to manipulate them as we would a mere thing (now the influence of the Kantian side of enlightenment reason is on full display). The aim here is to develop an overall account of rationality which encompasses instrumental reason in its proper place while moving beyond it. Systems, at first blush, seem to have presented a “can’t live with them can’t live without them” sort of dilemma. The solution for Habermas is that we can live with them so long as they are kept in their proper place. If the systems’ instrumental aims are decoupled from communicative reason or begin to steer us towards a purposive stance vis-a-vis other social actors they are out of place. The specific function of communicative reason within this account now requires more significant explication.

1.9) Communicative Reason and the Theory of Meaning

Communicative action—the meaningful interaction between persons—is the expression of communicative reason. Communicative reasoning is a dialogic, collective activity underwritten by a basic, and in principle universal, communicative capacity to engage in rational discussion. Communicative reason is the mechanism or capacity which allows social order to rest on shared meaning and understanding instead of the coercion of a sovereign, the unquestioned internalization of system-level imperatives, or the uncritical acceptance of the inherited tradition of the current lifeword background. According to this view:
Everyday communication makes possible a kind of understanding that is based on claims to validity and thus furnishes the only real alternative to exerting influence on one another in more or less coercive ways.68

In order, therefore, to understand communicative reason and the alternative theory of socio-political organization that stems from it one must consider Habermas’ theory of meaning and understanding.

Influenced by William James, John Dewey, Karl-Otto Apel and others, Habermas develops a pragmatic theory of meaning. The standard theory of meaning held that the meaning of a sentence was its truth conditions—one understood a sentence once they grasped what conditions would make it true or false. On this view, “meaning” strictly concerns objective relations between words and states of affairs. When the relation between the words in a sentence correspond to the relation between things in the world, the propositional content of the sentence is true—it states a fact. If it fails to do so the proposition is false. The likely result of such a view is that moral statements turn out to be meaningless. We can understand the meaning of statements such as “the apple is red” or “the apple’s mass is 700 grams” because we know what their truth conditions are. If we look at the apple and see that it is green or measure its mass to be 750 grams we know the propositions contained in the sentences are false because the condition for their truth (the apple exhibiting the property redness or being 700 grams) do not obtain. But, with sentences such as “murder is immoral” we cannot objectively look at or measure murdering to see if it possesses some natural property of immoralness. At most such sentences would be expressions of our personal evaluations. "Murder is
immoral” would essentially be saying something akin to “I do not approve of murder” or “I do not like it when people murder.”

In fact, on this account, much, if not most, of the language we use turns out to be meaningless. If we write a poem, for example, we might be disclosing our personal experiences rather than stating facts about the world. When we ask questions or make requests we are not stating facts about the world either. For example sentences such as “How are you today?” or “Please pass me the sugar” would fall outside the purview of this account, leaving them to be classified, strictly speaking, as nonsense.

Pragmatic theories, by contrast, emphasize the fact that meaningful language does many things besides fact-stating. While denoting the way the world is clearly comprises one of its functions (cognitive function) there are others (e.g. appeal function, expressive function). So, while it may be legitimate to apply the truth-conditional theory of meaning to language that is primarily intended as descriptive, it cannot be applied to a wide array of everyday communication we take to be meaningful. Thus, if, instead of reducing sentences to their propositional content (what they claim about states of affairs in the world), we take the utterance itself (the actual words used by a particular speaker to a particular hearer in a particular situation for particular reasons) to be the basic unit of meaning, as Habermas does, we now require a new account of the function of language and of both meaning and understanding. Habermas argues the primary, or most basic function of language is not to state facts but to make oneself understood and to allow interlocutors to reach a shared understanding—one which can serve as a basis for coordinating action.69
Making claims about facts in the world is merely a genus of this larger species of what all language does. The meaning of sentences, according to Habermas, does not depend simply on truth conditions but rather, more generally, on reasons. “We understand the meaning of a speech act when we know what would make it acceptable.”

Reasons are the support which can be given in order to justify the validity claims raised by all meaningful communications. “Ultimately, there is only one criterion by which beliefs can be judged valid, and that is that they are based on agreement reached by argumentation.” This is what Habermas calls the validity basis of meaning and it expresses the rationalist thesis at the core of his philosophy.

According to Habermas there are three different types of criticisable validity claims raised by all speech, where “criticisable validity claims” can be taken to represent “knowledge that can be argued about on the basis of reasons.” First there is the familiar propositional claim to truth. Second, there is the claim that we are being honest or sincere—that we are actually saying what we think/feel/believe. This is called the claim to truthfulness. Third, there is the claim that what we are saying is morally acceptable. This is called the claim to rightness. Specific claims to validity concern the relationship between the utterance and one of three “worlds”; in the case of truth, the objective world; in the case of truthfulness, the subjective world of the speakers own experience; and in the case of rightness, the social world of legitimately regulated interpersonal relationships. The different validity claims are explicitly thematized in different sorts of language use. For example if one were making an assertion the most obviously relevant validity
claim would be the claim to truth. When one is making a request, the claim to rightness seems to become most important. If one were making a declaration about one’s experience the claim to truthfulness takes center stage. Nevertheless, even if certain validity claims are not explicitly thematized, any genuine speech act makes all three claims and can therefore be challenged on all three fronts.

Take the earlier example of a simple everyday sort of sentence: “Would you like to come eat dinner with me at my house?” The question claims the truth of various facts about the world (e.g. that a house exists which belongs to me and that I have food there etc.). Implicit in the speech act there is also a claim to truthfulness—a claim that the purpose of my speech is what it appears to be (to offer you a shared meal) and that I am not lying to conceal some hidden alternate purpose, such as tricking you into following me so that I can rob you. It also makes a claim to the moral rightness of various norms underlying the statement. In this case one such norm is the idea that given our relationship and our respective status it is appropriate for me to make such an overture.

One can object to a speech act by questioning any of its validity claims. When I say “would you like to come eat dinner with me at my house?” you could say I have given a misleading picture of the facts and object by saying “First off, you do not have a house, you live in your parent’s basement and, what’s more, you don’t even know how to cook.” You could question my claim to truthfulness by saying, “You have never asked me over before, yet, now that I’m on the committee that will determine your promotion you invite me for dinner. You do not really want to be friends; clearly you are just trying to butter me up.” Finally, you could accept the
validity claims to truth and truthfulness but question the rightness of the statement saying “I am a stranger to you and you have no business assuming such a familiar attitude towards me. Moreover, I am your elder. As such, only I could initiate such a casual relationship. It is an act of disrespect for you to have done so.”

Part of what reinforced the truth conditional theory of meaning was the worry that, outside the model of the strict one-to-one correspondence of language to observable states of affairs, language, and any claim to truth or validity contained within, becomes groundless—hopelessly subjective—where things can be true for you and not for me and there would be no hope for reconciling these ultimately fundamental differences. However, thanks to an expanded conception of validity tied to the notion of communicative reason, abandoning the objective truth conditional model need not make meaning purely subjective—simply intersubjective; but how so?

Truth is a species of validity but it is not all there is to validity. Sentences are not simply true or false. They are also truthful or untruthful, right or wrong. If correct, this shatters the notion that all validity claims must be supported in the way truth claims are supported, by a symmetry between the world and the language which represents it. But, if it is possible to make validity claims at all—claims we expect others to recognize—they must be supported by something. That something is communicative reason. Habermas argues that what makes a validity claim valid is the fact that it is amenable to rationally motivated consensus, that is, it can gain the assent of participants in a properly prosecuted discourse (what constitutes a
properly prosecuted discourses is another question, one which we will be addressed in the next section).

In summary, for Habermas, the meaning of language is its function. Its function is to establish consensus. Language does this by making validity claims which can be supported, if necessary, by giving reasons. Whether we explicitly call it to mind or not, any time we use language we are always already making all three types of validity claims. Through this language use, our everyday agreements, and to a significant degree the social coordination of our individual actions, rests on “shared propositional knowledge” (in the case of truth), “mutual trust” (in the case of truthfulness), and “normative accord” (in the case of rightness). If we are to engage in meaningful communication we must presuppose, and give others to believe, that we are saying what is truthful, right and true. When we question any of the validity claims raised in communication disruptions to regular action and communication occur because we are no longer operating on the basis of a shared understanding.

Seeing as Habermas claims that social order rests on shared understanding, it is critical for his argument that these disruptions can be repaired, without resorting to extra-linguistic means. This is accomplished by moving to a higher level, reflective form of communication which Habermas calls Discourse. To grasp what is involved in the practice of reasoning communicatively one must understand what discourse is.
1.10) Discourse

In what we might call a regular situation, where the validity claims raised by our speech and action are accepted by others, we already share a common understanding of the situation and our interaction can proceed smoothly on the basis of this common interpretation. If, however, our claims are challenged, that is, if others reject our claims, they are demanding that we make good on our claims and we are required to defend them by giving reasons. When the validity claims that tacitly coordinate our action are challenged we move into a discourse situation, a form of speech which is reflective in the sense that participants are engaged in a communication about communication, a communication with the aim of achieving a rationally motivated consensus about the contested validity claims which cropped up in the course of our ordinary social interaction.\(^{76}\) Discourse is not meant to be merely an abstract academic exercise—a domain of experts removed from the goings on of practical everyday life. On the contrary, it is meant to be a common part of the functioning of the lifeworld. In fact, Habermas takes it to be a matter of empirical fact that discourse is the default mechanism for regulating conflict and establishing consensus in contemporary societies.

Influenced by Weber’s separation of modern life into three distinct value spheres, Habermas argues there are three different types of discourse corresponding to the three different types of validity claims and their respective domains of knowledge.\(^ {77}\) First, for validity claims to truth which fall within the natural science domain of knowledge, there is theoretical discourse; second, for claims to rightness within the sphere of morality and law, there is moral discourse;
finally, for claims to truthfulness within the sphere of the arts, there is aesthetic discourse.

Truthfulness is somewhat different, however, in the sense that no amount of argument can establish that someone is being truthful. Although validity claims to truthfulness are not redeemable discursively they are similar to discursively redeemable claims in that they bind the speaker to certain obligations, not to back the claim with supporting arguments if challenged, but to back it with consistent behaviour. In this way, validity claims to truthfulness, like all validity claims, create a “binding/bonding effect between speaker and hearer.” As I understand it, aesthetic discourse—because it concerns issues of taste rather than truth or justice and deals with evaluative questions from a particular group or individual perspective—must be of a significantly different kind in that its foci are “subject to rational discussion only within the unproblematic horizon of a concrete historical form of life or the conduct of an individual life” and are not universal as in, for example, the case of claims to rightness which are decided rationally “in terms of justice or the generalizability of interests.”

---

A If we suspect someone of being untruthful adducing further claims such as “I am an honest person” or “really I am telling the truth” cannot help their case because we would have the same reason to doubt these assurances. Incidentally I think it is often the case that the more emphatic the assurances the more we are likely to doubt the original claim. I recently saw a book at a rummage sale. The title, which looked like social philosophy, piqued my interest. I may have investigated it further but below it the cover read “honestly this is not a self help book” or something to that effect. I quickly passed over it assuming, almost certainly correctly, that it was a self help book, for nothing other than a self help book would have any reason to so adamantly and explicitly deny this fact on its cover. The only way to redeem claims to truthfulness is through a demonstration of their consistence with past actions and through continued interaction which maintains a consistency with that claim. In this case, if I were to investigate the book, the further I read through it without finding anything self-helpish the more the claim to truthfulness on the cover would be supported.
It is critical at this point to bear in mind that although each type of validity claim is explicitly thematized in the sort of language use associated with its respective sphere of knowledge, Habermas recognizes that life and language do not fall wholly apart into distinct spheres. Recall that all language makes all three claims. For example, discussion within the domain of natural science will involve claims to rightness and truthfulness.

Despite this sensitivity Habermas has been widely criticized for forcing these categories to line up too neatly. While it seems plausible that natural science is mostly concerned with truth and that morality is mostly concerned with rightness it seems doubtful that the most important thing about all art is that it should be truthful. The charge has sometimes been levelled against German philosophy in general that it exhibits a certain penchant for a tidy theory, and Habermas is sometimes thought to exemplify this trend by trying to force these categories into line just to make the theory fit together neatly. Even if everything does not fit quite so neatly, as I suspect it does not, the point I want to stress is that Habermas’ main argument, that there are different sorts of procedures for resolving the different sorts of problematic validity claims, is unaffected. It may, for example, simply be the case that there is no special sphere of knowledge related to truthfulness.

The inherent goal of discourse, of any kind, is what Habermas calls “rationales Einverständnis,” an agreement, a mutual unconstrained understanding, a rationally motivated consensus. The only relevant force here is the force of reason. “In discourse what is called the force of the better argument is wholly unforced...Conviction changes internally via a process of rationally motivated
attitude change." In order to ensure this outcome a discourse must be conducted according to certain rules. Habermas calls these the *idealizing pragmatic presuppositions* of discourse. They are idealizing in the sense that they move participants towards an ideal—in this case toward the “ideal speech situation” which allows for the realization of the more fundamental ideal of a consensus that is purely rationally motivated. A perfectly ideal speech situation is clearly counterfactual; real discourse will only ever approximate more or less closely to it. Yet, if clear and undistorted rational communication is the aim of discourse we must be able to have an idea of what that is. The ideal rules of discourse sketch a picture of this aim.

There are three levels of rules. First, there are logical and semantic rules such as the principle of consistency or non-contradiction. Second, there are norms governing the procedure. These would include the requirement that in discourse one must believe what one says (sincerity) and one must justify one’s assertions (accountability). Finally, there are rules designed to preclude the possibility of coercion, repression, inequality, or any other factor which would taint the outcome by allowing it to become dependent on anything other than the force of the better argument; rules such as—every competent speaker/actor can take part in a discourse, anyone can introduce or question any claim, everyone can express their

---

A At the time of *Moral Consciousness and Communicative Action* (1983), subtle refinements of his theory lead Habermas away from the phrase “ideal speech situation,” preferring the notion of an “unrestricted communication community.” While there may be good reason for this, for present purposes, I believe I can safely avoid unnecessary complication by using the better known phrase “ideal speech situation” since the underlying idea is essentially the same.
desires and attitudes and no one can be prevented through force, coercion or otherwise from exercising these freedoms.\textsuperscript{85}

\textbf{1.11) Transcendental Argument and Reconstructive Science}

It is clear that the rules of discourse are meant to specify a procedure which can be used to rationally test our claims and ultimately provide warrant or legitimation for our norms, beliefs, and actions. Successful discourse yields a consensus on the basis of reasons acceptable to all, since, if the rules of discourse are followed, the voices of all concerned parties are heard, no one is arbitrarily excluded and only the force of the better argument prevails. The issue remains, however, if rational justification is the outcome of a communicative process, how does one justify those rules which are meant to govern the process? How does reason justify its own principles? This is one of the central issues for Habermas and in some form it is an issue faced by any rationalist ethic.

In short, Habermas' answer is that, as \textit{pragmatic presuppositions}, rules of discourse are necessary and unavoidable. Habermas' justification of the rules of discourse is a species of transcendental argument, in the sense that he argues: these rules are preconditions which must obtain if discourse is to take place at all. He is reasoning from the very fact that we enter discourse— or the “fact of reason”\textsuperscript{86}— to the condition that must exist in order for this to be the case. Habermas claims that the attempt to explicitly deny any of the idealizing pragmatic presuppositions of discourse will result in performative self-contradiction. This is a Kantian intuition reinterpreted from a pragmatic point of view.\textsuperscript{87}
Early on, Habermas called his approach quasi-transcendental. It was explicitly distinguished, via the “quasi” prefix, from Kantian transcendentalism in that it was not meant to offer up an *a priori* philosophical argument designed to establish the absolutely necessary constitution of the human mind which would make any and all knowledge possible. Rather, somewhat more modestly, he was making arguments designed to establish that, given certain observable facts about the species, we can conclude that certain human interests and capacities—certain aspects of human nature—which are the product of contingent natural and social evolution—are, none the less, universal and “anthropologically deep seated.”

Although I think his project can still be meaningfully called transcendental in some sense, he later distances himself further by applying less and less frequently the descriptor “quasi-transcendental” in favour of the term “reconstructive science.” Reconstructions are still transcendental in what Habermas calls a “weak” sense. Reconstructive science is meant to describe an enquiry into our unconscious or intuitive capacities or abilities. It is possible, and often the case that we have know-how without knowing or being able to explain exactly what it is that we are doing or how it is that we are able to do it. Reconstructive-sciences simply seeks to reconstruct or explain our abilities, turning know-how into know-that. In particular, Habermas is obviously concerned with the universal capacities associated with our linguistic competencies. It is not transcendental in that it is not metaphysical or even *a priori*. Here philosophy cannot, prior to, and independent of, disciplines within other spheres of knowledge such as the sciences, establish with deductive certainty once and for all the conditions for the possibility of all knowledge. He
envisions a relationship of mutual fit—a hermeneutic circle—between theoretical reconstruction and scientific facts.90

Without the certainty promised by transcendental argument it may seem as though there could be no knowledge of a categorical duty, justice, or morality, and therefore nothing that runs deeper than the concrete and contingent ethical mores or traditions of a particular group or culture. What Habermas proposes is to deduce fundamental moral principles, not from first principles, but from the necessary, that is non-substitutable, mental operations always already performed by communicative subjects.91 His reconstructive approach seeks to explain the basis for our communicative action. These explanations themselves have a merely hypothetical status. Reconstructions are fallible, not foundational. They are subject to confirmation, refutation and revision in light of ongoing testing and debate. Section 2 considers in detail how Habermas employs this strategy to justify a normative theory. In a nutshell, the argument is that we cannot even enter communication unless we assume that something like an ideal speech situation is at least approximated. To put the intuition in the simplest terms: If you go into a conversation already thinking the other person is completely insane, or that they are going to threaten you if you challenge what they say, or that they are lying or trying to trick you, or that you will not be allowed to say what you think, then you cannot expect to have a real communication with the aim of reaching an understanding about something. You might decide to humour the other person, or be forced to keep up a charade for some reason, but you cannot truly expect to enter a conversation without making certain basic assumption; and if those assumptions
turn out to be wrong to a significant degree communication simply breaks down. I will now consider Habermas’ attempts to develop these necessary expectations into the basis for a moral theory.

2) Discourse Ethics

2.1) Introduction

The centerpiece of Habermas’ philosophy is his discourse ethics. Although it is not the topic to which Habermas has directly devoted the greatest number of pages in his corpus, the establishment or application of its principles stands in the background of everything he has written. The main focus of the preceding sections—aspects of Habermas’ interrelated sociological and linguistic theories—were, in part, preparatory for an understanding of the present focus—what can be called his ethics proper. Despite a genuine risk of oversimplification, Habermas, I think it fair to say, is a neo-Kantian. I suggest the best entry point to his ethics is to analyze how he reformulates Kant and where he differs from important contemporaries also influenced by Kant—perhaps most importantly Rawls. At least that will be that tack on which I set out. But first I will offer a brief overview of what discourse ethics is.

Above, I discussed Habermas’ pragmatic approach to communication in general (both meaning and understanding) and introduced the notion of idealizing presuppositions or rules of discourse in general. Discourse ethics relates specifically to the analysis of the communicatively rational process of moral discourse in which contested validity claims to rightness are discussed. As such, discourse ethics is, in the first place, a reconstructive science. Just as Habermas
reconstructs the rules of discourse in his general pragmatic theory, discourse ethics begins by trying to reconstruct the rules that govern moral discourse in particular. The result of his reconstruction is that moral discourse is governed by two rules which are the two fundamental principles of discourse ethics. They are 1) the discourse principle: only those norms are valid to which all possibly affected persons could agree as participants in rational discourse (henceforth “(D)”); and 2) the principle of universalization: a norm is only valid when the forcible consequences and side effect of its general observance for the interests of each individual could be freely and jointly accepted ahead of any known alternatives by all affected (henceforth “(U)”){92}. In other words a norm is valid only if the interests it represents are universalizable—not arbitrarily favouring one person or group to the detriment or disadvantage of another.

The reason why there are two principles, the precise difference between them, and the way they function with respect to one another within the framework of the discourse ethics approach are matters of longstanding, and I believe well-founded, controversy and confusion. The present investigation cannot concern itself with an evaluation of this topic. For present purposes, suffice it to say that (U), as a rule of argumentation, is taken to be part of the logic of practical discourse, while (D) is taken to be a more “fundamental”{93} or “basic idea of moral theory which does not form part of the logic of argumentation.”{94} One thing that Habermas does

---

{92} For example, Seyla Benhabib has rather persuasively suggested that Habermas ultimately has, or requires, only one principle to act as a test for universalization and the formulation of the theory in terms of two separate principles adds little more than unnecessary complication. Seyla Benhabib, “Communicative Ethics and Current Controversies in Practical Philosophy” in *The Communicative Ethics Controversy*, ed. Seyla Benhabib and Fred Dallmayr (Cambridge Mass.: The MIT Press, 1990)
make clear on this point is that (U) is “the only moral principle” and, it would seem, the one on which the hopes of a rationalist ethic ultimately stands or falls. Thus, the focus of this section will center around (U) and the attempt to “derive” it from the content of the presuppositions of argument.

2.2) Kant and Modernity

Besides following Weber in seeing modernity as a separation of the lifeworld into distinct value spheres, (1-scientific/technical, 2-moral/legal, and 3-aesthetic/expressive) Habermas emphasizes the fact that modernity, and the process of modernization, is characterized by the decreasing significance of traditionally and authoritatively defined roles and values and the increasing reliance on the shared meaning/understanding that springs from public sphere communication for the purpose of coordinating action and creating social order. On this view, the process of rationalization within the respective value spheres represents a development of the species. While this may have a certain Hegelian ring to it, here, society is not seen as an objectified form of a self-developing spirit and human self-realization is not the inevitable unfolding of reason in history, rather, more simply, processes of rationalization are characterized as developmental processes because they lead to capacities and social arrangements which are more functional—which allow us to better cope with our environments, to get along better with each other, and to achieve a greater degree of individual autonomy.

Nature is harsh and unforgiving; its operations are indifferent to human interests. In order to survive we must find refuge from the elements—stay warm
when it is cold, dry when it is wet. We must find shelter or defence from beasts which would devours us. We must feed and protect our children as well as teach them all that we have learned about how to survive our environment. If human beings do not materially reproduce themselves in this way they will do nothing else—for we will cease to be. That is why human beings have a fundamental, even existential, interest in the technological mastery over nature afforded by science, and why, moreover, rationalization pays-off in the scientific-technical sphere. But, according to Habermas, communication is so important to the species that it too can be called a fundamental interest of humanity. Not only do we have an interest in communication but in free and undistorted communication—for it is only on such a basis that we can emancipate ourselves from oppression by forming stable, peaceful, cooperative societies which do not depend on violence and manipulation. Moreover, it is only on such a basis that humanity can be free, living according to understandings, principles, and precepts which makes senses to us, which are accepted or rejected for reasons which are, or become through discourse, our own. This is why Habermas is right to say rationalization pays off in the moral sphere.

Given this understanding of modernity, Habermas claims that Kant is in fact the first fully modern thinker since he was the first to include a modern conception of morality. In general, notwithstanding the rapid enlightenment-era rationalization of the scientific/technological sphere, previous philosophers had failed to shift the basis of moral authority completely to reason—especially a reason germane to that sphere. In one way or another, they had continued to rest the core of their moral theory on the de facto acceptance of certain thick and substantive parochial
conceptions of the good life (as I have argued many contemporaries still do). The monumental achievement of Immanuel Kant was to finally base ethics on reason alone by making his fundamental moral principle a rational procedure for testing norms (witness the first formulation of the categorical imperative: Act only on that maxim which you can at the same time will to be a universal law) as opposed to the substantive prescriptions of any particular form of life. Kant is the first for whom the ground of moral authority is the purely formal test of universalizability. For any given maxim we must apply the categorical imperative’s test by asking whether that maxim could be universalized—if it cannot, it fails the test and should not be acted on. Further, Kant’s test is a purely logical one, such that moral wrongness is a kind of logical inconsistency. When we are immoral we have made a formal mistake in our own reasoning process. Habermas comes to think that because of this important separation of form from content, modern thinking allows us to separate out substantive questions of the good life (ethics) from formal questions of justices and rightness (morality).

2.3) Reformulating the Categorical Imperative

Habermas continues the Kantian tradition in several key ways. Like the categorical imperative, the principles of discourse ethics do not supply content per se, acting rather as a universalizability test for possible norms. Notice that the principles of discourse ethics do not give solutions to normative questions, rather,^A To avoid a possible confusion it should be noted that discourse ethics should technically be called the Discourse Theory of Morality but Habermas developed this general outlook before he incorporated the ethics/morality distinction and, perhaps for continuity sake, the name of the theory was not changed. Thus Discourse ethics is ethics in the general/popular/imprecise sense—not the narrow/technical sense.
creating solutions is the task of those affected by said solutions. Put differently, like the Kantian imperative, the principles (D) and (U) never directly address the question, what should I do? rather, they describe the conditions under which that question can be answered. Instead of justifying any validity claims to rightness they tell us what it means to justify such a claim; they are, in this sense, procedural norms.

For Habermas, our activity in the lifeworld is coordinated in part by validity claims to rightness—that is to say, the implicit moral norms—underlying our speech and action. So long as one's implicit norms—behavioural rules or principles of action—are unchallenged they remain part of the lifeworld, forming invisible tracks along which behaviour and social interaction can run smoothly. When norms are challenged and conflicts erupt the implicit normative content of the lifeworld is fed into the explicit medium of moral discourse in which participants make use of reason to repair the disruption and replenish the lifeworld by applying the universalizability test and establishing a new norm which all can understand and accept on the basis of reason alone. In this way the validity of moral norms is established universally. When accounting for how this has played out historically, Habermas explains that with the post-enlightenment expansion of rationalization within the moral sphere, the background culture of the Judeo-Christian lifeworld became increasingly subjected to the test of moral discourse—weeding out or modifying norms that failed it while preserving those that did not.

Not only is discourse ethics Kantian in that the fundamental principle is a test for, rather than a supplier of, specific maxims, but they are both testing for the same
thing—universalizability. The principles of discourse ethics ensure the process of creating solutions is a fair one—that is to say one which represents the general interest. (U) effectively functions to veto any norms that would favour or disadvantage any interested party.

Despite these agreements, discourse ethics diverges from Kantian ethics in several key respects insofar as Kant remained within the widespread and long-reigning, introspective paradigm of thought, fiercely rejected by Habermas, which he calls “the philosophy of consciousness.” First, Kant’s reason was monological in that his logical universalizability test was one that could be applied by an individual actor reasoning alone. Applying the categorical imperative is a solitary endeavour—a mental calculation wherein the individual actors sees if they can will a given universal maxim without contradicting themselves. On the contrary, for Habermas the test is not one which can be carried out alone. Since moral discourse is necessarily a social endeavour the appropriate form of reason is necessarily dialogical. Habermas would replace “what each can will without contradiction” with “what all can will in agreement.” As he puts it:

it is not sufficient...for one person to test whether he can will the adoption of a contested norm after considering the consequences and the side effects that would occur if all persons followed that norm or whether every other person in an identical position could will the adoption of such a norm. In both cases the process of judging is relative to the vantage point and perspective of some and not all concerned.

Actual participation, then, is the only way to ensure others are prevented from distorting one’s interests. Thus, moral justification is dependent on actual arguments for internal reasons.
Also, universalization cannot be a formal procedure carried out by an individual because this could not explain obligation. To violate the categorical imperative may constitute mistaken reasoning but that leaves unanswered—or at least does not answer satisfactorily—the question; why is mistaken logic morally wrong? Further, Kant appears to rely on the seemingly false assumption that the individual can assume the disinterested and transcendent third-person perspective of an ideal reasoner. Habermas, in siding with George Herbert Mead’s claim that “it is as social beings that we are moral beings”\(^{105}\) arrives at a somewhat different understanding of universalization.

In order to achieve universally valid norms, norms which are equally good for all, Habermas does not need to posit the existence of a neutral and purely rational third-person perspective. The requisite impartiality is achieved, without the need to transcend the first- or second-person participant perspective, as a result of the rational, dialogical process of moral discourse. In his words:

> The interpreter’s inevitable involvement in the process of reaching understanding does indeed deprive him of the privileged status of the objective observer...but for the same reason also provides him with the means to maintain a position of negotiated impartiality from within.\(^{106}\)

The rules of moral discourse, ideally followed, are sufficient to ensure that only norms which embody a genuinely universal interest are accepted by participants.

\textbf{2.4) Reconceiving Autonomy and Deontology}

For Habermas, as with Kant, we see interlocking conceptions of autonomy and rationality at the heart of their respective moral theories despite the fact that these concepts are understood differently by either thinker. As explored above, in
Kant’s view, moral reason is monological; in Habermas moral reason is dialogical. Relatedly, though it is of central concern for both thinkers, safeguarding individual freedom for Habermas takes a different form than it did for Kant, as Thomas McCarthy points out. For Habermas, individual autonomy cannot mean the freedom to act on our own insight arrived at as a solitary moral consciousness. As social beings we have no right to such a freedom; what we do have is a right to participate on par with others in the process by which we arrive at moral norms, the right to accept or reject any proposals or justifications and to introduce any topics, proposals, or justifications of our own—and it is this communicatively rational participation alone in which the equal respect for the dignity and autonomy of the individual consists.

A further divergence from Kant is represented in the fact that Habermas’ account is not strictly deontological. An exclusively deontological approach considers only agent’s intentions. The only morally relevant question is whether one’s motives are purely to act out of respect for the universal law or not, while every consequence of an action, including repercussions for the contingent interests and well-being of anyone, are explicitly and necessarily discounted from consideration. Morally, one’s only duty is to be in accord with reason—the moral law within—and to ignore even one’s own contingent, empirical inclinations and, for that matter, the merely contingent, empirical inclinations of anyone else.

In clear and direct contrast, a moral relevance for consequences is explicitly built into to the fundamental principle of Habermas’ deontological ethic (U). A norm is valid if and only if:
all affected can accept the consequences and the side effects its general observance can be anticipated to have for the satisfaction of everyone’s interests (and these consequences are preferred to those of known alternative possibilities for regulation).  

Nevertheless, it clearly remains deontological in the sense that the equality prerequisite for communicative action entails certain inalienable rights for each individual which cannot be violated for the sake of a particular outcome favoured by any person or group. Clearly though, Habermas’ moral reason is not pure in the Kantian sense.

2.5) Practical Reason

Following a line of thought running from Kant thought Weber, Habermas recognises three distinct types of practical reason. First there is strictly pragmatic or prudential reasoning, which is involved in judgments concerning what is prudent or practically expedient. It is purely instrumental and corresponds to the prudential ought—if you want X you ought to do Y.

Second, there is ethical reasoning which is involved in judgments concerning the substantive content of the good life. When ethical reason is applied in ethical discourse it asks: given my/our particular lifeworld context—history, traditions, self understanding etc.—what is best for me/us? It is a hermeneutic process of self clarification valid from a particular perspective. The ethical ought is not categorical, it asks only: what is good from the limited perspective of me or us? or “how should I/we live given the particulars of our life situation, history, values, and

---

A Pragmatic discourse can only be used here in the narrow sense of the communicative instantiation of instrumental reason. In a broader sense all meaning and all morality is pragmatic for Habermas in the sense that it severs a social function (Meaning is a way of forming relationships, creating social order etc. Moral discourse, in particular, functions to resolve conflicts over moral norms).
beliefs? Notice the similarities between these first two aspects of practical reason and the aspect of practical reason captured under Rawls notion of the rational.

While rational ethical discourse is possible on Habermas account, this ethical reason is significantly unlike the one in Aristotle’s account in that no general and universal answer can be given to questions concerning the full content of the good life on the basis of universal metaphysical insight into the ultimate essence of human beings. No philosophical argument can settle such questions in advance. Discovering the details of what constitutes a good life is something that must be constantly worked out and reevaluated in light of the particular circumstances of the participants in the ethical life of a community.

Finally, there is the moral form of practical reason which is the form directly relevant to judgments concerning what is right or just. In many ways this form of practical reason functions analogously to the reasonable in Rawls account of practical reason. Moral questions are differentiated from all other questions within the practical sphere in that they are decided rationally in terms of criteria of justice, that is, the universalizability of interests. The moral ought is indeed general and universal since norms, which can be accepted by all concerned solely on the basis of reasons they accept, are taken to be valid irrespective of any cultural or historical particularities. Moral discourse aims to answer not what is good for a particular group or individual but what is in the general interest of all. In other words it addresses the question what is right—categorically.

Discourse ethics is based in this moral form of practical reason. In discourse ethics “justificatory power resides only in the discursive procedure that redeems
normative claims to validity. And this justificatory power stems, in the last analyses, from the fact that argumentation is rooted in communicative action.” The program of discourse ethics consists of two steps, 1) the elaboration of a rule of argumentation to govern practical discourse which functions as the fundamental moral principle of discourse ethics viz. (U) (section 2.6), and 2) the justification of (U) (section 2.7), both of which have been discussed above, yet, given their central significance, warrant an attempt at further clarification and elaboration below.

2.6) The Principle of Universalization

Habermas’ theory of argumentation is not a formal one, nor is it one grounded in ultimate a priori certainty through any means. He accepts that purely deductive logic does not tell us anything substantively new and that the substantive content of arguments is necessarily based on our experience and interests (needs/wants), which are subject to various interpretations based on changing theoretical schematizations. Thus, for any rational argumentative practice to move from particular observation to general hypothesis it require what he calls a “bridging principle.”

A bridging principle is justified, not by argument, but a rule that make argument possible. As such, these are not susceptible to deductive justification; rather they are the instruments necessary to span the logical gulf in nondeductive relations. In theoretical discourse there is a principle of induction, and analogously, in moral discourse there is the principle of universalization proposed by Habermas. Incidentally, Habermas maintains that all cognitivist ethics ends up having to introduce a bridging principle as a rule of argumentation which contains
the same "basic idea" behind both (U) and the categorical imperative—namely, the idea of a necessarily general or impersonal character of moral norms such that all valid norms express something akin to a general or universal interest/will.\textsuperscript{115}

When one claims the validity of a moral maxim, either explicitly or indirectly through their speech or action, one exposes the maxim to potential challenges. When challenged a maxim must be submitted to moral discourse for purposes of discursively testing its claim to universality. Because only reasons count in an ideal discourse all imbalances in power are neutralized. Within discourse the reign of communicative reason and corollary neutralization of power imbalance is achieved to the extent that potential norms fed in from the lifeworld are jettisoned if not susceptible to consensus. In this way, (U) functions to block any potential norm that could not meet with the assent of everyone potentially affected by it. Normative justification then is a result of the reasoned agreement of the subjects of the proposed norms.

What is going on here may appear to be that discourse ethics essentially makes validity equivalent to social acceptance. This is not Habermas’ view; to the contrary it is one to which he is vehemently opposed. He maintains a distinction between what is amenable to assent in principle and what finds assent in practice. This means sound reasons are not equivalent to reasons considered sound.\textsuperscript{116}

We must distinguish between the social fact that a norm is intersubjectively recognized and its worthiness to be recognized. There may be good reasons to consider the validity claim raised in a socially accepted norm to be unjustified. Conversely, a norm whose claim to validity is in fact redeemable does not necessarily meet with actual recognition or approval.\textsuperscript{117}
The reason for this is that the ideal speech situation (introduced in sections 1.10-1.11) is a counterfactual that is only ever approximated. In real life situations most norms are accepted as legitimate because of a complex mixture of the unforced force of rational insight on the one hand, and actual force, of various kinds, on the other. Insofar as discourse is tainted by coercion or manipulation of any kind, its outputs will also be tainted; norms will be based on empirical acquiescence rather than rationally motivated assent. Notwithstanding these contaminations, in principle anyone can reach the same judgments with respect to given norms and actions, and will tend to do so in fact as discourse approximates more closely to the conditions defined by the ideal speech situation (or the more ideal the discourse community).

The ideal speech situation is just that—an ideal. This is, after all, a normative theory not a purely descriptive one. And, incidentally, Habermas can offer no guarantees that we will progress toward this ideal, for the realization of reason in history is not the destiny of mankind or the inevitable unfolding of our true nature, but a contingent matter. There can be no assurance that we will not descend into irrational superstition, barbarism and terror.

Nevertheless, if we are to actually move towards a closer approximation of the ideal of human relationships guided by communicatively rational/moral means, if, in other words, moral discourse guided by the moral principle (U) is to flourish, it requires certain conditions, including a prevailing outlook, fostered by enlightenment thinking, which Habermas calls a “decentered understanding of the world”\textsuperscript{118} characterized by the adoption of a “hypothetical attitude”\textsuperscript{119} toward the social world and a willingness to rely on, or submit to, the legitimating power of
rational procedures, namely competitions with arguments. The hypothetical attitude means that the lifeworld is always questionable in principle and entails that one does not dogmatically hold to a core of beliefs that have gained cultural currency. According to Habermas “Discourse ethics emerges when the conventional...is made reflective”120 and the social world dissolves, before discourse, into conventions in need of justification.121 The moral point of view can only flourish when “uprooted and now free floating norms”122 are subjected to the rationalizing scrutiny of moral discourse and either rejected or set upon a new basis—reason. Discourse is a rationalization, and at the same time—because of the normative content of the principle (U)—a “moralization of existing norms.”123

2.7) The Justification of the Principle of Universalization.

By now it is clear that according to discourse ethics:

Moral action is a case of normatively regulated action in which the action is oriented towards reflectively tested claims to validity. Intrinsic to moral action is the claim that the settling of action conflicts is based on justified reasons alone.124

Habermas has argued that for any subject capable of speech and action to reach an understanding about something there is no alternative to argumentation. In the case of conflict over norms that means moral discourse. The principle of universalization is a reconstruction of the everyday intuitions underlying this normative discourse. In other words, there is no attempt to deduce (U) from first principles. As a reconstructive endeavour the justification of (U) does not involve proving it in the sense of convincing anyone they should adopt it as a new principle on the basis of some chain of reasoning. What it does involve is making us aware of the conditions we are all already operating under when we argue.125 Our moral
reference point cannot be found in a first principle or “ultimate justification that
would lie outside the domain of argumentation.”

We cannot, without begging the question, simply provide reason and argument to prove the need for reason and argument. In favour of a pragmatic/hermeneutic inspired approach, Habermas abandons the attempt to do so, which he understands to be bound up with the philosophy of consciousness and which, he claims, has no way out of the epistemological labyrinth of the Münchhausen Trilemma—the idea that the quest for certainty must 1) rely on an infinite chain of reasoning which is impossible 2) prove its conclusion with arguments that ultimately rely on those very conclusions, resulting in a vicious logical circularity, or 3) rest all knowledge on a set of supposedly indubitable or self evident first principles not open to question, which arbitrarily cuts off the rational enterprise, reverting to a reliance on a seemingly arbitrary faith in unquestioned axiomatic assumptions.

In place of a logical derivation of (U) Habermas offers what he calls a “transcendental-pragmatic demonstration.” Here philosophy is no longer overburdened with foundationalism, since it no longer functions as the master discipline which provides the ultimate justification for the empirical sciences and assigns each subordinate discipline its appropriate and legitimate role within the enterprise of knowing/understanding. Instead there is mutual-dependence wherein philosophy looks for confirmation of its theoretical formulations from science and science acknowledges that it is built on philosophical assumption. Once the possibility of ultimate justification is abandoned we need not understand this hermeneutic circularity as vicious. The degree of coherent fit between
understanding derived from various spheres is what justifies our understandings and gives us a more or less reliable picture of our world. In the present case, for example, if empirical sociological or anthropological data were to contradict the universality of the competencies which Habermas reconstructs, the efficacy of those reconstructions would be called into question, but Habermas holds that, thus far, empirical data does fit, and therefore support, his reconstruction.

In what sense, then, can we understand Habermas to be *justifying* moral reason if he does not employ rational deductions? Again, the validity of the rule is not proven, what is proven is the fact that there is no alternative to them. And, as Habermas puts it, “What I cannot meaningfully dispute...for the same reason I cannot meaningfully justify by deriving it deductively.” (U) does not need to be derived because it is not a substantive norm that is being presented for approval; it is merely a careful, formal way of presenting pragmatic presuppositions which are intuitively known and implicitly adopted by every competent social actor. The content of (U) is located in the communicative presuppositions of argument which “always already presuppose those very relationships of reciprocity and mutual recognition around which all moral ideas revolve in everyday life no less than in philosophical ethics.”

Still, the question remains: how do we know what constitutes an unavoidable presupposition of rational moral argument? Following Karl Otto Apel, Habermas claims presuppositions are inescapable when denying them would result in a performative contradiction. If a presupposition cannot be deductively grounded yet cannot be denied without contradiction then it must belong to the class of quasi-
transcendental-pragmatic conditions that must exist if rational argumentation is possible.¹³⁵

To begin with the simplest case, consider the logical- and semantic-level rules of reasoning such as the law of non-contradiction and the requirement of consistency and non-equivocation. Suppose I agreed to settle a controversy through a rational discussion. After I assert “A & ~A” you explain that both “A” and “not A” cannot both be true at the same time to which I explain that I understand I am forwarding a contradiction but have not made a mistake, I simply do not recognize contradiction as problematic. If I do not recognize the law of non-contradiction, however, then I have not abided by the agreement to settle a dispute with reason. I certainly may wish to go ahead and assert “A & ~A,” but if I do I would be doing something else—perhaps a Zen Meditation, or simply trying to make you angry—but I am not having a rational conversation. Obviously for Habermas, there are, likewise, process and procedural level rules that define a situation of communicatively rational action.

Again suppose we agree to settle an issue of common concern—such as, how will we divide a piece of lasagne—on the basis of a rational discussion. You suggest we divide the lasagne in half, but I lie and say there is no need to divide the piece of lasagne because there is plenty more in the fridge. Later, when discover no lasagne in the fridge, you realize that my deception worked long enough to allow me to finish off the one portion. My refusal to be truthful was a refusal to settle the matter on the basis of a rational discourse. Likewise saying “give me the lasagne or I can have you fired,” would represent a refusal to settle the matter on the basis of a
rational contest in favour of strategically exercising my power and influence to
decide the matter in my favour. The upshot is that if we are going to take part in
communicative action we must accept certain preconditions of argumentation that
remove factors other than the force of reason. There is simply no argumentative
standpoint outside these presuppositions from which they can be “criticized.”

2.8) More Justification: Replies to Sceptical Counterarguments

Acknowledging philosophy’s inability to produce a universal metaphysical
basis for ethics, some contemporary moral theorists, including thinkers as
influential as John Rawls, have ultimately given up on the idea that the moral point
of view can or needs to be defended against a thoroughgoing moral scepticism.\(^{136}\)
This is thought to be asking too much of any moral theory. Habermas, however, has
been unwilling to make that concession. Without resorting to either intuitionist
value ethics or metaphysics, Habermas wants to be able to defend moral cognitivism
against all variants of moral scepticism—such as emotivism or decisionism.
Subjectivist approaches based on certain empiricist assumptions call into question
the moral-rationalist contention that moral issues can be settled on the basis of
intersubjectively binding reasons. According to this sort of sceptic, normative
statements are not decidable on the basis of reason; thus, when, in everyday life,
people argue about moral issues as though they were, people are simply confused
and when normative theorists devote themselves to analyzing such issues they are,
unfortunately, trapped even deeper in this mystification. If moral claims mean
anything at all they are either expressions of our personals feelings of approval or
recommendations designed to persuade others.\(^ {137} \) “It is the right thing to show up
and support this rally” really means “I like to support this rally” or “I want you to support this rally.”

Traditional attempts (intuitionist, material value, natural law), which try to argue that normative statements can be true or false, began with the same contested assumption that all validity can be equated with propositional truth. Habermas, of course, maintains that normative statements, though not reducible to descriptive statements capable of being true or false, are capable of being valid or invalid. Since the validity of moral claims can only be established through moral discourse the possibility of moral judgment presupposes the possibility of moral discourse. But in order for moral discourse to be possible it must be governed by a rule or principle. That is why “the justification of (U) demonstrates at the same time that moral-practical issues can be decided on the basis of reasons.” A lot hangs in the balance here. If (U) can be justified to the sceptic, then the possibility of moral discourse will be upheld along with the possibility of its outcomes—universally valid moral norms. At this point I will briefly consider some of the available responses to several sceptical counterarguments Habermas’ justificatory strategy faces.

One sceptical claim is that (U), rather than being universal is actually particular to Western culture; therefore, discourse ethics is ethnocentric. The reply is straightforward: this is an empirical claim which is not borne out by observable facts. Habermas maintains that every culture on earth coordinates activity in part through the exercise of communicative reason, which, regardless of context, rests on a certain universal minimum of pragmatic presuppositions from
which the principle of universalization can be derived or reconstructed. The communicatively rational stance is not a personal choice or a social convention. Communicative reason/action are modes of thought/action fundamental and essential to any form of human life, and acutely so in a decentred contemporary world characterized by the hypothetical attitude.\textsuperscript{142} Habermas readily admits that the status of his reconstructions are hypothetical and subject to revision/criticism as opposed to ultimate \textit{a priori} deductions but claims that reconstructions are less hypothetical the more general the competencies being reconstructed.\textsuperscript{143}

Another possibility for the sceptic is to simply opt out of rational discourse (see my discussion of Rorty in Chapter 3 Section 2). While it may be true that denying the necessary pragmatic presuppositions of rational argumentation and at the same time engaging in the activity of rational argumentation throws up contradictions, a detractor who does not engage in communicative action is a perfectly consistent sceptic whose opposition finds expression not through his or her arguments but through his or her life. Instead of taking a yes or no position on moral norms a sceptic can adopt the role of the ethnologist, the outside observer who describes without judging. Conceivably, at times they may even, consistently, pretend or give the appearance of acting communicatively for purely strategic reasons. They may act to intervene in given contexts to achieve favoured results without holding that anyone’s actions stand in need, or are even capable, of being rationally justified to others.

Habermas replies that the sceptic’s refusal to engage in rational argument is an empty gesture because all human beings grow up and reproduce their life in
socio-cultural forms of life necessarily built upon a web of communicative action from which no individuals can voluntarily sever themselves.\textsuperscript{144} Extricating oneself from all communicative action is an existential dead end:

The symbolic structures of every lifeworld are reproduced through...processes [which]...operate only in the medium of action oriented toward reaching understanding. There is no other, equivalent medium in which these functions can be fulfilled.\textsuperscript{145}

Individuals may have a choice between strategic and instrumental action in individual cases but “do not have the option of a long term absence from contexts of action oriented toward reaching understanding.”\textsuperscript{146} In short, it is not simply the case that entering discourse commits us to certain prerequisites of communicative reason; our very participation in the everyday social world binds us to those same prerequisites because the coordination of our social world is inescapably based, in large measure, on communicative action.

Another significant line of criticism discourse ethics faces is the charge that it merely expresses, like Kant, an empty formalism. In other words, because it is purely formal it can offer no real/substantive guidance. We cannot look to a normative theory that does not tell us what is right or wrong for guidance on how we ought to act. Thus, a purely formal morality is no morality at all.

Before considering how this charge might be answered I think it is necessary to attempt some clarification of Habermas’ employment of the concepts of form and content in this context, which, I think, can be somewhat confusing. Habermas is emphatic about that fact that his ethic, like Kant’s, is formal and is to be distinguished from any substantive ethic, such as Aristotle’s or Rawls’ (notwithstanding the way Rawls understands his own view), which maintains a
particular vision of the good life at its normative core. The normative core of his
theory—the formal rules of augmentation—are not meant to prejudge substantive
content. All norms are content and as such are justified only through the rational
process of communicative action. The presuppositions of argumentation are not
norms but formal and general principles which operate at a different level.

Clearly, if Habermas’ normative core were indeed devoid of any content it is
difficult to see how it could offer any moral guidance; yet, it does offer guidance
because it does, by his own admission, have content in at least some sense.
Speaking of the rules of argumentation he claims that “anyone who participates in
argumentation has already accepted these substantive normative conditions.”

So then, what does Habermas mean by form and content in these contexts? I
believe we can understand the fundamental principle of discourse ethics as formal
in the sense of procedural, as distinct from both the lifeworld imputes to the
procedure and the rationalized output of that procedure. Despite this, the test
applied by that procedure does itself have a powerful, albeit it quite general,
content—namely the impartiality for which the principle of universalization tests.
According to Habermas this “impartiality is rooted in the structures of
argumentation themselves and does not need to be brought in from the outside as a
supplementary normative content.” This is owing to the fact that “the
normatively integrated fabric of social relations is moral in and of itself” These
forms of reciprocity built into social interaction form “the naturalistic core, so to
speak, of moral consciousness.”
Although the universalization principle applies a normatively substantive test it remains formal in the sense that it is a test that is applied to particular norms. It cannot spontaneously generate norms. Potential norms are thrown up by the cultural background or lifeworld context. Norms that survive the rationalizing process of moral discourse emerge as legitimate on account of their embodiment of genuinely universal interest.

So, with a correct understanding of what is meant by “formal” we need not understand Habermas’ normative theory as hollow. It has real normative teeth, in the first place, because Habermas is not trying to replace thick, substantive ethics with formal morality. This theory recognizes a legitimate role for both. Our lifeworldic ethical self understanding of what is good for I/us can exist alongside justice, and the lifeworld can supply material for potential norms. Second, the application of communicative reason in moral discourse brings to bare genuinely substantive demands of justice inherent in the form of communicative action. “Discourse ethics grounds morality on a pattern inherent in mutual understanding in language from the beginning.”

2.9) Emotions and Motivations

While “Moral solutions retain only the rationally motivating force of insight” it remains the case, according to Habermas, that the cognitive operations necessary for the ideal role taking involved in discourse are “internally linked” with both motives and emotional dispositions such as empathy. I believe this point is critical.
First, consider the internal link with motives.\(^A\) Like all ethics based on the human capacity to reason, discourse ethics stipulates that “to act morally is to act on the basis of insight.”\(^{156}\) According to any such view “to say that I ought to do something means that I have good reasons for doing it.”\(^{157}\) For rational creatures, such as human beings, reasons for doing something just are motives. That does not mean that we do not have other motives pulling us in opposing directions but is does mean there is not some unbridgeable gap between reason and motivation. This at least partially addresses the counterclaim made against rationalist ethics that such a moral theory may provide reason for behaving morally, but that is of little or no use since we can easily say “I recognize the reason why I ought to do something but I choose not to do it.” In other words rationalist moral theory does not explain what it needs to—the practical motivation that drives actual behaviour.

This, however, is to misunderstand the nature of reasons. If we truly grasp that something is a reason it cannot help being a motive. If you say you have a reason for doing something that does not motivate you at all it is difficult to understand in what senses it could be called a reason. If it has zero motivation at all you probably have not understood it as a reason for you at all. This relates to the fact that according to Habermas’ theory of meaning “only to the extent to which the interpreter also grasps the reasons why the author’s utterance seemed rational to the author himself does he understand what the author meant.”\(^{158}\) This is the reason why there can be no disinterested, third person, etic-perspective (observer only-perspective) on reasons.

\(^A\) This link was touched on in connection with the “internalism requirement” (Chapter 2 Section 2.7).
Suppose you run into a colleague's office across the hall and explain that you are evacuating the building because it is on fire. You even explain that the fire alarm is sounding, people are rushing outside and it is not an exercise because you can see and smell smoke in the hall. But your colleague says “yes o.k. I understand you, I’ll talk to you later” and then casually returns to the work on her desk. You would have every reason to assume that in fact your colleague did not understand you because, assuming you know she is not trying to commit suicide, if they understood the reason why you were leaving they would recognize at the same time that those were also reasons for them to leave and those reasons would motivate them into action.

All communicative action requires that we understand the reasons behind the claims made in an utterance, and understanding reasons means that we take them seriously as reasons by taking a yes or no position, judging them to be good reasons or bad reasons. This necessary judgment can never be done in some disinterested way as if the conclusions do not pertain to our world. We cannot say “it is a fascinating claim that the building is on fire and I fully understand your reasons for believing it but I choose to remain neutral on the matter.” We necessarily find ourselves in action contexts in which we must accept a reason or not, and once we accept a reason that recognition is a motivation to believe and act in accord with it.

Next, consider the internal link with emotional dispositions. The “moral point of view that proceeds all controversies originates in a fundamental reciprocity that is built into action oriented toward reaching understanding.” It is this “idealized form of reciprocity” which is the “defining characteristic of a cooperative
search for truth."\(^{160}\) If there is to be a universal conception of justice it can only be gleaned from this reciprocity underlying communicative reason. Just norms are valid because they necessarily embody generalizable interests and the reason they do so is because they are the product of an impartial, communicatively rational endeavour. Because reason requires ideal role taking, empathy is also built, as it were, into the process.

The basic idea that reason requires an impartial attitude is seen everywhere from Rawls' veil of ignorance to Leibniz' maxim that we must put ourselves in the place of every other.\(^ {161}\) This attitude, though not the same thing as an emotional state, is naturally, closely related (internally connected) to certain other-regarding empathetic, and even compassionate feelings. Thus Habermas speaks of the "emotional prerequisite for the cognitive operations expected of participants in discourse."\(^ {162}\) Habermas' rationalist ethic is therefore not intended as a cold and purely cerebral affair in the sense that it does not see an ethic of justice as categorically separate from something like an ethic of love or care.\(^ {163}\)

**Conclusion**

Habermas' revitalization of reason consists, most importantly, in two things.

1) Since its enlightenment roots, the rationalization of Western society has given rise to relatively isolated value spheres with their own respective expert cultures, all of which, to varying degrees, are cut off from each other and the day-to-day functioning of the lifeworld. In all everyday communication our moral expectations and evaluations, self expressions, and cognitive interpretations overlap and interpenetrate. Thus, reaching understanding in the lifeworld requires ranging
across the spectrum of moral-practical, aesthetic-expressive and cognitive-instrumental reasoning. Armed with an understanding of the different types of validity claims and the position they occupy in social life we can better understand how to overcome the isolation of various forms of reason and join them to the lifeworld without detriment to their “regional rationality.”

2) Habermas resists and offers an alternative to a one-sided scientistic understanding of reason that has informed much of the 20th century’s social and economic theory and practice to the horrific detriment of the needs and interests of most people on earth. His alternative is to put a spotlight on communicative reason. By conceptually disentangling instrumental and communicative rationality Habermas seems to have found a way to discard the cloudy bathwater and hang on that fragile child of enlightenment reason, and in doing so has hit upon a way in which critical theory might continue to claim, now free from contradiction, that enlightenment reason has been deeply damaging and that the only solution is for current social and political condition to be subject to rational criticism. The problems with reason and the processes of rationalization that Weber and Frankfurt school theorists identified are problems with a one-dimensional conception characteristic of one prominent modern understanding—instrumental reason. The reason that is championed in its place is a balanced conception of reason, a conception in which instrumental reason has its rightful place—not, pace Hume, merely in the service of the passions but in the service of aims established on the basis of practical, communicative reason. Earlier critical theory did not recognize and give due credence to this rationality, which Habermas argues is the only way to
provide an inner logic and corollary inner force or momentum to the "resistance against the colonization of the lifeworld by the inner dynamics of autonomous systems." Thus, we need not any longer consider the contradictions of enlightenment reason intractable aporias.

In revealing the problems with a “deeply ingrained reductionist concept of rationality” that has come to predominate thought coming out of the enlightenment tradition, Habermas essentially diagnoses modern society with the same great madness that afflicted Captain Ahab. Permit me a restatement of the epigraph with which I began this dissertation:

Human Madness is oftentimes a cunning and most feline thing. When you think it fled, it may have but become transfigured into some still subtler form...not one jot of Ahab’s broad madness had been left behind; so in that broad madness, not one jot of his great natural intellect had perished. That before living agent, now became the living instrument. If such a furious trope may stand, his special lunacy stormed his general sanity, and carried it, and turned all its concentrated cannon upon its own mad mark; so that far from having lost his strength, Ahab, to that one end, did now possess a thousandfold more potency than ever he had sanely brought to bear upon any one reasonable object.

The broad and noble tradition of philosophy has led humanity from the dark ages of ignorance and superstition toward enlightened thinking and set us on a path of development—of rationalization. Though we may have considered ourselves on a course to set those dark days behind us we see that they were but “transfigured into some still subtler form.” Like Ahab the strong and brilliantly effective captain, we as a species now peruse our aims with remarkable rationality—a truly astounding genius. But again, like Ahab, despite this genius our “special lunacy,” (a specific sort of moral irrationality) means that we are often reduced to a “living
instrument” to the extent that our aims remain beyond the scope of our reason—our communicatively realized capacity to decide things amongst ourselves. Just as Ahab subverts the reasonable interests and intentions of his employers and his crew by leading the ship and men he commands on a perilous, doomed, and preeminently irrational quest for vengeance on an animal, the displacement of lifeworld communicative reason by the purely instrumental reason associated with systems, cuts our own aims adrift from the scrutiny of reason. Witness how well the 20th century has pursued its “mad mark”—the rationality, the remarkable, strategic efficiency and effectiveness, by which poor are made poorer, so that rich can become richer, the rate at which we can decimate one another thanks to the destructive capacity of our technologies of war, and even the technological and administrative sophistication with which we can nearly exterminate whole races of people in the many genocides of which the Nazi’s Holocaust is only the most remarkable example. Witness also, how, completely blind to the interests of the species, unchecked markets compel us to ravage our earth by pushing cycles of wasteful production and consumption to their natural breaking point in a mad bid to make investors as much money as possible as quickly as possible.

If we are sufficiently persuaded by Habermas’ analyses, our prognosis is not as dire as Melville’s infamous captain’s. Ahab was truly beyond help, no amount of reason or compassion could make him waver in the least from directing his formidable powers towards his obsessive passion to kill the white whale that took his leg. Unlike that dead part of Ahab, the communicative side of human reason, though weekend by colonization from systems, is still intact, and, in fostering the
discursive space for this reason to flourish, humanity has the potential to continue developing—to create a better world, a more reasonable world, a more moral world, one that is more fair, more just and more free—and the reason why communicative reason has this power to hold out the promise of a potentially better world is the universalizing normative force inherent in this very capacity. Thus, an emphasis on the communicative nature of human reason will be central to the outline of moral rationalism which I try to develop in the next chapter.
Notes: Chapter Five

6Habermas, *STPS*, 41.
8Habermas, *LC*, 45.
9Habermas, *STPS*, 56.
11Chambers, *ACCS*, 93.
12Habermas, *TCA2*, 27.
13Habermas, *TCA2*, 153.
14Habermas, *TCA2*, 124.
15Habermas, *TCA2*, 125.
16Habermas, *TCA2*, 122.
17Habermas, *TCA2*, 122.
18Habermas, *TCA2*, 179.
19Habermas, *TCA2*, 212.
20Habermas, *TCA2*, 113.
21Habermas, *TCA2*, 102.
22Habermas, *TCA2*, 102
23Habermas, *TCA2*, 318.
24Habermas, *TCA2*, 165.
26Habermas, *TCA2*, 181.
29Habermas, *TCA1*, 42.
30Habermas, *TCA2*, p. 151.
31Habermas, *TCA2*, p. 390.
32Habermas, *TCA2*, p. 359.
33Habermas, *TCA2*, p. 196.
37Wendling, *HI*, 356.
38Habermas, *TCA1*, p. 101.
36 Habermas, *TCA1*, p. 342.
37 Habermas, *TCA2*, p. 265.
40 Habermas, *OPC*, 170.
41 Habermas, *OPC*, 118.
42 Habermas, *OPC*, 118.
43 Habermas, *OPC*, 118.
44 Habermas, *OPC*, 118.
45 Habermas, *OPC*, 118.
46 Habermas, *OPC*, 118.
47 Habermas, *OPC*, 118.
48 Habermas, *OPC*, 118.
49 Habermas, *OPC*, 118.
50 Habermas, *OPC*, 118 & 93.
51 Habermas, *OPC*, 222.
52 Habermas, *OPC*, 220.
53 Habermas, *OPC*, 118.
54 Habermas, *OPC*, 217.
55 Habermas, *OPC*, 222.
56 Habermas, *TCA1*, 288.
57 Habermas, *OPC*, 315.
58 Habermas, *OPC*, 300.
59 Habermas, *OPC*, 205.
60 Habermas, *OPC*, 122.
61 Habermas, *OPC*, 225.
62 Habermas, *OPC*, 223.
63 Habermas, *TCA2*, 154.
64 Chambers, *ACCS*, 95.
65 Habermas, *TCA2*, 325.
66 Habermas, *TCA2*, 359.
67 Habermas, *TCA2*, 141-143. See also: Habermas, *LC*, 45-50; Habermas, *TCA1*, 369.
68 Habermas, *MCCA*, 19.
69 Habermas, *TCA1*, 287.
70 Habermas, *TCA1*, 297.
73 Habermas, *MCCA*, 58.
74 Habermas, *TCA1*, 100.
75 Habermas, *MCCA*, 136.
76 Habermas, *TCA1*, 42.
77 Habermas, *MCCA*, 107.
78 Habermas, *TCA1*, 41.
79 Habermas, *MCCA*, 59.
80 Habermas, *MCCA*, 108.
81 Habermas, *MCCA*, 108.
83 Habermas, *MCCA*, 160.
84 Habermas, *MCCA*, 88.
85 Habermas, *MCCA*, 87-88.
86 Habermas, *MCCA*, 130.
87 Habermas, *MCCA*, 36.
89 Habermas, *MCCA*, 32.
90. Habermas, MCCA, 14.
91. Habermas, MCCA, 2.
93. Habermas, MCCA, 121.
94. Habermas, MCCA, 93.
95. Habermas, MCCA, 93.
96. Habermas, MCCA, 107.
97. Habermas, MCCA, 198.
98. Habermas, TCA1, 164-168.
100. Habermas, MCCA, 10-11.
101. Habermas, MCCA, 67.
102. Habermas, MCCA, 65.
103. Habermas, MCCA, 67.
104. Habermas, MCCA, 57.
106. Habermas, MCCA, 29.
142 Habermas, MCCA, 82-83.
143 Habermas, MCCA, 95.
144 Habermas, MCCA, 100.
145 Habermas, MCCA, 102.
146 Habermas, MCCA, 102.
147 Habermas, MCCA, 94.
148 Habermas, MCCA, 130.
149 Habermas, MCCA, 76.
150 Habermas, MCCA, 164.
151 Habermas, MCCA, 170.
152 Habermas, MCCA, 103.
153 Habermas, MCCA, 163.
154 Habermas, MCCA, 178.
155 Habermas, MCCA, 82.
156 Habermas, MCCA, 162.
157 Habermas, MCCA, 49.
158 Habermas, MCCA, 30.
159 Habermas, MCCA, 163.
160 Habermas, MCCA, 163.
162 Habermas, MCCA, 182.
163 Habermas, MCCA, 182.
164 Habermas, MCCA, 18.
165 Habermas, MCCA, 19.
166 Habermas, MCCA, 45.
167 Habermas, TCA2, 333.
168 Habermas, MCCA, 77.
Chapter Six

Conclusions: Outline for a Reconstructive Approach to Moral Rationalism
Until the philosophy which hold one race superior and another inferior is finally and permanently discredited and abandoned...until there are no longer first class and second class citizens of any nation...the dream of lasting peace, world citizenship rule of international morality will remain but a fleeting illusion to be pursued, but never attained.

From War- Bob Marley & The Wailers

I want now to take stock and prepare the ground for a synthesis by going over some of the main conclusions of the foregoing chapters with the aim of providing a summary sketch of the positive insights which I argue should be taken from the major contributions to moral rationalism and incorporated into an adequate theory of moral reasoning, while at the same time highlighting some of the major, previously-encountered pitfalls which, ideally, any such account should be able to avoid. To be clear, I make no claim that the result of this distillation process will be anything aspiring to such a grand position as “the most adequate theory of moral reasoning” or even a fully-formed conception of moral reasoning. Rather, by taking account of the most important lessons learned from my analysis of the development of moral rationalism, what I do hope this process will yield is a good picture of what we should be looking for, that is to say, an account of what major features one would hope to include, and what major features one should try to avoid, in any attempt to formulate an adequate version of moral rationalism. If this is successful it should provide a guide to the broad features one should expect in any adequate account of moral rationalism appropriate to the current state of our knowledge and sensitive to the current state of our world. Crudely put, the outcome I envision is a sort of rough and ready check-list of the key pros and cons of competing accounts so that we can size-up any proposed account by seeing how well it is able to hang on to what seem to be true or
indispensable features and how well it is able to avoid what seem like the most serious flaws that have cropped up with prior accounts. With this guide in hand I will then go on to hazard several further conjectures about a version of moral rationalism that might be able to meet these expectations.

1) Taking Stock

1.1) Practical Questions & Rational Capacities:
The Examined Life as the One Worth Living

The ancient Greek philosophical tradition, exemplified in Aristotle, established in the Western world the central idea of moral rationalism: that human being’s life-activity need not be wholly subjected to determination by external powers or authority owing to our possession of cognitive capacities which allow individuals to ask and answer, for themselves, the fundamental questions of life: what is a good life? how can we live such a life? and a host of related questions. Earlier I acknowledged the foundational significance of Socrates (Chapter 3, Section 2.1) and, despite real differences that exist, the clear influence of this moral-rationalist tradition, via Plato, on the thinker whom I have taken here to exemplify this tradition—Aristotle (see my footnote at the beginning of Chapter 1). At various stages I have presented arguments suggesting that this central idea of moral rationalism is both necessary, if we are to find ways to prosper and live at peace with one another, and defensible against both primitive and postmodern forms of irrationalism. While Aristotle forcefully defended this central idea he was mistaken, in certain respects, about how practical reason operates as well as the substantive conclusions to which it leads.

With respect to the latter, I argued that there is no reason to conclude, with Aristotle, that the natural essence and good life for a human being is contemplative reason.
To make these substantive conclusions Aristotle relies on untenable transcendent assumptions, namely that there are divine beings, that these deities are pure intellect, and that our intellect is something radically apart from the material, and moreover that this is the part of us which is of the highest good since it is most like the divine. For Aristotle, contemplative reason is our ultimate ethical end because it has this transcendent nature. At several stages I argued against this as well as other oversimplified uni-dimensional characterizations of the good life (E.g. Chapter 1 Section 1.4, Chapter 3 Sections 1.2 and 3.2). What an adequate account does need is a defence of what I call the core idea of moral rationalism, which is procedural in the sense that it posits reason as a set of capacities or a method which can be deployed to answer fundamental practical questions, while rejecting a substantive version of moral rationalism which equates the good life with reason itself.

Even disentangled from its transcendentally derived substantive rationalism, Aristotle’s account of the procedures of practical reason is not free from suspect metaphysical assumptions. At the root, so to speak, of Aristotle’s practical reason is theoretical reason, a capacity to immediately see the end or purpose of things in the ordered scheme of nature. Our capacity to make practically wise choices, choices that allow us to become good human beings, is predicated on our ability to grasp what a human being is by first seeing what the telos of a human being is, whatever that may turn out be. Even leaving aside what that telos is, Aristotle’s rational procedure assumes that reality is an ordered purposive whole in which all things have unique characteristics which define their essential function, their proper place in the scheme of nature, and moreover that human reason includes a capacity to behold the essential purpose of things. At the very least we can say that there is room to reasonably dispute these assumptions. To make a
Rawlsian inspired point, it would be unreasonable to suppose that public decision making procedures should be dependent on a form of reasoning that requires the acceptance of deeply metaphysical assumptions of such a contested and rationally contestable nature.

Nevertheless, as I have argued already (Chapter 1 Section 1.3), there are certain tensions in Aristotle's own thought which might be resolved in different directions. Already present within his thinking are strands which suggest a less deeply metaphysical or less transcendent account of moral rationalism. These strands contain ideas which should be retained in the face of critical alternative accounts. What I have in mind here includes, for one thing, instead of uncritically relying on the given religious assumptions of his day and falling back on the mystical idea that pure contemplative reason is transcendent and divine, Aristotle could, more promisingly I suggest, have consistently followed through on his conception of mind and body as the form and matter of a unity, which is consistent with an immanent understanding of mind and reason similar in many respects to later developments in the thought of Spinoza, Marx and others. For another thing, Aristotle acknowledges the role science and dialectics play in reason's function (to grasp the nature of things) though, I suggest, he did not fully (or at least consistently) grasp the implications of this recognition. The certain, stable, determinate reference point of Aristotle's system is our secure knowledge of essences, but if fallible, culturally-mediated social practices like science and ever-unfolding discursive argumentative practices like Socratic dialogue are part and parcel of the function and development of human reason, then the stable reference point of a finally knowable human essence is not possible and more importantly not necessary in order to retain the core insight that moral decisions can be made on the basis of rational insight.
It is important to mention here one final issue that might also be thought of as a tension; despite its ultimate uni-dimensional focus on the pure contemplative form of reason, in another sense, Aristotle's view does have what I view as the merit of not reducing practical reason to some pure contemplative or speculative form. Although he does make pure rational reflection the highest end on an ultimate level, on the level of the whole human life he does operate with a multifaceted conception of human beings as having embodied and socially embedded existences of particular kinds which means that practical reason must involve choosing wisely with respect to the various spheres of life in which we are necessarily engaged (See Chapter 1 Section 1.4). I will return to some of these ideas again in a moment when taking stock of the Neo-Aristotelian contributions, but first I want to attempt to summarize the lessons learned through an encounter with Kant's account of practical reason.

1.2) The Rational Will: Lessons From The Copernican Revolution in Ethics

Kant defends a fundamentally unified conception of reason which has undergone a revolutionary shift of focus towards the subject. In its theoretical application, reason makes perception possible by structuring and organizing raw sensation. We can know the world of sensation because it is constituted within subjective structures. Likewise, in its practical application reason gives normative standards, not because it is capable of grasping and then working to realize what is objectively good in the world, but because the rational choices of the subject are the source of value. Here one can note the close connection, for Kant, between reason and a range of enlightenment notions of freedom seen in thinkers such as Rousseau and Spinoza. For Kant, to be free, in the positive sense that we are not only free from determination by any external rule or force but are actually
the cause of our own action and author of our own law, means that our action is
determined by a form of causation apart from the laws governing phenomenal experience,
a form of causation deriving from our purely rational true-self. Thus, the only genuine form
of freedom is to be rational, and the free/rational subject, as such, has a special status, a
unique dignity, as the seat and source of value, and these subjects together make up a
kingdom or community of ends which Kant calls the *Corpus Mysticum*.

As I hope to have demonstrated already (E.g. Chapter 2 Section 2.7), this version of
moral rationalism faces several serious difficulties which must be avoided, hopefully while
maintaining the connection between reason and freedom. Chief amongst these is the
problem of transubstantivation. Reason, for Kant, amounts ultimately to logic. Practical
reason, expressed in the categorical imperative as a test for logical consistency, does not
yield the sort of substantive normative guidance Kant suspects it can. This alone would
require some fundamental retooling of his account. Nevertheless, I hope to have
demonstrated at several stages that certain key Kantian ideas seem like insights worth
preserving: 1) The idea of freedom as following a law that we give ourselves and the
correlate idea of political legitimacy deriving from the application of rational laws that are
thus our own. 2) The subjectivist idea of the rational will as the source of value and the
seat of a special kind of dignity and equality accorded to it as an end in itself. 3) The idea,
or the hope, that rational beings and their rational interests can progress towards
coexistence in a harmonious community, harmonized via the medium of reason—though
not on a transcendent plane.

However, it is problematic to think reason could achieve this harmony in the way
Kant supposed. Reason is able to realize this harmony, for Kant, because the formal
principles of reason are the same for us all; upon sober, individual, logical reflection we recognize our subjective freedom consists in strict observance of a moral law that is one and the same for us all. Kant’s rational harmonization is not the achievement of any actually existing community but a result of our preexisting rational unity as part of a mysterious corpus mysticum. This presupposes a metaphysical conception of human beings as supersensible and mysterious things-in-themselves beyond natural causation.

It is this problematic metaphysical account, which grounds human freedom and reason in a realm of things beyond the world of sense, that allows Kant’s pure deontological approach to make the sharp demarcation between moral/practical reason, which obeys categorical maxims for their own sake, and any form of consequentialism; it is what allows him to classify all instrumental striving to satisfy material desires or interests as forms of heteronomy. In this sense, Kant is just as antithetical to Hume as Aristotle.

I argued at considerable length against the inadequacies of the reduction of practical reason to a role in the instrumental service of pre-given desires (especially Chapter 3 Section 1) and tried to show the force behind various other systems of thought (Kant, Habermas, Rawls, Aristotle, Sen, Nussbaum) insofar as they agree, at minimum, on the point that ends are subject to rational scrutiny. Despite those arguments, which I need not rehearse here, I also suggested that while practical reason cannot be reduced to instrumental reason, Humean type arguments contribute to establishing a more modest point widely recognized by various contemporary accounts of moral rationalism, that desires, consequences, and instrumental reasoning are components relevant to moral decision making. In making room for these considerations, the need for which is made
particularly clear in Rawls’ reformulations, an adequate account must part with a purely deontological Kantian approach.

1.3) Capabilities & Grounding Experiences

I argued that Sen and Nussbaum go some distance towards incorporating much of value while avoiding key difficulties in previous versions of moral rationalism. They incorporate what I claim is an Aristotelian insight and avoid what I claim is a Kantian mistake by refusing to define practical rationality in terms of any logical formula. For Sen, more generally, practical reason is simply seen as the process of subjecting both values and choices to rational scrutiny. Picking up on the multidimensional strain in Aristotle while moving away from the ultimately uni-dimensional aspect, Sen and Nussbaum broadly define the focal variable of value in human life as functionings (any doings and beings) and the capability to function. Ultimate value is no longer problematically equated with rational contemplation of truth (Aristotle) or reduced to a specific account of freedom—determination of the will solely out of concern to act in a logically consistent way by following the categorical imperative (Kant).

Unfortunately, at this stage, the theory has the significant demerit of lacking substantive normative content in order to show what capabilities are valuable. As I pointed out already, it is akin to Aristotle’s golden mean but without a standard (theoretical reason in Aristotle’s case) with reference to which the mean is to be determined—in other words, not a satisfyingly thorough, and ultimately useful, normative account. I claim this is too high a price to pay for avoiding the aforementioned difficulties. Recognizing this, Nussbaum goes on to specify a way of determining an account of valuable functionings consistent with one of the promising strands of thought I identified in
Aristotle. I looked at her Aristotelian argument for the idea that a criterion which is non-detached (immanent and not requiring an Archimedean point beyond history/experience) and objective (transcending the isolation of individual parochial particularism) is to be found in a series of universal spheres of human life called grounding experiences. It is evident that this is a genuinely post-enlightenment version of Aristotelianism in that it is now sensitive to Kantian revelations regarding the active role of the subject in experience. This plays out in the fact that grounding experiences are not taken as uninterrupted givens immediately knowable as Aristotelian essences, but starting points which feed into a dialogue which can work out, in general terms, what sorts of functions are valuable and accordingly what are the best norms, decisions, social and institutional arrangements, etc.

While I suggested that Nussbaum is correct that grounding experiences provide non-subjective elements that feed into moral dialogue, I argued they were not enough (Chapter 3 Section 3.2). The identification of various spheres of life in which we function is not sufficient to guide a moral discourse to normatively binding conclusions regarding what good functioning consists of within those spheres—additional objective content is necessary. Nussbaum, I claim, sneaks in this content with her Neo-Aristotelian conception of the person. Instead of arbitrarily stipulating what we happen to think good functionings/capabilities are, we look to the objective grounding experiences, but when we get to the stage of discourse, our reflective procedures are to be guided by the idea that a good functioning with respect to a given aspect of life is that which makes us fully human. This normative criterion of humanness which is understood in a roughly Aristotelian way, as a life organized by reason, is not made fully explicit and not convincingly defended. Because the bulk of the normative work-load is shifted from the grounding experience to
this criterion of humanness, the account is only convincing if Nussbaum has some certain
immediate insight (which we can all be expected to share) into the true essence of
humanity. Ultimately then Nussbaum is back facing the same problem Aristotle faced.

My proposal is that we might modify Nussbaum’s scheme. The additional objective
and substantive content that is required might reside in an element not fully analyzed by
Nussbaum—the rational reflective procedures themselves. The picture I have of
Nussbaum’s account, together with an alternative augmentation can be represented
schematically as follows:

**Fig. 1**

![Diagram](image)

Granted, having already dealt with the inability of a pure, subjective, logical procedure to
yield substantive guidance, the strict Kantian alternative is already closed. However, there
may be room in an account of practical reason to incorporate procedural content,
particularly if the rational reflective procedures employed in moral considerations are
understood in a different, less formal way. This, I maintain, means looking to certain Neo-
Kantian interventions.

1.4) **Pluralism & Public Reason**

Rawls’ Neo-Kantianism already avoids the major deficits associated with Kant’s
moral rationalism. He avoids the need for transubstantiation because he does not attempt
to derive the normative substance of his theory from reason in any formal, calculative
sense. At the same time Rawls has an explanation of where normative content comes from, which appears to be an advancement over Kant’s account, recalling that, since Kant’s categorical imperative merely offered a formal logical test to apply to maxims there is no satisfying account of where maxims themselves are meant to come from or why they matter.

As I presented it, the central idea behind Rawls’ account of how normative content derives from practical reason is a heavy emphasis on the claim that practical reason is comprised of two irreducible elements viz. the reasonable and the rational respectively. The rational is our capacity for a thick conception of the good life (the key idea that we can critically evaluate our ends which I already suggested is an indispensable core-concept of moral rationalism) and encompasses the instrumental reason employed in the pursuit of those ends (the Humean means/ends reasoning that I also argued needs to find a place in any complete picture). While the rational is associated with ethics, the reasonable is associated with the more restricted idea of morality or justice which concerns, not a full account of the good life, but, more narrowly, what we owe to other members of society, or more specifically, to fellow citizens of a liberal-democratic constitutional state. Thus, we are reasonable when we act out of a desire to justify our public political actions (those affecting others) on grounds others could not reasonably reject, and in doing so constrain the reasonable and our full conception of the good by a publicly shared basis of reasoning.

To put this in Kantian terms, the full content of maxims and the good we hope to advance through action and cooperation, is supplied by the rational as a capacity of individuals as well as the many non-public reasons of various ethical communities within a public political culture. The specifically moral content of public, political constraints on our
pursuit of our particular conception of the good, with which Rawls is primarily concerned, comes from the reasonable. Operating analogously to the categorical imperative, the reasonable (modeled in Rawls’ account by the structural feature of the original position) acts as a test, screening out unreasonable, and thereby unjust, principles of action.

On the positive side of the ledger, in addition to the virtues already mentioned, Rawls appears to incorporate—while avoiding the need to transubstantiate or posit a noumenal realm—what I identified as key strengths of Kant’s moral rationalism; in particular, he appears to articulate a basis for the hope that rational beings and their rational interests can coexist in a harmonious community, harmonized via the medium of a shared public form of practical reason. He is concerned, not only with honouring the special dignity of rational beings by respecting their freedom and equality, as was Kant, but—and this constitutes a distinctively Rawlsian focus which I maintain needs to form part of any adequate account—with facing up to the fact of pluralism or deep reasonable disagreement. He does these things by trying to show how a stable, just society is possible on the basis of practical reason in spite of, and while respecting, reasonable disagreement. This, in turn, is accomplished only by making the useful and pertinent separation between public reason, morality, justice and political value, on one hand, and non-public reasons connected with full ethical value on the other. I grant that the strict separation Rawls imagines expects too sharp a compartmentalization of the individual’s rational capacities, nevertheless, it is hard to see how any workable conception of public practical reason, capable of coordinating interaction in modern democratic societies, could get by without drawing some analytic distinction along this line.
Despite these distinctive strengths, I argued this approach is not fully adequate (chapter 4, Section 8). To begin with, like Aristotle, Rawls’ substantive theory of justice, by focusing on primary goods, is subject to Sen’s criticism of any approach which chooses a focal variable too narrow to capture the range of things rational beings value. Here, however, rather than focusing on the content of the theory of justice, I want primarily to highlight deeper issues with Rawls’ general strategy of justifying moral-political principles with a public form of practical reason. As with previous versions of moral rationalism a practical reason which grasps or generates moral norms is taken to be self-authenticating—it is the point where justification or explanation necessarily stops. In Rawls’ case it may make sense to make this claim about the aspects of practical reason that fall under the rational—in which Rawls does not expect us to find a shared basis for public life—but not those described by the reasonable, which is doing the normative work in Rawls’ moral-political theory. The reasonable, which is not meant to be derived from the rational, depends on a substantive set of ideas about society and the individual, the so-called conceptions of practical reason, which are self-admittedly equiprimordial with the basic principles of rationality in Rawls’ theory but without any strong justification for why they should be taken as such.

So, instead of recapitulating the difficulty associated with the strict Kantian attempt to derive substantive principles from formal reason, Rawls makes what is essentially the same mistake as Nussbaum by illegitimately introducing the necessary substantive content by deriving it from an external (extra-rational) source—specifically, what was a Neo-Aristotelian conception of the person in Nussbaum’s case and a community’s publicly shared conceptions of society and person in Rawls’ own case. Instead of relying on
substantive notions outside the account of practical reason Rawls smuggles\(^A\) undefended substantive concepts into the very notion of practical reason, which amounts to the same thing (Rawls’ general justificatory strategy is represented in Figure 2). The major difference is that Nussbaum employs her discourse guiding normative criterion of humanness in a way that requires it to be objective and universal while Rawls does not claim any such status for the concepts of practical reason that underpin his theory of justice. Rawls does not make claims to universality or objectivity on which he is unable to deliver but that does not let him off the hook. The question remains: why do we have to rely on an acceptance of these concepts and moreover why should they be shielded from all questioning?

Fig. 2

While rationality is presented as a fundamentally unified concept in the style of Kant, what Rawls delivers is a fundamentally divided conception; one in which,

\(^A\) By this I do not intend to imply that Rawls is actively trying to hide his intellectual manoeuvres.
problematically, the reasonable, and hence any public forms of reasoning require that we unques
tioningly accept certain culturally salient fundamental concepts without reasons and constrain what we sincerely believe, for what we take to be good reasons, by principles derived from these ideas. In this way, Rawls' theory makes Kantianism fail in the same way I argued Macintyre makes Aristotle fail—by letting go of objectivity and uncritically and conservatively deriving the normative substance of the theory from the current practices and self-understanding of a particular community. All reasoning becomes radically relative to the self understanding native to a particular group. His practical reasoning is only meaningful for a bounded and particular preexisting political community; specifically the United States and countries very much like it (developed Western liberal democracies). In this way—as an account of how the basis of human cooperation and social organization can, and should so far as it is possible, be shifted to reasoned agreement rather that inherited tradition—Rawls’ account is not successful. An adequate account must avoid the conservative mistake of looking for the substantive basis of a shared point of view in tradition, just as Rawls avoided the mistake of looking for it the logic of individual beings whose true rational selves are already in accord on the supersensible plane of things-in-themselves.

1.5) Discourse Beyond the Philosophy of Consciousness

Although we cannot perceive it directly or with certainty and finality, Habermas argues that human practical reason includes a capacity—called ethical reasoning—to reflect on our final ends or the nature of the good life. Whatever our specific idea of the good happens to be, Habermas seems to endorse the idea that human beings have a fundamental interest in gaining some degree of control over the forces of nature. In
pursuing our ends and, of particular concern here, in gaining control over the forces of nature, we employ a second form of practical reasoning—the familiar prudential or means-ends reasoning. In addition, human beings, as profoundly social, have another fundamental interest in peaceful coexistence and smooth, domination-free coordination of action without recourse to manipulation or violence. Further, as I hope to have demonstrated, this account forcefully maintains that because the only alternative to manipulation, coercion and violence—a fundamentally instrumental mode of interaction—is coordination on the basis of a communicatively achieved, shared set of reasons, human beings have a fundamental interest in rational, communicative interaction. In a move similar to the one I argued Rawls had good reasons to introduce, and connected to this fundamental interest in rationally-mediated, domination-free interaction, is a third form of practical reason called moral reasoning, which separates off from ethical reasoning (concerned with the self-understanding of a particular individual or closely connected community) a sphere concerned with what is right or just.

Like Rawls and Kant, this sphere is arrived at on the basis of a rational procedure which operates by applying a universalizability test, yet one that incorporates improvements over other accounts. Unlike Kant this test is applied neither formally nor monologically. Unlike Rawls it is not applied, again, monologically, while at the same time it claims to be a universally valid procedure as opposed to one that can only ever be appropriate for a pre-existing community fortunate enough to already operate with a shared understanding of itself at the heart of an overlapping consensus.

In these respects, Habermas already holds out the promise of incorporating many of the positive features associated with other approaches. Of course the key issue remains:
from whence does the substantive, and in this case objective, element of the
universalizability test and antecedent claims to moral validity originate if not a moral law
implanted in us all, a direct acquaintance with the aim of a good life, or the established
traditions of a preexisting community? The answer, which I have already endeavoured to
elaborate, is the pragmatic presuppositions which unavoidably underwrite our
communicative interaction, and herein lies the distinctive and, I suggest, positive
contribution of Habermas’ theory which ought to be incorporated into any adequate
account of moral rationalism.

Language, and with it a communicative, as opposed to an instrumental, mode of
interaction, is a human universal. This basic human capacity is underwritten by a set of
pragmatic presuppositions which are unavoidable when we use language and engage in
communicative action. Habermas does not justify these presuppositions with further
rational arguments; they can neither be disputed nor justified in this way. They are
justified only in that they are pragmatically necessary, in the sense that, insofar as we
engage in communicative interaction (which is unavoidable for human beings) they cannot
be denied without performative contradiction.

Like nearly all prior accounts, the self-authenticating aspects of reason include
logical and semantic rules which have often been understood as applicable by an individual
consciousness. On Habermas’ communicative account, the self-authenticating principles
underlying reason also include procedural norms and principles which require equality
and function to exclude from actual dialogue any force save that of the better argument—
the indwelling and therefore unforced (non-heteronomous in Kantian language) force of
reason. These norms include the principle of discourse (D) and the principle of
universalization (U), which is the ultimate source of non-subjective and substantive moral content in Habermas’ moral rationalism. His general strategy of deriving moral principles from practical reason can be represented, in its most simple form, as follows:

**Fig. 3**

<table>
<thead>
<tr>
<th>Pragmatic Presuppositions</th>
<th>Reflective Procedure</th>
<th>Rationally Binding Norms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Of Communication (Universal &amp; Substantive)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Habermas’ universalizability test for the moral validity of actions and norms functions to ensure that valid norms embody interests that are genuinely universal. The question, “does a norm or action represent a universalizable interest?” cannot be answered in advance, in armchair fashion by an isolated individual. This test is intended to be applied to and through real world moral discourse between actual interested parties.

This account also provides an explanation of the place of instrumental reasoning and consequentialist insights in moral theory, and how they are deontologically circumscribed by (U) as a principle of communicative reason which categorically demands, in Kantian spirit, that we do not adopt an instrumental attitude towards rational beings whose status as ends in themselves is respected, not by the monological Kantian freedom to act on the basis of the moral law within, but by the ability to participate on par with others in a process of establishing universalizable norms and making decisions on the basis of no power, privileged status, or force other than reason.

2) Synthesis

I have suggested how the contemporary reconstructive approaches to moral rationalism that I have considered in detail, all incorporate key insights from landmark
contributions to this general perspective which, I think it not an exaggeration to say, has been the central guiding thread of Western philosophical ethics. At this stage I want to consider how the positive elements of these accounts of moral rationalism—which I have tried to pick out and defend—might hang together once various problematic parts—which I have identified and argued against—have been pruned away. Figure 4\textsuperscript{A} aggregates simplified summaries of the general justificatory strategy of Nussbaum’s Neo-Aristotelianism together with the Neo-Kantian approaches of both Rawls and Habermas. Each theory incorporates an aspect which I have argued introduces non-subjective elements which feed into the establishment of moral norms which could guide our collective existence on the basis of reason.

\textbf{Fig. 4}

\begin{center}
\textit{Nussbaum}

\begin{itemize}
  \item Grounding Experience
  \item Concept of Person
  \item Reflective Procedures
\end{itemize}

\textit{Rawls}

\begin{itemize}
  \item Existing Culture
  \item Conception of Society and Person
  \item The Reasonable
  \item Practical Reason
  \item The Rational
  \item Rationally Binding Moral Norms
\end{itemize}

\textit{Habermas}

\begin{itemize}
  \item U
  \item D
  \item Pragmatic Presuppositions
  \item Reflective Procedures
  \item Other Principles
\end{itemize}
\end{center}

\textsuperscript{A} It is my hope and intention that it will prove useful to refer back to this schematization in order to clarify some of this section's main arguments.
For Habermas these include the logical and semantic rules which make up the non-procedural norms of reason. Since reason cannot be justified using reason, it is fairly easy to see how these principles necessarily represent a terminus point for rational justification. Perhaps even the procedural principle (D), which gives expression to the idea that rationally valid norms must be freely acceptable on the basis of no force other than reason, has a fair claim to represent a fundamental and irreducible aspect of rationality. Any form of assent that deserves to be called a “rational assent” must be understood as given freely rather than forced. Similarly, for Rawls, there is “the rational” which includes the basic practical capacity to reflect on our ends (which was then particularly well elucidated and defended by Sen) as well as our instrumental capacity to pursue ends (which was thoroughly analyzed by Hume’s approach, albeit one which focused too narrowly on this function). I have argued already, at several stages, that these capacities have a fairly solid claim on being understandable as basic aspects of human reason (Chapter 2 Section 1, Chapter 3 Sections 2 & 3, Chapter 4 Section 8, Chapter 5 Section 2.7). Finally, for Nussbaum, there are the objective grounding experiences which define the basic spheres in which human beings function either well or poorly. Although these do not have the status of being directly and infallibly intuited by reason, as they might have for Aristotle, many of them have been fairly well established and would be widely accepted as anthropologically deep seated aspects of human existence (Chapter 3 Section 3.2).

I will propose that the most adequate account of moral rationalism ought to, and can consistently, incorporate all these features. Even taken together, however, these features will not be enough. All these approaches have had to grapple, implicitly or explicitly, with the fact that these aspects are insufficient to rationally ground the normative projects each
is pursuing. Logic without *procedural principles* including (U); the rational not limited by the *reasonable*; and spheres of human functioning without a *normative account of the human being* to define good functioning; simply prove inadequate. This is the reason why these approaches introduce additional morally substantive content in the form of *procedural principles, the reasonable, or a normative account of the human being*, and this is where things get particularly tricky. Despite the difficulty, any adequate account of moral reasoning must account for A) what is the substantive content of these additional elements? and B) what is their epistemic status, that is, how are they justified? These key issues must be addressed before proceeding any further with a proposed synthesis.

With respect to the former, I want to argue there is room in this whole debate for significant clarification. The additional substantive content in each theory is supplied by Habermas’ (U), Rawls’ conception of society and person, or Nussbaum’s own normative account of the person, respectively. While I do not wish to maintain that the substantive content of each aspect is equivalent or functions to produce precisely equivalent results, I do wish to point to the fact that each element shares at least one critical core idea (worked out in various ways). Moreover, I want to put forward the proposition that this core idea is the only necessary piece of substantive content required over and above the basic non-subjective elements already enumerated above and, relatedly, I want to suggest that, to the extent that the additional substantive content of each theory is adequate, it is so on account of the fact, or to the extent, that it properly works out the implications of this one core idea. Further, a great deal of the understandably significant controversy and complexity involved with working out the precise implications of this idea obscures the fact that, at the deepest level, there is a single key notion at play, and relatedly, that there may be more
fundamental levels of substantive agreement within these versions of moral rationalism that are not generally acknowledged. The basic idea I have in mind is the idea of the fundamental equality of rational agents.\(^A\)

Nussbaum’s argument that every human life necessarily functions within certain spheres of experience does not contain or entail the egalitarian assumption that we ought to strive for the equal achievement of individuals within the space of human functionings and capabilities. Rather, the idea of the fundamental moral equality of rational actors is a part of both Sen and Nussbaum’s conceptions of the person and the elucidation of the functionings/capabilities sphere as the evaluative space for egalitarian consideration has no point or function without this assumption. Similarly, Rawls’ notion of the rational does not embody any assumption regarding equal moral consideration. One can form and pursue goals in a way that is completely rational while refusing to acknowledge or respect the interests and considerations of anyone else. The implication of the equal dignity and respect which is to be accorded to anyone capable of forming rational interests is only worked out in the notion of the reasonable limits on the purely rational. Finally, Habermas was clear that of all the rational rules and principles presupposed by communicative action, the only one with moral content is the principle of universalization (U), and it is (U) which embodies the idea of fundamental equality that is then worked out in terms of universalizable interests.

Though it includes other ideas, the most basic postulate contained in Nussbaum’s conception is the fundamental equality of persons. The reasonable for Rawls (though it

---

\(^A\) Whatever else it entails, fundamental equality itself can be expressed as the proposition that all rational actors (persons) deserve equal moral consideration.
may rely on other ideas) and (U) for Habermas are above all the attempts to spell out what we owe each other as a direct implication of our fundamental equality. Furthermore, I have shown in previous chapters how each of these approaches fails without the additional substantive content of a notion which I now claim revolves around this central notion of equality. Thus, it seems that if any version of moral rationalism is going to succeed, the one indispensable idea that is needed beyond (D), beyond logical and semantic rules, beyond the rational, and beyond grounding experiences, is fundamental equality. The key question will now concern the status of this idea, viz. is fundamental equality a freestanding assumption that goes beyond moral rationalism or is the assumption contained somehow within an account of practical reason?

It is not surprising that this is the most difficult aspect of morality to subsume under the conception of practical reason. To begin with, there is every reason to suppose that this idea stands apart from the other principles of practical reason. For instance, when I disagree with a libertarian about whether the fact of basic moral equality implies that the focal variable of practical equality seeking should be simply certain negative rights, I stand ready with counterarguments; likewise, if a misogynist doubts that women should be considered persons, I am prepared to argue that if men are persons women must be as well, but when one such as Nietzsche expresses a profound disgust for the very idea of equality, claiming instead that it is proper for the strong to dominate and assert their power over the weak, I’m not certain there is a way to respond that he has failed to respect some logical principles of inference, evidentiary criteria of justification, or any similar standards of correctness. Nevertheless, even if we cannot show that Nietzsche is being irrational in this respect, we could say he is being unjust, uncompassionate or
unreasonable, in Rawls’ sense—but does this entail a violation of anything that deserves to be called practical reason? I suspect that it does, and I will strive now towards an explanation of this.

2.1) Incorporating Equality

Luckily for this or any cosmopolitan project there is a wide base of agreement on the principle of basic equality. Despite a great deal of room for dispute with respect to how we should interpret this basic equality and what its implications are, most moral theories, religions, and cultures contain major strains of thought that embrace it, if only in principle; and this thin level of universal agreement could be enough to get a transcultural process of discussion and negotiation off the ground. This is a good thing and it is a fact which should be exploited in the practical interests of working towards even greater levels of agreement and forms of solidarity. However, whether or not the injunction to respect equality rests simply on the fact of explicit agreement within a culture or even between cultures is a separate matter.

I have shown how, for Rawls, instead of starting with our own thick cultural universal and its substantive conception of what the good life is, we look for agreement on a thin norm that has more to do with the demands of justice or fairness than a full picture of what the best sort of life is for a human being (Chapter 4 Section 5.1). This thin but substantive set of cultural norms would act as a sort of minimum filter which would determine what deeper and more specific values and choices can pass through. Equality is the most fundamental component of this thin account. For Rawls then, equality, as the key underlying idea of the reasonable, depends precisely on a factual agreement. On this view there is some aspect of justice (the most substantive part) which evades a universal
rational explanation that could convince a rational moral sceptic through argument and evidence. From this perspective the moral sceptic cannot be answered by deriving principles of justice from rationality. On the basis of arguments developed in Chapter 4, the status of culturally-inherited-norms was shown to be inadequate to the task of incorporating a principle of fundamental moral equality within any genuine form of moral rationalism, despite Rawls’ claim that the reasonable is part of practical reason.

Unlike Rawls, Nussbaum wants the norm of equality to have a universal status but fails to offer an explanation of how the norm is able to claim this universal status. I have already argued that her idea of grounding experiences is too thin to yield the sort of universal normative claims she wants to make about the nature of a well functioning human life. So, looking at figure 4, it should now be apparent that once we have excluded Nussbaum’s neo-Aristotelian ideal of personhood, and Rawls’ culturally inherited conceptions of society and person the only general route remaining by which the notion of fundamental moral equality could feed into the establishment of rationally valid norms is (U). Fortunately, Habermas’ general justificatory strategy for (U) looks like a promising account of how fundamental equality can be incorporated as a basic component of moral rationalism.

If legitimate universals are the outcome of a process, then we must imagine that process to be a fair one. That would involve the process itself being regulated by certain norms or ideals—most likely norms such as: everyone deserves a free and equal opportunity to participate etc. I believe this points to the fact that even a thoroughly sophisticated articulation of procedural reason and open ended universals will not be able to get the normative ball rolling on their own without morally substantive universals which
guide the procedure. In short, it cannot be the case that every norm is always derived from reasonable procedures because there needs to be some prior universal governing the procedure in the first place—hence, we are led to conclude the need for some, at least thin, universals underlying communicative action as a pragmatic presupposition. In addition to other principles of practical reason I suggest that fundamental equality has this status.

The universal I am proposing as necessary is very thin, in fact it is one principle—a single claim: *all persons are moral equals*. With the addition of this one simple commitment I propose we can go on to work out discursively all the provisional principles that would be needed—although they are never fixed or immune from contestation themselves—to govern the requisite discursive processes of cultural translation and negotiation, principles such as: everyone involved should have a chance to participate, to have the conversation put in terms understandable to the participants addressed (language, terminology, level of abstraction etc.), participants should be free from coercion, manipulation, deceit etc., everyone should be required to listen to and work to understand the position of other interlocutors, and likely many other conditions.

While the particular implications may be worked out in different ways, if we are going to participate in a communicative mode of interaction, and coordinate our activity on the basis of reasoned discourse—which is a species-level capacity of human beings and a universal feature of human social interaction—then we need to begin from the basic notion that participants are members of a discourse community, a sort of earthly kingdom of ends, by virtue of their unique status as rational agents or persons, which makes them fundamentally equal to other participants. Part of what it means to view someone as a person and engage with them rationally means not invoking power, status, or any other
form of superior social or material position in order to influence belief or compel compliance. Notice here that I am not committing to defend the full details of Habermas’ particular formulation of the pragmatic presuppositions of communication but the idea that these are correctly seen as trying, in part, to work out the implication of equality, which is a pragmatic presupposition of practical reason on the most fundamental level.

To recap, the contemporary reconstructive approaches on which I am drawing here contain elements—the rational, grounding experiences, (D) and other pragmatic principles presupposed by communicative action—which function as non-subjective resources for the establishment of norms on the basis of reason. These approaches agree, if not explicitly then in the form of their account, on the fact that some additional, morally substantive content is necessary to get moral rationalism off the ground. Despite the complex and varied concepts and principles proposed, at root, what is being introduced, or what needs to be introduced, is simply the fundamental moral principle of equality which becomes complicated only in trying to spell out its precise implications. Fundamental equality is an unavoidable pragmatic presupposition of a communicative mode of interaction. If we are going to interact with others as rational agents, subjects who are the source and seat of value and not merely things, then we must first accept their inclusion under this special status of person, fundamentally different from other physical objects of the natural world—a status which is one and the same for ego and alter alike. This makes appeals to thick ethical accounts of the human being and to inherited cultural traditions unnecessary for moral rationalism. The general approach I am recommending, and which I hope incorporates the strengths of previous accounts without inheriting any of their more
serious drawbacks, is schematized in Figure 5. I label this the discourse-capabilities approach and I will now attempt to clarify its broad features.

2.3) Overview of the Outline

The first and most obvious thing I want to highlight about a discourse-capabilities approach is the central place given to reflective procedures. While all the accounts on which I draw accord some significance to reflective procedures of some sort or another, this approach centers on communicative reason in the strongest sense. Instead of starting from some fixed, thick understanding this approach has moved, with Habermas, to a more procedural understanding of reason in which aspects of the procedure can be opened to endless renegotiation. The procedural reason that I look towards has universal normative content embodied in the reflective process itself. This is clearly similar to what Habermas calls ‘discourse’ and may also bear certain similarities to what Judith Butler has called cultural translation. Understood in this way, reasoning can fulfill what I characterized as one of moral rationalism’s central aspirations, by providing a standard that is “both immanent (not to be found outside of concrete language games and institutions) and transcendent (a regulative idea that we used to criticize the conduct of all activities and institutions).”¹ Once reasoning is understood in this more procedural way the universals which this never-ending process of discourse, debate, and renegotiation generate will have the character of remaining provisional and never being fully or finally articulated—of being, as Butler puts it, an “open ended ideal,” a “not yet.”² The ability to criticize and move beyond the conventional formulation and parochial instantiation of a norm always remains possible, according to the discourse-capabilities account, because of the transcendent potential of procedural reason—the ability to constantly have ideas challenged and subject
to reasoned contestation from without, even from sources that are radically other, including those arising from different cultural contexts.

Fig. 5. The *Discourse-Capabilities Approach*

Now, the topic or objective of this rational discourse is concerned with the central questions of practical life including specific questions such as; what is the right action to take in a specific context? and even, more fundamentally, general questions such as; what would a good or well functioning human being be like? These latter are not fixed in advance. Answering them is the unavoidable and ongoing task of humanity, and individually of every human life. The specific sort of questions we ask are, however, anchored to anthropologically deep seated facts of our existence—for instance, the fact that we love and conflict, eat and breathe, live and die. Questions about what it is to function well, and to have the facility to function well, must take the form of reflection on how we are able to function with respect to the various spheres in which we necessarily operate as human beings.
These rational reflections must be conducted according to various rules and principles, which are themselves underwritten by deep-seated features of our nature as social, language-using, and communicatively-acting beings. These features include: 1) Principles of Reason, which encompass logical and semantic rules; 2) a Freedom Condition of some kind, which expresses the internal connection between freedom and reason by asserting that, for any assent to be considered rational it must be given freely; and 3) an Equality Condition which asserts that all free rational actors involved in the reflective procedure participate on an equal level, requiring reciprocal responsibilities and entitlements of various kinds, the concrete details of which may need to be worked out, negotiated and renegotiated in the course of debate.

It is not possible, without performative contradiction, to engage in rational discussion, or coordinate activity on the basis of mutual understanding, without accepting these thin universal conditions, at least in principle. This has consequences in terms of the sort of orientation which can be adopted vis-a-vis irrationalists who oppose these conditions. Any groups or individuals not sharing in these universals will have beliefs radically incommensurate with those who do and therefore, there could never be any question of compromise between the two or of one trying to accommodate the other. I maintain that the failure of a normative theory to accommodate, that is, the incompatibility of a proposed thin universal with certain other views, is not a weakness but a strength. This is because, although it is certainly a limit, it is a commendable limit. The hard limit of an egalitarian normative theory that seeks to embrace difference must be inegalitarianism (any view that fails to acknowledge the equal moral worth of every person), for toleration of any such theory or practice would amount to the opposite of toleration. For example,
tolerating the practice of genocide would be antithetical to the basic egalitarian orientation of this version of moral rationalism. The idea of embracing otherness would disintegrate, if it had to mean embracing theories and practices that view or treat certain others as less than full persons. There should never be any question of trying to embrace all views, as suggested by certain philosophies of pure difference or otherness, for this would leave us without any normative or transformative potential. It would be the antithesis of a critical social theory and the negation of the possibility of progress. Views incompatible with the thin universal must not be accommodated but opposed—dragged, wherever possible, into the light of reasoned scrutiny.

To sum up this overview, the moral norms that govern our collective existence are rationally binding and legitimate to the extent that they can be seen as the direct outcome of free and fair reflective procedures. At the same time, as indicated by the thick arrows in figure 5, reflective procedures can reflexively make the form and content of the reflective procedures themselves the subject of rational reflection. However, at the deepest level it seems impossible to conceive participation in communicative reason without accepting the validity of, in addition to principles of reasoning, the freedom and equality conditions which must govern any reflective procedure that deserves to be considered rational. The principles of reasoning and the freedom condition have been successfully incorporated in different ways into many of the different approaches, but the equality contention has proved both necessary and difficult to explain. Employing a Habermasian form of justification I have incorporated it by arguing for its status as a pragmatic presupposition at a fundamental level even deeper than the particular implications we can rationally derive
from it—which may include Habermas’ articulation of (U) or something like Rawls notion of the reasonable.

Having given a rough overview of the structural features of the version of moral rationalism represented in figure 5, I want now to go through it once more in order to clarify how I see it as embodying, in a clear and coherent form, an amalgam of core insights from the philosophical tradition of moral rationalism which I set out to analyze. I will do this by showing how the discourse-capabilities approach is the outcome of following through a progression of fundamental moral questions and turning to different strands of moral rationalism for answers along the way.

2.3) Question and Answer: Elaborating the Outline

Once we overcome the inadequacies of instrumental rationality (discussed in chapters 2 and 3) and recognize that actual ends themselves must be rationally scrutinized, I argued that questions concerning practical rationality become unavoidably linked to questions concerning what is of value. In other words, what I characterized at the beginning of this thesis as the most fundamental and practically unavoidable questions of life—what should we do? and, how can we figure out what we should do?—led to the questions, what is valuable? and how can we discover what is valuable? Taking this as a starting point, I believe I am now in a position to show how the outline of a normative basis for social theory, which I outlined above, can be seen as having unfolded as a response to a progression of five fundamental questions beginning with:

1) What is of value?

To begin answering this question I think it is best to draw on the fundamental Kantian arguments, presented in chapter 2, for the special status of persons—rational
agents who bestow value through the act of valuing. This has two key consequences. First, that all rational beings bear a unique status. Persons, in virtue of their rational agency, have an inviolable dignity as valuable ends in themselves—the only thing valuable not on account of someone else valuing them. Therefore, whatever else we value or chose to pursue, persons have a highest order value which trumps all derivative forms of value. This means that one ought never to adopt a purely instrumental attitude towards others. Habermas then showed (chapter 5) how this injunction not to treat others as mere objects, by instrumentalizing them, is a presupposition of a fundamental form of human interaction in which people act communicatively.

The second consequence of this is that the things people have reason to value are thereby invested with ethical value. The Capabilities approach then offers a particularly sophisticated analysis of this value, arguing that what is valuable, and ultimately what should be the target of egalitarian concern, cannot be simplistically captured by rights, basic goods, happiness etc., but must cover the full spectrum of human capabilities. My approach incorporates this insight by following the capabilities approach in saying that what is valuable—beyond rational agents with their equal status as ends—includes all valuable states of being and doing, and, moreover, that respecting the moral equality of persons necessarily involves paying attention to their actual achievements in terms of these rational interests. This is part of the reason for the central significance of capabilities in the discourse-capabilities approach.

However, I argued that the only way to make the capabilities approach part of an adequate account was to embed it within some other normative theory that could specify a criterion for determining what counts as good, as opposed to poor, functioning. Now, I
want to stress the fact that this can be done without losing or trivializing the insights of the capabilities approach. For example, a capabilities approach could be underpinned by a utilitarian criterion stipulating that functionings and capabilities are valuable in proportion to their ability to produce a net gain in pleasure over pain. Notice that this would not wipe out the significance of this approach by simply reducing it to a variant of utilitarianism. A significant theoretical gain over utilitarianism would have been introduced by the fact that what is of inherent value is no longer reduced to utility or any other single phenomenon. What is valuable would still be functionings/capabilities, not pleasure. The utilitarian principle would only come into play as a criterion for determining which functionings are valuable. So, while a capabilities approach based on a utilitarian criterion would say that the states of being or doing that cause pleasure are the actual locus of value, utilitarianism would say that achieving some state of being or doing which makes us feel happy is only of instrumental value and could be bypassed if, for instance, we could experience the same pleasure without achieving those states (e.g. by being misled into thinking we had achieved those states).

Nussbaum's version of Sen's capabilities approach does something analogous to this. Instead of claiming that leading a life of contemplative reason is the one and only thing of ultimate value, Nussbaum acknowledges the wide range of capabilities that make up the sort of life human beings value. However, as a criterion for determining which capabilities are to be valued she turns to a principle stating that valuable capabilities are ones that allow us to live a fully-human life, which is a life organized by reason. So, what are valuable, according to Nussbaum, are capabilities which comprise part of a rationally ordered human life. I claim this represents a marked theoretical departure, indeed an
improvement, over the straight Aristotelian view according to which reason itself is the ultimate value.

What I want to do to capability ethics is something which is again analogous to the above strategies, only, I hope, leading to an even more adequate final result. Ultimately, to foreshadow a little, I want to claim that the locus of value includes both rational agency and the full spectrum of capabilities which we find to be part of a flourishing human life in the course of a free and open debate governed by and embodying a communicative form of reason (hence the need to stress the discursive in the discourse-capabilities approach). I will explain how I hope to arrive at this, beginning with an elaboration of my answer to the question which Sen left unanswered—the second in the progression of my current inquiry: 2) *How do we know which beings and doings are the ones of value?*

On one level (though I will revisit it on another when I consider question 4) I am in accord with Nussbaum and the Aristotelian tradition which claims that there are certain features of special importance, without which a life could not be recognized as fully human, on the basis of which universally valuable human functionings can be defined as the beings and doings which contribute to a rich and fully-human life. I have already made clear my objections to the idea that reason affords direct and final insight into a timeless human essence which specifies our place in the world once and for all; however, I cannot foresee the possibility of a theory with any critical normative force operating with a completely blank or hollow conception of human nature. In keeping with a tradition that I see running from Aristotle, through Spinoza, Marx, Dewey, and other moral rationalists, I see this approach as in-line with a normative theory based in a version of Humanism—what early critical theory considered a science of humanity. In short then, I want to answer the
question “which capabilities are valuable?” thusly: valuable capabilities are the ones that are best for us—the ones which we have reason to value, given the kinds of beings we are. Hence, we are led to ask:

3) What sort of beings are we?

Humanism, first of all, can be contrasted with pure transcendent ethics based on submitting to a transcendent authority or standard; humanity becomes the author of its own norms. Instead of norms aimed at denying and repressing oneself, becoming subservient and obedient to some external power, ethics is focused precisely on the realization of human potentials. The danger, once value revolves around humanity as opposed to the reverse, is that norms seem to become relative to the arbitrary whims of individual preference. The possibility of a non-subjective normative account of the human being mitigates this result. Just because value derives from human beings does not mean anything goes, that there cannot be objective value judgment about what is good or bad for human beings that hold for everyone, provided we know something about the nature of human beings; and knowing about human beings need not involve appeal to metaphysics. In the words of Eric Fromm:

Human nature can never be observed as such, but only its specific manifestations in specific situations. It is a theoretical construct which can be inferred from empirical study of the behaviour of man. In this respect the science of man in constructing a ‘model of human nature’ is no different from other sciences which operate with concepts of entities based on, or controlled by, inferences from observed data and not directly observable themselves.\(^3\)

This sort of humanism (what Fromm designated with the now anachronistic title “science of man”) is distinguished from foundationalist forms of essentialism in that it maintains that “this science does not start out with a full and adequate picture of
what human nature is; a satisfactory definition of its subject matter is its aim, not its premise.” It is a contingent, fallibilist enterprise, the concepts of which are continually open to contestation and revision. So, instead of answering the question of our nature with a full description that can serve as an epistemic foundation of moral rationalism, I advocate something of a perspectival shift of focus towards the method for answering this question. This represents an important departure from Rawls, for whom we reason publicly within a political community with an inherited set of shared ideas about society and the individual. On my account, such fundamental ideas are never taken as foundational or shielded from rational scrutiny. Here, moral discourse covers the topic of capabilities while penetrating deeper, to underlying notions of person and society.

I think one can see Nussbaum as allied with this humanist tradition insofar as she constructs a model of human nature based on inferences from observable data and shows how the overarching question of human nature can be broken down into a series of more specific questions tied to observable grounding experiences. Instead of just asking “how does a human being function well, as such?” it becomes possible to ask how does a human being function well with respect to the fact that he or she is mortal, and so on. This is an advantage; however there remains a lack explanation of the method for answering the question “what is virtuous behaviour with respect to the various spheres of human functioning?” And, since no conception of human nature can offer a certain foundation for determining what proper human functioning is, we are led back to ask:
4) How do we figure out both: who we are and ought to be, and relatedly what capabilities are most valuable or worthy of pursuit?

There will always be significant disagreements about what good human functioning consists in. Despite this fact, I have argued that we are, fortunately, rational beings capable of discursively trying to figure out who we are. We are capable of providing reasons, and using language and argument to convince and understand one another—in fact, we are practically bound to do so. Hence, normative theory, rather than being imposed by an authority, can be conceived of as an outcome of a discursive, dialogical process of refining or revolutionizing our self understanding and our norms based on this understanding. It is possible, then, to deal with that fact that there is difference of opinion about our nature, not by snuffing it out, but by recognizing that such difference is unavoidable and engaging others in dialogue.

That we possess such discursive rationality is knowable, I have already suggested, through a form of rationalist humanism that follows Habermasian reconstructive science, which attempts to reconstruct, or make explicit, deep universal aspects of human nature that lie beneath and make possible our normal daily interactions in social life. I stress, that reconstructive science is neither foundationalism nor essentialism. Its insights into human nature are not themselves immune from debate and criticism. Reconstructive science is thoroughly fallibilist—understanding our competencies as subject to change and development—not a form of transcendental argument designed to establish necessary, ahistorical conditions. In short, my proposed answer is that we figure out what we ought to be and what is best for us in a perpetual rational process of communicative discourse which requires no foundation beyond the pragmatic presuppositions of communicative
reason. I believe I have already said enough about the justification of communicative reason (Chapter 5) and the central place accorded to reflective procedures; what I want to make clear at this juncture is how this fits into the framework I am proposing here.

In essence, what I have proposed doing is embedding the capabilities perspective within a version of communicative ethics. Nussbaum embedded Sen within a Neo-Aristotelian account of the person and, with the elaboration of grounding experiences, introduced a non-subjective element to frame and concretize the questions addressed in moral discourse, but she did so without adequately accounting for how to determine what constitutes virtuous functioning in the various universal spheres of human life and without explaining the presupposition of moral equality. I take on board the insights of both Sen and Nussbaum by embedding this whole approach within an account that is capable of explaining how we can non-subjectively address questions about capabilities through the use of communicative reason, which at the same time succeeds in introducing fundamental moral equality via what I called the equality condition.

Given the significant room for reasonable disagreement on the fundamental moral questions that are the subject of reflective procedures it is necessary to ask a final question of particular relevance to the possibility of progressive social and political theory:

5) Is it realistic to hope that, guided by communicative reason, we could arrive at a consensus regarding the nature of the good life which could form the moral basis for a stable and peaceful form of collective life under which free and equal individuals can flourish?

I take Rawls, who has grappled so deeply with the specific problems posed by the fact of pluralism, to be the first source of inspiration here. In short, I want to be able to answer this question the same way Rawls would: by saying “no, but that is not necessary.”
In other words, the moral basis for a stable and peaceful form of collective life under which free and equal individuals can flourish does not require that we agree on a more or less thorough account of the nature of the good life, only that we share a conception of justice that requires us to constrain our private conceptions when reasoning in the public sphere on matters of collective concern.

Rawls wants to constrain, absolutely, each citizen’s rational reflections on the good life, together with their instrumentally rational pursuit of those goals, by the reasonable injunction that they respect what they owe to other free rational citizens as moral equals. The Discourse-Capabilities model proposed here introduces the same constraint, only here the reasonable constraints are decidedly not taken from a shared political conception but introduced as freedom and equality conditions that are basic and universal presuppositions of rational interaction with a status not entirely dissimilar to other principles of reason such as the law of non-contradiction. This has an important implication for how the constraint functions.

For Rawls, any rational individual can work out the basic principles of justice that fall out given the currently shared political conception. It is then the responsibility of each individual to only present publicly, arguments which appeal to the already established consensus. The unfortunate conservatism and inherent anti-minoritarianism of this approach is avoided if constraints need not be imposed primarily by individuals but in the process of public, rational discourse. I grant that citizens have a certain civic responsibility to filter their public discourse by presenting arguments others can sincerely be expected to share by seeing the reasons for themselves, but there is no need to demand, or reason to suspect, that individuals could apply a rigid and impermeable filter on their own sincere
beliefs. Ultimately individuals can be free to introduce whatever arguments they like, even the most outlandish, because only arguments which respect equality by embodying a genuinely universal interest will be accepted by all interested parties, in a reflective procedure that approximates ideal conditions. While we could never expect these procedures to produce full ethical agreement on what capabilities make up the good life, norms can be upheld, decisions can be taken, structures and institutions can be created and maintained, which embody a genuinely universal interest.

According to this type of account, I claim the main focus of justice becomes the establishment of the socio-political conditions that must obtain in order for there to exist a space wherein the force of the better argument trumps actual material force and a mechanism by which the rational outcomes of this space determine the material implementation of decisions. Within this family of communicative approaches I suggest Nancy Fraser’s analysis articulates, particularly accurately, the social and political conditions for communicative rationality to function. Thus, in the next section, I will appeal to Fraser’s specific account of participatory parity simply as a starting point for the purpose of making a final point arising from a consideration of what I see as the conditions for any moral rationalism based in reflective procedures—a version of which I see myself as defending.

2.4) A Final Note

Fraser argues that the condition for communicative rationality to function is the participatory parity of everyone to whom the norms and procedure under discussion are to apply. Like the other approaches I have been discussing, I argue the indispensable linchpin of Fraser’s theory is a commitment to the equal moral worth of all human beings. Her
conception of justice is really just an attempt to further flesh out the implications of that equality. The result of this derivation, on the level of moral theory, is the notion of “participatory parity.” She argues that the moral equality of humanity entails the real freedom to participate on par with others in social life. Moving into the realm of social theory, she then argues that the ability of some members of society to participate on par with others can be negatively impacted in two analytically distinct ways—either by being denied certain necessary resources or the requisite social standing. Fraser argues that there are, in fact, two mutually irreducible, yet practically intertwined ordering dimensions that give rise to these two ways of producing injustice. First there is maldistribution rooted in economic system mechanisms and second, misrecognition rooted in institutionalized patterns of cultural value. Fraser incorporates both dimensions of justice under the normative principle of participatory parity. She formalizes this by stipulating two conditions for participatory parity to obtain—the objective condition that individuals have sufficient material resources to ensure their independence and “voice” and the intersubjective condition that institutionalized cultural values offer equal respect and opportunity for esteem. The possibility for moral-rational development will have to depend in no small measure on these conditions.

The point I want to make about this is the following: I think the fact that participatory parity requires these well-being achievements reveals a sort of circularity. Rational reflective procedures, in which we jointly participate in making decisions that enhance the common good by fostering and promoting what we find to be valuable human

---

capabilities for all, requires participatory parity, or some reasonable approximation thereof. In order for participatory parity to obtained, however, requires that participants have already achieved a significant level of flourishing in terms of capabilities. On the down side this entails a negative feedback loop in which unequal flourishing leads to unequal participation in rational decision making procedures which leads to deeper inequality, and so on. On the other hand it also entails a positive feedback according to which greater equality in terms of capabilities leads to a closer approximation of participatory parity, which leads to more fair rational decision making procedures that embody more perfectly the universal interests of participants, which leads to even deeper equality in terms of valued capabilities, and so on.

At the outset of this thesis I expressed my worry that the world is an irrational place and have argued that human development is realizable through a process of making our world a more rational place—a fairer place in which a greater degree of human flourishing is possible. This development, I submit, will consist primarily in keeping this feedback moving in the positive direction by working towards the realization of more equality in terms of capabilities—which is necessary for the participatory parity that makes rational reflective and decision making procedures fair and their outcomes legitimate—and by participating in the ongoing activity of justifying the norms which guide our collective forms of life, without allowing any privileged, elite group to simply assert its will over any other by resorting to irrational force.

To sum up, in Section 1 I summarized the benefits and difficulties associated with some of the most important developments in moral rationalism. In Section 2 I proposed a framework which aims to incorporate the best of this tradition. I argued that many of the
key insights of Aristotle, Kant, and in some ways Hume, have already been incorporated into certain, important contemporary approaches on which I then draw. As I see it, my framework sets out on a Rawlsian tack by trying to explain how a moral-political framework based on a form of reason can allow for a stable environment in which free and equal individuals can flourish despite the fact of pluralism. In essence my proposal is a version of the capability approach which, at its normative core, turns to a version of discourse ethics to provide a procedural account of how we arrive at an understanding of human nature and determine what capabilities are important. I admit that this is a rough outline rather than a completely worked out account of how practical reason can address moral concerns. I see this as a first step in a developing project, many details of which remain to be worked out, and many others which I hope, and expect, can be revisited and refined in light of criticism and debate.
Notes: Chapter 6

4 Fromm, 32.
5 Nancy Fraser & Axel Honneth, Redistribution or Recognition? A Political-Philosophical Exchange (New York: Verso, 2003), 231.
6 Fraser, 218.
7 Fraser, 36.
Bibliography


