Public Unreason: Essays on Political Disagreement

by

Aaron James Ancell

Department of Philosophy
Duke University

Date:_______________________

Approved:

___________________________
Walter Sinnott-Armstrong, Supervisor

___________________________
Allen Buchanan

___________________________
Wayne Norman

___________________________
David Wong

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy in the Graduate School of Duke University

2017
ABSTRACT

Public Unreason: Essays on Political Disagreement

by

Aaron James Ancell

Department of Philosophy
Duke University

Date:_______________________

Approved:

___________________________
Walter Sinnott-Armstrong, Supervisor

___________________________
Allen Buchanan

___________________________
Wayne Norman

___________________________
David Wong

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy in the Graduate School of Duke University

2017
Abstract

Why is political disagreement such a persistent and pervasive feature of contemporary societies? Many political philosophers answer by pointing to moral pluralism and the complexity of relevant non-moral facts. In Chapter 1, I argue that this answer is seriously inadequate. Drawing on work from psychology, political science, and evolutionary anthropology, I argue that an adequate explanation of political disagreement must emphasize two features of human psychology: tribalism and motivated reasoning.

It is often assumed that disagreements rooted in bias and irrationality can be ignored or idealized away by philosophers developing *ideal theories*, that is, theories that aim to sketch the normative outlines of an ideal society. In Chapters 2 and 3, I argue that this assumption is mistaken because even ideal theories are subject to constraints, and idealizing away disagreements rooted in certain kinds of bias and irrationality violates these constraints.

In Chapter 4, I turn to the ethics of political compromise, focusing specifically on compromises that involve making serious concessions to injustice. I consider and attempt to reconcile two seemingly inconsistent approaches to evaluating such compromises: one that emphasizes fundamental moral principles versus one that emphasizes pragmatic considerations.
Dedication

In memory of my dad.
Contents

Abstract.................................................................................................................................................iv

Acknowledgements ................................................................................................................................. x

Introduction .................................................................................................................................................. 1


   1.1 Divergent Values ................................................................................................................................. 6

   1.2 Disagreement About Non-Moral Facts ............................................................................................ 14

   1.3 The Divergent Judgments Account ................................................................................................. 18

   1.4 Why does it matter why people disagree? ......................................................................................... 20

   1.5 Features of Political Disagreement ................................................................................................. 24

      1.5.1 Pervasive ................................................................................................................................... 24

      1.5.2 Expansive .................................................................................................................................. 25

      1.5.3 Persistent .................................................................................................................................. 27

      1.5.4 Obstinate .................................................................................................................................. 28

      1.5.5 Can the Divergent Judgments Account explain these features? ........................................... 29

      1.5.6 Antagonistic .............................................................................................................................. 30

      1.5.7 Clustered .................................................................................................................................. 33

      1.5.8 Polarized .................................................................................................................................. 38

      1.5.9 Where does this leave the Divergent Judgments Account? ..................................................... 42

   1.6 Political Ignorance ........................................................................................................................... 43

      1.6.1 Disagreement About Well-Established Facts ......................................................................... 44
1.6.2 Do people know what they are disagreeing about? ........................................ 49
1.6.3 The Illusion of Explanatory Depth ................................................................. 52
1.6.4 Political Ignorance as a Problem for the Divergent Judgments Account .... 54
1.7 Party Before Policy ............................................................................................. 55
1.8 Political Tribalism ............................................................................................... 65
1.8.1 Tribalism and Social Identities ...................................................................... 65
1.8.2 Tribal Politics .................................................................................................. 69
1.9 Motivated Reasoning ......................................................................................... 77
1.10 Conclusion: The Tribal Rationalizers Account ............................................. 85
1.10.1 The Advantages of the Tribal Rationalizers Account ............................... 86
1.10.2 The Limits of My Arguments ..................................................................... 89
2. Utopianism and Political Irrationality ............................................................... 91
2.1 The Charge of Utopianism ............................................................................... 94
2.2 Probability, Possibility, and Hopeless Aspirational Theory ............................ 97
2.3 The Limits of Estlund’s Argument ................................................................. 102
2.4 Estlund Against Epistocracy .......................................................................... 105
2.5 Thinking Straight is not like Dancing like a Chicken ..................................... 111
3. The Fact of Unreasonable Pluralism ................................................................. 117
3.1 The Fact of Reasonable Pluralism ................................................................... 123
3.2 The Sources of Unreason ............................................................................... 127
3.2.1 How the Sources of Unreason Make Us Unreasonable ............................. 128
3.2.2 Two Types of Unreasonableness ............................................................... 131
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3</td>
<td>The Limits of Ideal Theory</td>
<td>134</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Ideal Theory and Unreasonable Pluralism</td>
<td>137</td>
</tr>
<tr>
<td>3.3.2</td>
<td>The Constraints of a Realistic Utopia</td>
<td>139</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Applying the Constraints</td>
<td>142</td>
</tr>
<tr>
<td>3.4</td>
<td>Can we Overcome the Sources of Unreason?</td>
<td>143</td>
</tr>
<tr>
<td>3.4.1</td>
<td>The Historical Record</td>
<td>145</td>
</tr>
<tr>
<td>3.4.2</td>
<td>The Pitfalls of Human Reason</td>
<td>148</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Am I an evoconservative?</td>
<td>152</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Conclusion and Implications</td>
<td>154</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Implications for Ideal Theory</td>
<td>155</td>
</tr>
<tr>
<td>3.5.2</td>
<td>The Stability of a Liberal Society</td>
<td>158</td>
</tr>
<tr>
<td>4.1</td>
<td>Making Dirty Compromises with Clean Hands</td>
<td>163</td>
</tr>
<tr>
<td>4.1</td>
<td>What is a dirty compromise?</td>
<td>164</td>
</tr>
<tr>
<td>4.2</td>
<td>An Example of a Dirty Compromise</td>
<td>172</td>
</tr>
<tr>
<td>4.3</td>
<td>Three Problems for Dirty Compromises</td>
<td>174</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Are dirty compromises ever morally acceptable?</td>
<td>175</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Do dirty compromises undermine integrity?</td>
<td>180</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Do dirty compromises produce dirty hands?</td>
<td>185</td>
</tr>
<tr>
<td>4.4</td>
<td>How to Evaluate a Compromise</td>
<td>188</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Evaluating the Content of a Compromise</td>
<td>192</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Evaluating Compromises as Actions</td>
<td>199</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Compromising with Integrity</td>
<td>203</td>
</tr>
</tbody>
</table>
4.4.4 Keeping Clean Hands ................................................................. 207
4.5 Harder Cases and Further Issues ..................................................... 209
References ......................................................................................... 215
Biography ......................................................................................... 231
Acknowledgements

I wrote this dissertation, but I am not the only one who deserves credit for its existence and for whatever positive merits it has. I have been helped, encouraged, and inspired along the way by my family, friends, colleagues, teachers, and mentors. Over the past six years I have accrued more debts, both personal and intellectual, than I can recount here. A few that should not go unmentioned are: to my parents for their unwavering support; to Devin Gibbs for believing in me more often than I believe in myself; to Paul Henne for sharing both good times and bad (and for the Sugar Ray poster); to my friends who made Durham feel like home; to Walter Sinnott-Armstrong for his guidance, tireless support, and comments on countless drafts; to Allen Buchanan and David Wong for their graduate seminars that inspired this project, and for their guidance, feedback, and support as members of my committee; to Wayne Norman for the music, the philosophy, and for always coming through for me; and to my colleagues in the Department of Philosophy, the Moral Attitudes and Decisions Lab (MAD Lab), the Rethinking Regulation Graduate Student Working Group, and the Political Theory Reading Group from whom I have learned more than any classes could ever teach.

Pieces of this dissertation were presented at the University of Copenhagen’s 2015 “Compromise and Disagreement” conference, the 2016 North Carolina Philosophical Society Annual Meeting, the 2016 Canadian Philosophical Association Annual Congress, and the 2017 American Philosophical Association Central Division Meeting. I am grateful
for the many questions and comments raised by audiences that pushed me to think more carefully, more broadly, and more charitably.

The first four years of my doctoral studies were funded in part by a Doctoral Fellowship from the Social Sciences and Humanities Research Council of Canada. I am also grateful to donors to the Robert K. Steel Family Graduate Fellowship Fund and the Katherine Goodman Stern Endowment Fund that provided me with two summer research fellowships and a dissertation fellowship.
Introduction

Under what conditions should it be legal to have an abortion? Will cutting corporate taxes benefit the middle class by spurring job creation? Which bathroom should transgender people use? Is healthcare a right or a privilege? To what extent is climate change caused by human activity? What should we do about civil wars in other countries that are displacing millions of people? Are genetically modified foods safe? Should it be legal to privately own an assault rifle? Does affirmative action constitute reverse racism? Should we allow oil pipelines to be constructed in delicate ecosystems? Which economic inequalities are unjust? Why are so many people of color being killed by police officers? Do free-trade agreements hurt blue-collar workers? Should hate speech be outlawed? What kind of restrictions should we have on immigration? These are just a few of the questions that mark the battle lines of contemporary democratic politics. They are the subjects of editorials, presidential debates, protest movements, Twitter fights, and heated arguments around holiday dinner tables. Why is it so hard for us to agree about the answers to such questions?

Many political philosophers answer by pointing to deeper disagreements about fundamental values and the difficulty of determining relevant non-moral facts. In Chapter 1, I argue that this answer is seriously inadequate because it fails to explain several typical features of political disagreement, and it often gets the causal chain backwards. I contend
that an adequate explanation of political disagreement must give central roles to two features of human psychology: tribalism and motivated reasoning.

When it comes to actual politics, political philosophers are usually willing to grant that people are frequently ignorant, biased, irrational, and dogmatic. But the actual world is not the world that political philosophy aims to describe. Political philosophy is primarily a normative discipline; it aims to tell us how the world should be rather than (just) how it is. To do this, much political philosophy takes the form of describing what an ideal society would be like; a project known as ideal theory. And, it is widely assumed that in the course of pursuing this project philosophers can safely set aside many facts about the very non-ideal world we live in. When it comes to political disagreement, this assumption leads philosophers to focus on the kinds of disagreement they believe would exist in an ideal society composed of ideal citizens, and they assume this excludes disagreements rooted in bias, irrationality, or other such unsavory causes.

In Chapters 2 and 3, I argue that this line of reasoning is mistaken. If ideal theory is to avoid being utopian (in the pejorative sense) it must work within constraints imposed by human nature and the world we inhabit. Moreover, these constraints are necessary in order for prominent ideal theories of liberalism and democracy to get off the ground; without them, many of the problems liberalism and democracy are supposed to be responses to could simply be idealized away within our theories. Having shown the
necessity of some constraints on ideal theory, I then argue that idealizing away disagreements rooted in certain kinds of bias and irrationality violates these constraints. This is because these kinds of bias and irrationality are deeply rooted in human psychology and very difficult to overcome.

In Chapter 4, I turn to one of the most important methods of dealing with political disagreement: compromise. I focus specifically on a troubling class of compromises I call *dirty compromises*: compromises that involve making serious concessions to injustice. I consider two seemingly inconsistent lines of thinking about such compromises. The first focuses on fundamental values and moral principles and concludes that all such concessions are morally unacceptable because they violate fundamental principles of justice. The second focuses on pragmatic considerations and the range of realistic alternatives and concludes that some such concessions are morally acceptable because they are part of the best available means of improving on the status quo. I argue that each side gets something right, and that these seemingly inconsistent lines of thought can be reconciled by recognizing that each is suited to a different target of evaluation.

These four chapters were originally conceived to stand as independent pieces, but I have edited them to highlight the connections between them and to remove redundancy. Chapter 4 can still be read independently, but Chapters 2 and 3 now refer to discussions in previous chapters.

Why do we disagree so incessantly about politics? Political philosophers often answer by pointing to deeper disagreements about fundamental values and relevant non-moral facts. In this chapter, I examine these explanations and find them wanting. The problem is not that they are wholly false, but rather that they are inadequate and often backwards. They are inadequate because they fail to explain some of the most salient and troubling features of political disagreement, and they are often backwards because deeper disagreements about values and non-moral facts are often consequences, rather than causes, of the political disagreements they are supposed to explain. I argue that an adequate explanatory account of political disagreement must include at least two additional factors, political tribalism and motivated reasoning, that are absent from most contemporary philosophical accounts.

I begin by describing in greater detail the two major explanations commonly given by political philosophers. I then combine these explanations into a single account of political disagreement that I call the Divergent Judgments Account. After a brief interlude to explain why an adequate account of the causes of political disagreement is important for the purposes of political philosophy, I begin setting out my case against the Divergent
Judgments Account. In Section 1.5 I describe several salient features of political disagreement and argue that the Divergent Judgments Account falls short of explaining three of them: antagonism, clustering, and polarization. I then turn to two major bodies of empirical literature which further reveal the inadequacy of the Divergent Judgments Account. The first is the extensive literature on political ignorance, and the second is a recent wave of evidence suggesting that people’s political views are typically caused by their partisan allegiances, rather than the reverse. Sections 1.8 and 1.9 then explain the two key factors, partisan tribalism and motivated reasoning, that I contend are necessary pieces of any adequate explanation of political disagreement. These factors are central to an alternate account of political disagreement that I call the Tribal Rationalizers Account.

In order to avoid misunderstandings, it is useful to emphasize at the outset that I do not claim that all political disagreement is best explained by the Tribal Rationalizers Account. Moreover, the Divergent Judgments Account and the Tribal Rationalizers Account are not mutually exclusive. Many political disagreements are best explained by an amalgam of the two accounts, and even an amalgam of the two accounts is not sufficient to fully explain all political disagreements. Political disagreement is massively complex and heterogeneous, and no account that can be summarized in a single chapter—let alone a few pages of a single chapter—is sufficient to explain all of it. My primary aim here is simply to highlight the disconnect between the Divergent Judgments Account—
i.e. the explanations of disagreement given by many political philosophers—and the actual political disagreements that pervade contemporary societies.

1.1 Divergent Values

At least since the rise of the deliberative democracy and public reasons paradigms in the 1980s and 1990s, it has been common for political philosophers to explain political disagreement by pointing to deeper disagreements about fundamental principles and values. Amy Gutmann and Dennis Thompson, for example, open their landmark book *Democracy and Disagreement* by announcing that, “Of the challenges facing American democracy today, none is more formidable than the problem of moral disagreement. Neither the theory nor the practice of democratic politics has so far found an adequate way to cope with conflicts about fundamental values.”

Similarly, in *Democracy and Moral Conflict*, Robert Talisse writes, “We are divided over our most fundamental moral commitments. We disagree about moral basics, and accordingly disagree about the precise shape that our politics should take.” Then there is John Rawls and the many philosophers who have followed him in emphasizing what Rawls called *the fact of...

---

reasonable pluralism: the fact that people living in a free society will inevitably come to hold diverse and competing moral and religious worldviews.5

What does it mean to say that people disagree about “moral basics” and “fundamental values”? One possibility is that people have different sets of values such that some people are committed to fundamental values that others do not share. Political rhetoric often frames disagreements in this way. People say things like, “If you support laws limiting offensive speech, then you hate freedom,” and “Anyone who opposes affirmative action doesn’t care about equality.” Such accusations suggest that some fundamental value such as freedom or equality is absent from the moral framework of the opposing party. This is rarely true. Almost no one genuinely hates freedom or has no concern for equality. Indeed, the fundamental values at stake in most political disagreements—values like freedom, equality, security, privacy, human wellbeing, and so on—are shared by almost everyone.

This is not to say that there are no cases of disagreement in which one side does hold a fundamental value that the other does not. Consider, for example, the value of religious faith. At the 2012 Republican National Convention, Senator Marco Rubio

3 John Rawls, Political Liberalism, Expanded ed. (New York: Columbia University Press, 2005), xvi. As Jonathan Quong notes, the idea of reasonable pluralism lies at the heart of the public reasons paradigm and, “the most influential conception of such pluralism or disagreement is Rawls’s account of the burdens of judgment and the subsequent fact of reasonable pluralism.” See Jonathan Quong, “Public Reason”, Summer 2013 ed., The Stanford Encyclopedia of Philosophy (2013).
proclaimed that, “faith in our Creator is the most important American value of all.”\footnote{“Marco Rubio’s Rnc Speech (Text, Video),” Politico, accessed September 14, 2017. http://www.politico.com/story/2012/08/marco-rubios-rnc-speech-text-video-080493.} Contrast this with the words of Richard Dawkins, a trenchant atheist, who writes, “I think a case can be made that faith is one of the world’s great evils, comparable to the smallpox virus but harder to eradicate.”\footnote{Richard Dawkins, “Is Science a Religion?,” The Humanist Jan./Feb. (1997): 26.} Rubio and Dawkins clearly disagree about the value of faith, and we can aptly describe this disagreement by saying that that faith is a fundamental value for Rubio but not for Dawkins. I will call cases of this type cases of divergent value sets.

The vast majority of political disagreements are not cases of divergent value sets. As David Wong says, “it is seriously misleading to characterize moral disagreement as typically involving a difference in the ultimate principles held.”\footnote{David Wong, Natural Moralities: A Defense of Pluralistic Relativism (New York: Oxford University Press, 2006), 247. Wong argues that most differences between moralities are better explained by what I call divergent value priorities. I think Wong is correct that such divergent priorities explain moral differences of the kind he focuses on most heavily, for example, differences between traditions of moral thought such as Confucianism which emphasizes inter-personal relationships and personal duties as opposed to more Western systems that place greater emphasis on individuals and impersonal duties. It should also be noted that Wong does not contend that all moral and political disagreements are best explained by divergent values or other epistemic factors. Indeed, in recent work he offers a much richer picture that points to several of the factors I highlight here including tribalism and motivated reasoning. See David Wong, “On Valuing Moral Disagreement,” (unpublished manuscript).} Even in cases that involve deeper disagreements about moral principles and values, such as disagreements
about the permissibility of abortion, it is not typically the case that one side holds a fundamental value that the other does not. It’s not as if people who are pro-choice have no concern for the value of life, nor is it true that people who are pro-life have no concern at all for women’s rights to control over their own bodies (though activists on both sides often try to frame the debate in such terms). If disagreement about fundamental values explains more than a small minority of political disagreements, it must be because people disagree about fundamental values in some other way.

A second kind of disagreement about fundamental values is far more common: disagreement about how prioritize or trade off such values in cases where they conflict. I will call such disagreements cases of divergent value priorities. Many political philosophers point to divergent value priorities to explain political disagreement. Gutmann and Thompson, for example, say that “the problem of moral conflict originates not only between persons but between values themselves.” They explain that, “Finding the right resolution becomes more difficult when moral values conflict, and conflict among values readily turns into a conflict among persons, as citizens come to different conclusions about the same decisions and policies.” William Galston also points to competing values as a primary cause of political disagreement. He writes:

7 Gutmann and Thompson, Democracy and Disagreement, 23.
8 Ibid., 24.
During policy disputes, the contending forces usually appeal to competing goods and principles, each of which has prima facie moral weight. In many cases, there is no obvious way of reducing these diverse considerations to a common measure of value or of giving one kind of claim lexical priority over others. The most difficult political choices are not between good and bad, but between good and good.9

Similarly, Rawls writes, “In being forced to select among cherished values, or when we hold to several and must restrict each in view of the requirements of the others, we face great difficulties in setting priorities and making adjustments. Many hard decisions may seem to have no clear answer.”10

A metaethical debate lurks in the background here. Some philosophers, Wong and Galston among them, are value pluralists who contend that the reason that it is difficult to adjudicate conflicts between different values is that values themselves are “qualitatively heterogeneous and cannot be reduced to a common measure of value.”11 This contrasts with the view of value monists, such as utilitarians, who believe that all values can ultimately be reduced to a common measure such as pleasure, preference-satisfaction, or well-being. Other philosophers, like Rawls and Gutmann and Thompson, are agnostic.

10 Rawls, Political Liberalism, 55.
11 Galston, Liberal Pluralism, 30. Or as Wong puts it, “there exists an irreducible plurality of basic moral values, where such values are not derivable from or reducible to other moral values.” Wong, Natural Moralities, 6. For a classic statement of value pluralism, see Isaiah Berlin, “Two Concepts of Liberty,” in Four Essays on Liberty, ed. Isaiah Berlin (Oxford: Oxford University Press, 1969).
Gutmann and Thompson write, “We do not know whether, if we enjoyed complete understanding, we would discover uniquely correct resolutions to problems of incompatible values... But we should be able to recognize that all of us lack that kind of understanding at present.”\textsuperscript{12} For present purposes, this metaethical issue can remain in the background. We can consider the extent to which political disagreement is explained by divergent value priorities while remaining neutral about whether values themselves are incommensurable.

In addition to divergent value sets and divergent value priorities, there is a third kind of disagreement about fundamental values that I will call \textit{divergent value specifications}. In such cases, people share some fundamental value at an abstract level, but disagree about how the value is best interpreted, applied, or specified.\textsuperscript{13} The value of equality is a good example here. At an abstract level, almost everyone believes that equality is morally important. But people disagree about what equality means and about what is to be equalized. Should we have equality of outcomes? Equality of opportunities? Which


\textsuperscript{13} Christopher McMahon has a sophisticated account of political disagreement based on divergent value specifications. In his view, political disagreements are rooted in different understandings of moral concepts, which are in turn rooted in different understandings of the meaning of moral terms. See Christopher McMahon, \textit{Reasonable Disagreement: A Theory of Political Morality} (New York: Cambridge University Press, 2009).
opportunities? And among whom must the opportunities be equal? Does equality require uniform treatment, or does it require compensating for undeserved disadvantages in starting positions and natural abilities? Disagreement about the answers to questions like these seems to underlie disagreements about more concrete issues like whether affirmative action increases equality by correcting for discrimination, or diminishes it by holding people to different standards.

Put together, divergent values sets, divergent value priorities, and divergent value specifications provide ample scope for disagreement about fundamental principles and values. And it is certainly true that people on opposing sides of political disagreements often appeal to different values, prioritize values differently, or have different ways of specifying or applying the same value. But what is at stake here is a causal claim: that these deeper disagreements cause political disagreements. Consider, for example, how Gutmann and Thompson explain disagreement about the permissibility of abortion. They write:

[Pro-life and pro-choice advocates] arrive at radically different conclusions about abortion because they cannot agree on whether the fetus is a full-fledged constitutional person, whether a woman’s right to control over her body takes priority over any claims the dependent fetus may have, and what responsibility a woman has to realize the human potential of a fetus that lacks consciousness and sentience.14

14 Gutmann and Thompson, Democracy and Disagreement, 74. My emphasis.
Notice the causal claim: people disagree about abortion because they disagree about deeper moral issues about personhood and the priority of various rights and obligations. Indeed, Gutmann and Thompson suggest that disagreement about abortion is inevitable because, “short of imagining a radically different world... citizens [will] continue to face a fundamental conflict between the life of the fetus and the liberty of the woman.”

I have no doubt that the persistence of disagreement about the permissibility of abortion stems in part from the difficulty of resolving underlying questions about the moral status of the fetus and the relative weights of the right to life and the right to bodily autonomy. But I also suspect that the causal chain frequently runs the other way: people disagree about the moral status of the fetus and about the relative weights of the rights involved because they antecedently disagree about whether abortion is morally permissible. More generally, whereas political philosophers typically point to deeper disagreements about fundamental values as a cause of political disagreement, I think we need to ask to what extent the reverse is true. That is, we need to ask to what extent these deeper disagreements are actually effects, rather than causes, of political disputes. In the latter half of this chapter, I survey some empirical evidence suggesting that the deeper disagreements that look like the causes of political disagreements are very often...

15 Gutmann and Thompson, Democracy and Disagreement, 24.
consequences of political disagreements that are caused, first and foremost, by differing partisan allegiances. But for now, I turn to the other common explanation of political disagreement: disagreement about non-moral facts.

1.2 Disagreement About Non-Moral Facts

In addition to moral values, political issues also involve a vast range of non-moral facts. As Arthur Lupia writes:

The number of facts that can be relevant to the operations of government is infinite. This number includes not just facts about laws but also facts about economics, history, and constitutions that can influence our decisions and expectations. This number also includes facts about how laws affect different types of people, facts about how these effects will change over time, facts about how people feel about these laws, and facts about whether these feelings might produce pressure to implement different laws in the future.16

To Lupia’s list we could add much else including facts about the environment, the impact of new technologies, the causes of various social problems, and the platforms and records of various politicians and parties. The second standard explanation of political disagreement given by political philosophers is that people disagree about such non-moral facts.

Philosophers seeking to downplay the extent of fundamental disagreements about values are especially prone to point to disagreements about non-moral facts as the primary cause of disagreement. Hélène Landemore, for example, writes that disagreement about fundamental values “should not be blown out of proportion” and that “many functional democracies rest on a basis of shared values.” She suggests that much of what looks like disagreement about fundamental values in fact arises from disagreement about non-moral facts. She writes:

A lot of apparent “value pluralism”—for example, the disagreement between Democrats and Republicans over the legitimate size of government—can arguably be explained by a disagreement about facts, including complicated facts such as the causal relationship between big government and efficient spending in a given social and political context. This disagreement about the facts of the world leads to a disagreement about political principles that are dependent on these facts.

I have argued elsewhere that Landemore underestimates the potential for genuine moral disagreement among citizens who share fundamental values, but for now her more basic point is well taken: a lot of political disagreement appears to hinge on disagreement about non-moral facts rather than divergent fundamental values.

18 Ibid., 216.
But why do people disagree about non-moral facts? We live in an age of unprecedented access to information. Why can’t we just look up the answer to any question of fact that might be in dispute? More generally, factual questions seem to be resolvable in a way that moral questions are not. Socrates notes this in one of my favorite passages of Plato:

SOCRATES: What are the subjects of difference that cause hatred and anger? Let us look at it this way. If you and I differ about numbers as to which is the greater, would this difference make us enemies and angry with each other, or would we proceed to count and soon resolve our difference about this?
EUTHYPHO: We would certainly do so.
SOCRATES: Again, if we differed about the larger and the smaller, we would turn to measurement and soon cease to differ.
EUTHYPHO: That is so.
SOCRATES: And about the heavier and the lighter, we would resort to weighing and be reconciled.
EUTHYPHO: Of course.
SOCRATES: What subject of difference would make us angry and hostile with each other if we were unable to come to a decision? Perhaps you do not have an answer ready, but examine as I tell you whether these subjects are the just and the unjust, the beautiful and the ugly, the good and the bad. Are these not the subjects of difference about which, when we are unable to come to a satisfactory decision, you and I and other men become hostile to each other whenever we do?
EUTHYPHO: That is the difference, Socrates, about those subjects.20

Here Socrates suggests that disagreements about values are much harder to resolve than disagreements about matters of fact. In the case of factual disagreements, Socrates

suggests, we can simply check the facts. We can measure, or count, or weigh, or—in this day and age—look it up on our smartphones. So why do we disagree so much about the non-moral facts relevant to political issues?

The standard answer is that the relevant facts are often difficult to know. Rawls, for example, points to the fact that the evidence bearing on them is often “conflicting and complex, and thus hard to assess and evaluate.”21 Consider the debate about capital punishment in the United States. People on either side of this debate often disagree about whether the death penalty deters violent crime. And the evidence about whether it does so is indeed mixed.22 So it is plausible that the difficulty of determining whether the death penalty deters violent crime is one of the primary causes of disagreement about the merits of the death penalty.

However, just as with deeper disagreements about values, we need to ask which way the causal chain runs. The conventional view is that people disagree about political issues because they disagree about relevant non-moral facts—that disagreement about whether the death penalty deters violent crime leads to disagreement about the merits of the death penalty. Or, in Landemore’s example, disagreement about the effectiveness and

21 Rawls, Political Liberalism, 56.
efficiency of government interventions leads to disagreement about the legitimate size of government. But it could be the reverse; it may be that many disagreements about relevant non-moral facts are not causes of the political disagreements they are supposed to explain, but rather effects. Later I will present evidence that suggests this is often the case; people come to disagree about the merits of policies or politicians first, and then they interpret the non-moral facts in whatever way fits these pre-existing views.

1.3 The Divergent Judgments Account

Combining the these two widely invoked explanations of political disagreement—divergent values and disagreements about non-moral facts—yields what I call the Divergent Judgments Account of political disagreement. This account runs roughly as follows:

People’s political views are grounded in their moral principles and values together with their understanding of the relevant non-moral facts. Political disagreement arises in one, or a combination of, two different ways: First, since people have divergent value sets, divergent value priorities, and divergent value specifications, they often reach different conclusions about which laws, policies, and courses of action are morally desirable, obligatory, or forbidden. Second, people often have different understandings of the relevant non-moral facts because the evidence
bearing on such facts is often complex, conflicting, and open to interpretation. These different understandings of the relevant non-moral facts cause people to disagree about the causes of social problems and about the probable effectiveness of proposed solutions. Disagreement about the moral merits and probable effectiveness of various laws, policies, and courses of action in turn leads to disagreement about the merits of particular parties and politicians who represent different political views.

I believe that this account captures a common sense understanding of the causes of political disagreement and one that underlies much contemporary political philosophy. At the very least, it reflects how many prominent philosophers, such as those quoted in the previous two sections, explain political disagreement.

My primary aim here is to show that this account is seriously inadequate as an explanation of the political disagreements that pervade contemporary democratic societies. I say “seriously inadequate” rather than “false” because I grant that the Divergent Judgments Account at least partially explains at least some political disagreements. However, as a general account of political disagreement, the Divergent Judgments Account is at best very incomplete, and at worst more or less backwards. It is incomplete because it fails to explain several salient features of political disagreement. And it is often backwards because the causal chain frequently runs in the opposite
direction: People begin by identifying with a party or politician, they adopt the political views that correspond to their partisan allegiances, and then they construct justifications for these views after the fact, interpreting values and non-moral facts in whatever way best fits their pre-existing views. I will begin arguing for these claims shortly. But first, let me explain why it matters.

1.4 Why does it matter why people disagree?

Why should political philosophers care about why people disagree about politics? An analogy with medicine is helpful here. Suppose that you go to the doctor complaining of a high fever, headache, and joint pain. The doctor diagnoses the cause of your symptoms as a bacterial infection and prescribes a course of antibiotics. Whether the antibiotics are likely to help depends on whether the doctor correctly diagnosed the cause of your symptoms. If it turns out that the cause of your symptoms is actually malaria, which is caused by protozoa rather than bacteria, then the antibiotics are unlikely to help. In fact, the antibiotics might just make things worse if you end up suffering side effects in addition to the symptoms you already had. The same is true of political disagreement. If we misdiagnose the causes of the political disagreement, then the solutions we propose may not be effective and might even make things worse. So, if our normative political theories are going to offer practical guidance about how to deal with political
disagreements, then it is crucial that those theories be informed by an adequate account of the causes of political disagreement.

In trying to develop an adequate account of the causes of political disagreement, we must be wary of engaging in what Robert Nozick derisively called *normative sociology*: “the study of what the causes of problems *ought to be,*”23 The world might be a better place if political disagreements were overwhelmingly caused in the way that the Divergent Judgments Account says. Certainly, it would be a better place if political disagreements were seldom rooted in prejudice, partisan bias, and irrationality. But these normative claims tell us nothing about the causes of political disagreement in the actual world. Moreover, I suspect that philosophers have a tendency to explain political disagreements in terms of deeper disagreements about moral principles and values because disagreements about moral principles and values are things that the tools of traditional analytic philosophy are well-suited to analyze and help us understand. As the old adage goes, when your only tool is a hammer, everything looks like a nail. Of course, some things are nails, and some political disagreements do stem from deeper disagreements about moral principles and values. But we must not neglect baser causes of political

disagreement like tribalism and irrationality simply because they are unsavory or because they are less amenable to our traditional tools.

Many political philosophers believe that they have a principled reason for ignoring such baser causes of disagreement. These philosophers take themselves to be engaged in ideal theory, that is, in the task of sketching the normative outlines of an ideal society. And it is often assumed that ideal theory need not attend to the causes of political disagreement in our actual, very non-ideal, world. Rather, what matters for ideal theory, it is assumed, are the kinds of disagreements that would arise between the idealized citizens of an ideal society who reason about politics in more or less rational and unbiased ways. In other words, it is only reasonable disagreement that matters for ideal theory. Rawls is explicit about this. He writes:

We might suppose, say, that most people hold views that advance their own more narrow interests; and since their interests are different, so are their views. Or perhaps people are often irrational and not very bright, and this mixed with logical errors leads to conflicting opinions. But while such explanations explain much, they are too easy and not the kind we want. We want to know how reasonable disagreement is possible, for we always work at first within ideal theory.  

For ideal theorists like Rawls, what needs to be explained is how and why political disagreement will exist even in an ideal society. So, even if my claims about the actual

\[\text{Rawls, Political Liberalism, 55.}\]
causes of political disagreement are true, many political philosophers will deny that these claims have any implications for their normative theories.

In the next chapter, I argue that this justification for ignoring baser causes of disagreement is unsound because even ideal theory is subject to constraints of realism, and these constraints preclude idealizing away deeply ingrained features of human psychology such as tribalism and motivated reasoning. But even if my argument in the next chapter fails, my argument in this chapter still bears on the claims of most ideal theorists. This is because most ideal theorists contend that ideal theory provides practical guidance to those of us living in actual, non-ideal, societies.25 Jonathan Quong, for example, writes that ideal theory gives us “powerful reasons to behave in certain ways, and not in others, in our current world”26 and that ideal theory should guide us as we “strive to bring our own political behaviour and institutions closer to ideal conditions.”27 But suppose that the nature and causes of political disagreement in the actual world are very different from the nature and causes of political disagreement in the ideal world specified by an ideal theory. In that case, there is no reason to assume that the strategies and principles that the ideal theory proposes for dealing with disagreement will be good

25 David Estlund is a notable exception. He argues that ideal theories need not provide practical guidance. See David Estlund, “Utopophobia,” Philosophy & Public Affairs 42, no. 2 (2014). I consider some of Estlund’s arguments in the next chapter.
27 Ibid., 158.
strategies and principles for dealing with political disagreement in the actual world. To return to the analogy with medicine; there is no reason to think that treatments developed for chicken pox will be effective treatments for psoriasis. So let’s look at the symptoms and see how well the Divergent Judgments Account fits as a diagnosis.

1.5 Features of Political Disagreement

In this section, I describe several features of political disagreement. These features are not necessary conditions, and I do not claim that all political disagreements have all of the features I describe. My aim is to identify some features that are typical of political disagreement as it exists in contemporary liberal democracies and to examine the extent to which the Divergent Judgments Account can explain these features.

1.5.1 Pervasive

Political disagreement is pervasive in two ways. First, it is found everywhere; there is no country on Earth in which people do not disagree about how their society ought to be run, who ought to rule, what the laws ought to be, and much else that falls within the domain of politics. Second, within a given society political disagreement is a common occurrence rather than a fringe phenomenon. In other words, political disagreement is not just a product of there being a few extremists who dissent from the rest of society; ordinary
people disagree with each other about a vast range of political issues. If you stood on the subway platform of any major city and asked people passing by about the political issues of the day, you would inevitably hear a wide range of divergent opinions.

1.5.2 Expansive

Political disagreement spans a vast range of issues both horizontally and vertically. By “horizontally,” I mean that political disagreement spans numerous different issues; abortion, gun rights, affirmative action, electoral reform, campaign finance, supreme court nominations, monetary policy, tax rates, free trade agreements, and foreign military intervention, just to name a few of the most prominent. By “vertically,” I mean that political disagreement spans many different levels from abstract principles and values to the finer details of particular policies.

Even within a single political issue, there is usually scope for disagreement across multiple facets and levels. Consider, for example, just a few of the many different ways in which it is possible to disagree about the permissibility of the death penalty:

a) We may disagree about whether fundamental moral principles and values prohibit the death penalty under all circumstances.
b) We may agree that the death penalty is sometimes permissible, but disagree about the range of cases in which it is so. For example, we may disagree about whether large-scale drug trafficking should be punishable by death.

c) We may agree that the death penalty is sometimes permissible, but disagree about why it is so. For example, we may disagree about whether it is justified on retributive grounds or because it acts as a deterrent.

d) We may agree that the death penalty is permissible if it acts as a deterrent, but disagree about whether it does so.

e) We may agree that the death penalty is permissible in principle, but disagree about whether it should be allowed in practice because we disagree about the rate of wrongful convictions.

f) We may agree about the rate of wrongful convictions, but disagree about how high the rate of wrongful convictions must be in order to make the death penalty impermissible in practice.

g) We may agree that the death penalty is permissible in principle, but disagree about whether it should be allowed in practice because we disagree about whether the penalty is applied in a racially discriminatory way.

Notice that some of these ways of disagreeing involve deeper moral disagreements, whereas others involve disagreement about non-moral facts, and some may involve both.
This is yet another way in which political disagreement is expansive: it spans both moral and non-moral claims.

1.5.3 Persistent

Political disagreement is persistent both in the sense that political disagreement in general is a constant feature of public life, and in the sense that particular political disagreements often persist for decades or even longer. For example, in the United States, public attitudes about abortion have remained largely unchanged in the four and a half decades since Roe v. Wade.28 Of course, not all political disagreements are so entrenched. Public attitudes do change, sometimes even in the span of just a few years. Public attitudes towards same-sex marriage are an example of this. Just ten years ago, the majority of Americans opposed legal recognition of same-sex marriages, whereas today more than two thirds of Americans support it. Nonetheless, same-sex relationships remain a contentious political issue. Indeed, even today, one in four Americans believes that not

only same-sex marriage, but any “same sex relations between consenting adults,” ought to be forbidden by law.29

1.5.4 Obstinate

Political disagreements are often not just persistent, but obstinate. By this I mean that attempts to resolve them by changing people’s political views typically fail, at least within timeframes shorter than several years. Exchanging arguments and evidence often does little to bring people’s views closer together. In fact, reflecting on arguments and evidence often has the opposite effect; it leads people to adopt more extreme positions, thus driving them further apart.30 I am not saying that arguments and evidence never resolve political disagreements; only that they are typically ineffective. Anyone who has had a debate about politics around a holiday dinner table knows this (which is perhaps why discussing politics in such settings is widely considered taboo: it’s rarely productive, and people just get angry and frustrated).

1.5.5 *Can the Divergent Judgments Account explain these features?*

The Divergent Judgments Account offers plausible explanations of the pervasiveness, expansiveness, persistence, and obstinacy of political disagreement. These explanations run roughly as follows: Political issues are frequently moral issues; they are about what is just and unjust, right and wrong, good and bad. Indeed, most political issues have at least one moral dimension. Given the great latitude for disagreement about moral values, there is great latitude for people to reach different conclusions about the moral dimensions of political issues, and thus to arrive at different and often opposing political views. Moreover, political issues often turn on difficult-to-determine non-moral facts, and the latitude for disagreement about such facts creates even greater latitude for disagreement about politics. This explains why political disagreement is so pervasive and why people disagree about such a wide range of issues. Political disagreement is vertically expansive because there are many different levels at which we may disagree about both moral values and non-moral facts, and thus many different levels at which we may disagree about corresponding political issues. Political disagreement is persistent and obstinate because specifying values and resolving conflicts between them is very difficult, as is determining many of the non-moral facts relevant to political issues. The ongoing exchange of reasons and evidence often does little to settle disputes about these
underlying moral and factual matters because our best arguments and evidence fall short of being rationally conclusive.

These explanations are all very plausible, and I grant that they are at least partially right. As I’ve said, my objection to the Divergent Judgments Account is not that it is wholly false, but rather that it is inadequate: while it may explain some disagreement (or some features of disagreement), it falls well short of explaining everything that an adequate explanatory account of political disagreement should explain. I now turn to some features of disagreement that seem more difficult to explain in terms of the Divergent Judgments Account.

1.5.6 Antagonistic

When people disagree about politics, they often believe that their opponents are not just wrong, but stupid, ill-intentioned, or even dangerous. In the United States, 70% of Democrats and 52% of Republicans say that members of the opposing party are “more closeminded than other Americans.” Significant minorities on both sides (between 30% and 47%) also say that members of the other party are less intelligent, less honest, and less moral. More than half of people on both sides report that the opposing party makes them feel frustrated, and nearly half also say the opposing party makes them feel angry and
afraid. More than a quarter of Democrats and a third of Republicans say that the other party is not just wrong, but “a threat to the nation’s wellbeing.”

Such antagonism, while perhaps worse than in previous decades, is not entirely new. As we saw in the passage from the *Euthyphro* I quoted in Section 1.2, Socrates observed more than 2000 years ago that justice and morality are “subjects of difference that cause hatred and anger.” Moreover, work by psychologist Linda Skitka and her colleagues suggests that we generally tend to dislike people who have moral and political views that differ from our own. Summarizing this work, Skitka writes, “people do not want to work with, live near, or even shop at a store owned by someone who does not share their morally mandated opinions.” And the stronger people’s convictions, the more intolerant they become:

...as people’s moral convictions in a given attitude domain increases, so too does their intolerance of attitude divergence within that domain. Moreover, people are less willing to share and more willing to behaviorally discriminate against those who do not share their moral point of view.

---

33 Plato, Euthyphro, 7b-d.
35 Ibid., 274.
Of course, not all political disagreements are marked by such hostility and mutual intolerance, but many are, and an account of political disagreement should provide some explanation of why this is so.

Can the Divergent Judgments Account do this? On the one hand, if political disagreement is rooted in divergent values, then we might expect it to be antagonistic. For example, suppose that James and Judy disagree about whether abortion should be legal because they disagree about how to specify the bounds of moral personhood. James believes that moral personhood begins at conception, whereas Judy believes that it begins much later, say, in the late second or early third trimester. To James, abortion is a form of murder—the intentional killing of an innocent person. To Judy, abortion, at least if performed early enough, does not involve killing a person at all. It is easy to see how James might come to feel hostility towards Judy; from his point of view, Judy is advocating legalized murder. Judy too might feel some hostility towards James because, from her point of view, James advocates unjustified limits on women’s control over their bodies. In this case, mutual antagonism is quite intelligible.

On the other hand, if political disagreements are rooted in difficulties we all face—the difficulty of specifying and adjudicating conflicts between values, and the difficulty of determining the relevant non-moral facts—then why do we have such a hard time understanding how intelligent and well-intentioned people could come to very different
political conclusions? Indeed, far from finding political issues difficult, many people seem to think that the answers are obvious and that anyone who denies them must be close-minded, stupid, or evil.\textsuperscript{36} In defending their political views, people often say things like “It’s just plain common sense!”\textsuperscript{37} Why is this? The Divergent Judgments Account does not seem to have an answer.

I think an adequate explanation of this antagonism requires citing factors that are not part of the Divergent Judgments Account. In particular, in section 1.9 I suggest that motivated reasoning is largely to blame: because we interpret and weigh values, facts, arguments, and evidence in systematically biased ways, it often seems to us that our own views are obviously true, even when they are far from it.

\textbf{1.5.7 Clustered}

People’s views about different and seemingly unrelated political issues tend to be correlated. As Michael Huemer observes, “you can often predict someone’s belief about

\begin{flushright}
\end{flushright}

\begin{flushright}
\textsuperscript{37} For a discussion and several examples of such appeals to common sense in political rhetoric, see chapter 4 of Scott F. Aiken and Robert Talisse, Why We Argue (and How We Should): A Guide to Political Disagreement (New York: Routledge, 2014).
\end{flushright}
one issue on the basis of his opinion about some other completely unrelated issue. For example, people who support gun control are much more likely to support welfare programs and abortion rights.”38 In the United States, the extent to which political disagreements break down along basic party lines is striking. The Pew Research Center asked both Democrats and Republicans whether they agreed with the position of their party on seven major issues: “health care policy; policies to deal with illegal immigration; policies to deal with climate change; policies to deal with ISIS, the Islamic militant group in Iraq and Syria; abortion policy; gun policy; and policies to deal with the economy.” They found that about 40% of people on both sides agreed with their party on all seven issues, and a further 30% agreed with their party on at least five out of seven.39 Among people that Pew defines as “highly engaged”—people who vote regularly and have volunteered for or donated to a political campaign at least once—there was even greater conformity along party lines. Within this group, 50% of Republicans and 63% of Democrats agreed with their party on all seven issues, and only about 10% disagreed with their party on more than two issues.40

40 Ibid.
The more general point here is that if two people disagree about abortion, then they probably disagree about gun laws, environmental regulations, and foreign military intervention as well. Moreover, we can predict with accuracy much greater than chance which side they take on those other issues simply by knowing which side they take on abortion. This is striking because these different issues are logically independent. Whether or under what circumstances abortion ought to be legal is logically independent of whether stricter gun laws are permissible or likely to be effective, and both of these issues are independent of whether foreign military intervention is a good idea in any particular case. Yet, people’s beliefs about these three different issues are correlated. Of course, these correlations are far from perfect and they occur at a fairly coarse level. For example, about a quarter of Democrats and nearly a third of Republicans say they disagree with their own party’s stance on abortion. And even among those who agree with their party’s general stance, there is room for disagreement about more specific questions. Nonetheless, an account of political disagreement should explain why, for instance, people who are pro-choice tend to support stricter gun-control, and people who are pro-life tend to oppose stricter gun-control.

Can divergent values explain these correlations? Huemer argues that the answer is “no”. He writes:

On the Divergent Values Theory, we should expect prevalent political belief clusters to correspond to different basic moral theories. Thus, there should be some
core moral claim that unites all or most “liberal” political beliefs, and a different moral claim that unites all or most “conservative” political beliefs. What underlying moral thesis supports the views that (a) capitalism is unjust, (b) abortion is permissible, (c) capital punishment is bad, and (d) affirmative action is just? Here, I need not claim that those beliefs always go together, but merely that they are correlated (if a person holds one of them, he is more likely to hold another of them); the Divergent Values hypothesis fails to explain this.41

While I agree with Huemer that divergent values do not adequately explain the observed correlations between political views, I think that Huemer’s argument for this conclusion is too hasty. The Divergent Judgments Account need not say that there is a single moral claim that unites liberal morality versus conservative morality. It could be that what unites these moralities is a set of value priorities. Indeed, there is some evidence that this is the case.

According to the Moral Foundations Theory developed by Jonathan Haidt and colleagues, moral judgments are based on six intuitive moral foundations that Haidt likens to different taste buds: harm/care, fairness/reciprocity, liberty/oppression, in-group/loyalty, authority/respect, and purity/sanctity.42 Haidt contends that most moral differences between liberals and conservatives are explained by differences in the emphasis that liberals and conservatives place on these different foundations. He writes,

“the two ends of the political spectrum rely upon each foundation in different ways, or to different degrees. It appears that the left relies primarily on the Care and Fairness foundations, whereas the right uses all five.” In other words, what unites liberal morality is a distinctive emphasis on issues related to harm and fairness, whereas conservative morality places less emphasis on these issues, and more emphasis on issues related to loyalty, authority, and purity.

Haidt’s theory might seem to vindicate the Divergent Judgments Account’s claim that political disagreements are largely explained by divergent value priorities. But again, we need to ask which direction the causal chain actually runs. Do people disagree about politics because they prioritize different fundamental values, or do they prioritize different fundamental values because they antecedently disagree about politics? Insofar as it is the latter, the Divergent Judgments Account (and Haidt’s theory) gets things backwards. And as we’ll see, the empirical evidence suggests this is often the case.

Even setting aside the issue of the direction of causation, there are lingering questions about the clustering of political views that Haidt’s account does not answer. For example, liberals tend to support legalized abortion and to provide arguments for this

---

that emphasize women’s freedom of choice. Meanwhile, conservatives tend to oppose legalized abortion (or at least favor greater restrictions on it) and to provide arguments for this that emphasize the harm done to the unborn. But given Haidt’s theory, it seems we should expect the opposite: It ought to be liberals emphasizing harm and conservatives emphasizing liberty. Haidt might respond that the reason that conservatives oppose abortion is because they emphasize the sanctity foundation and so have greater reverence for the sanctity of life. But then why does this concern for sanctity not extend to the sanctity of the environment? Why is it that conservatives tend to want more government interference aimed at preserving the sanctity of life, and less government interference aimed at preserving the sanctity of the environment? What we need is an explanation how and why people frame political issues in terms of the values (or moral foundations) they do. Later I will suggest that the explanation often lies in partisan conformity and motivated reasoning—people frame issues in terms of whichever values get them to the conclusion they want to affirm, and that conclusion is often specified in advance by their partisan loyalties.

1.5.8 Polarized

“Political polarization” can mean many different things, but here I will focus on what I take to be the central meaning: people are increasingly clustered at the poles of the
political spectrum with fewer people falling in the middle between the poles. In one of the most comprehensive reports on trends in public opinion in the United States, the Pew Research Center reports that the number of Americans at the extreme ends of the political spectrum has doubled since 1994, and the number of people in the center has sharply declined. Nowhere is this clearer than in the ideological chasm that has opened between self-identified Democrats and Republicans. Pew reports that:

...this shift represents both Democrats moving to the left and Republicans moving to the right, with less and less overlap between the parties. Today, 92% of Republicans are to the right of the median (middle) Democrat, compared with 64% twenty years ago. And 94% of Democrats are to the left of the median Republican, up from 70% in 1994.

Democrats and Republicans now disagree fiercely about issues about which they used to agree. For example, environmental protection used to enjoy broad bipartisan support, but it is now one of the most contested issues. Consider what Pew calls the “partisan gap”—the difference between the percentage of Democrats and the percentage of Republicans who agree with various statements. When it comes to the environment, Pew reports:

Views on the importance of environmental protection have arguably been the most pointed area of polarization. When these questions were first asked 20 years ago, there was virtually no disagreement across party lines. Even as recently as

---

44 For a rich discussion of several different meanings of “polarization”, see the editor’s introduction in Nathaniel Persily, ed. Solutions to Political Polarization in America (New York, NY: Cambridge University Press, 2015). Persily refers to what I call polarization as hyperpartisanship.

45 Political Polarization in the American Public.
2003, Republicans and Democrats were, on average, only 13 points apart on questions related to the environment. That gap has now tripled to an average of 39 points—one of the largest value gaps in the study.\textsuperscript{46}

Significant partisan gaps have also opened on immigration (from just 4 points in 2002 to 24 points in 2012), the appropriate scope of government (from just 6 points in 1987 to 33 points in 2012), and support for the social safety net (from 23 points in 1987 to 41 points in 2012). More generally, Pew found that the average partisan gap across 48 different issues nearly doubled between 1987 and 2012 going from 10 points to 18 points.\textsuperscript{47} As Nolan McCarthy and his colleagues put it, “In the middle of the century, the Democrats and Republicans did dance almost cheek to cheek in a courtship of the political middle. But over the past forty years, the parties have deserted the center of the dance floor in favor of the wings.”\textsuperscript{48} And it does not appear to be only the U.S. that has become politically polarized. Although good studies are hard to find, there is evidence that the rise of more

\textsuperscript{47} Ibid.
ideologically extreme parties and politicians across Europe is related to increased polarization.49

To explain polarization, the Divergent Judgments Account must say that the underlying disagreements that it takes as the primary causes of political disagreement will grow over time. That is, it must say that disagreement about fundamental values will increase over time, or that disagreement about relevant non-moral facts will increase over time, or that both forms of disagreement will increase over time. But the Divergent Judgments Account does not seem to have the resources to explain why any of these things would be true. Why should we expect increasing disagreement about how to weigh and specify values? Why not expect the opposite? That is, why not expect people to resolve at least some of their disagreements about fundamental values through discussion and debate? And why expect increasing disagreement about non-moral facts, especially if such disagreement is caused by complex and conflicting evidence? Shouldn’t we expect greater convergence on the facts over time as more evidence comes in and tips the balance one way or the other? The Divergent Judgments Account lacks any compelling answers.

to these questions and thus fails to provide any explanation of one of the most salient and troubling features of contemporary political disagreement.

1.5.9 Where does this leave the Divergent Judgments Account?

The Divergent Judgments Account offers plausible explanations of the pervasiveness, expansiveness, persistence, and obstinacy of political disagreement. It is less equipped to explain why political disagreement is frequently antagonistic and why people’s political views about logically independent issues tend to be correlated. And it seems totally unequipped to explain political polarization. This shows that the Divergent Judgments Account is incomplete; there are significant features of political disagreement that it fails to explain. But this does not yet show that the Divergent Judgments Account is false or even seriously inadequate. It may be that explaining certain features of political disagreement requires pointing to factors that are not part of the Divergent Judgments Account but that these additional factors can be added without disturbing the core of the Account. To show that the Divergent Judgments Account is not just incomplete, but seriously inadequate, requires a different kind of evidence. One has to show that the Divergent Judgments Account is often wrong about the causes of political disagreement. The next two sections make the case for this by examining two different bodies of empirical evidence.
1.6 Political Ignorance

According to the Divergent Judgments Account, disagreement about relevant non-moral facts is one of the primary causes of political disagreement, and the reason that people disagree about such facts is that the evidence bearing on them is complex, conflicting, and open to interpretation. So, insofar as the Divergent Judgments Account is the right explanation of political disagreement, we should find two things. First, we should find that the relevant non-moral facts about which people disagree are ones for which the evidence is complex, conflicting, and difficult to evaluate. The Divergent Judgment Account fails to explain disagreement insofar as people also disagree about relevant non-moral facts for which the evidence is clear, unequivocal, and easy to access. Second, insofar as political disagreement is caused by people reaching different conclusions about the relevant non-moral facts, we should find that people have at least a basic and minimally coherent understanding of such facts. Insofar as we find that people do not know even very basic, undisputable, and obviously relevant facts about the policies or politicians about which they express opinions, we have reason to doubt that their opinions are actually grounded in prior assessments of such facts. And insofar as people’s opinions are not actually grounded in prior assessments of the non-moral facts, disagreement about such facts cannot explain why those opinions diverge. In this section, I show that the Divergent Judgments Account faces serious trouble on both of these fronts:
There is widespread disagreement about non-moral facts for which the evidence is clear, unequivocal, and easy to access; and people often have little understanding of even very basic facts about the policies and politicians they express opinions about.

1.6.1 Disagreement About Well-Established Facts

“Post-truth” was the Oxford English Dictionary’s Word of the Year for 2016. As the OED defines it, the term denotes “circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief.”

While the term is new, the phenomenon it describes is not. Almost two decades ago, political scientist James Kuklinski and his colleagues observed that, “Unfortunately, what is true in principle often has little relevance when the discussion turns to what are represented as facts pertinent to public policy.” Indeed, even facts that are easy to verify are widely disputed and many people’s “factual” beliefs about political issues seem to have little basis in reality.

For example, people’s beliefs about key economic indicators like inflation, unemployment, and economic growth often bear little resemblance to the facts. The

problem is not just that people do not know the exact numbers. Rather, the problem is that people’s beliefs about even basic economic trends—e.g. whether unemployment has increased or decreased—seem to be based more on partisanship than reality. The 1988 American National Election Study (ANES) asked people whether inflation and unemployment had gone up or down during President Reagan’s two terms in office. The correct answers—easily found in newspapers and official government announcements—were that unemployment and inflation had both declined. In the case of inflation, the decline was especially dramatic, falling from 13.5% to 4.1%. Yet, when Larry Bartels looked at the 1988 ANES data, he found little agreement on these facts. He writes:

Democrats were strikingly impervious to the good economic news. For example, more than 50% of “strong” democrats claimed that inflation had gotten somewhat worse or much worse over the preceding 8 years, despite the fact that the actual inflation rate had declined by more than two-thirds. Fewer than 8 percent said it had gotten much better.  

Strong Republicans were far more likely to answer correctly: only 13% said inflation had gotten worse, and 47% knew that it had gotten much better. Bartels also found similar (albeit opposite) trends in data from the 2000 ANES survey which asked about the

53 Ibid.
economy under President Clinton: Democrats were far more likely than republicans to correctly answer that the economy had improved and that the budget deficit had shrunk.\textsuperscript{54}

The economy is not the only issue about which people’s beliefs seem to track partisanship more than they track the facts. Consider climate change. Whereas scientists overwhelming say that climate change is largely caused by human activity, only 48% of Americans agree and 20% say there is no evidence that climate change is occurring. This disagreement breaks down along party lines: only 23% of Republicans say that climate change is caused by human activity compared to 69% of Democrats.\textsuperscript{55} Granted, the actual scientific evidence demonstrating that climate change is occurring and largely caused by human activity may be difficult for non-scientists to interpret and understand. But non-technical and easily accessible summaries of this evidence provided by authoritative sources are not hard to find.\textsuperscript{56} Moreover, people do not even agree about the widely-publicized fact that there is a scientific consensus. Only about a quarter of Americans

\textsuperscript{54}Bartels, "Beyond the Running Tally," 137.
believe that “almost all” climate scientists agree that climate change is largely caused by human activity, and more than a third of Americans believe that “fewer than half” of climate scientists agree on this.\(^57\) This is so despite the fact that people could easily learn that there is an overwhelming scientific consensus by, say, reading the Wikipedia article on “Scientific opinion on climate change”, which is one of the top internet search results for “Is there are scientific consensus on climate change?”\(^58\)

Disagreement about the safety of genetically modified (GM) foods is a similar case. Despite a consensus among scientists and health regulators that GM foods are safe, nearly 40% of Americans say that genetically modified foods are “worse for your health than foods with no genetically modified ingredients,” and 20% characterize the health risks posed by GM foods as “high.”\(^59\) There is evidence that much of this disagreement is rooted in basic misunderstandings of how genes work and whether genes are unique to genetically modified foods. In a study by Steve Sloman and Philip Fernbach, “about a

\(^57\) Funk and Kennedy, *The Politics of Climate*.

\(^58\) The article reports, citing numerous sources, that “The scientific consensus is that the Earth’s climate system is unequivocally warming, and that it is extremely likely (meaning 95% probability or higher) that this warming is predominantly caused by humans. It is likely that this mainly arises from increased concentrations of greenhouse gases in the atmosphere, such as from deforestation and the burning of fossil fuels, partially offset by human caused increases in aerosols; natural changes had little effect.” See “Scientific Opinion on Climate Change,” Wikipedia, accessed September 14, 2017. https://en.wikipedia.org/wiki/Scientific_opinion_on_climate_change.

quarter of respondents agreed that ‘a gene inserted into a food can migrate into the genetic code of humans who consume the food.’” And, “another quarter of respondents said they weren’t sure if this weren’t true, but this could be true.”60 Less than half of Americans know that “ordinary tomatoes contain genes,”61 and a survey conducted by the department of agriculture at Oklahoma State University found that 80% of people support mandatory warning labels on all food that contains DNA—in other words, on all food.62 It is difficult to attribute these misunderstandings to complex and conflicting evidence. It seems more likely that many people simply haven’t looked at the evidence.

A final example comes from the “Brexit” referendum on whether the United Kingdom should leave the European Union. Much of the debate leading up to the referendum focused on concerns about immigration, so if people’s views about Brexit are based on the facts, we should find that they have at least a rough idea of the extent of immigration into the U.K. Yet, when asked “What percentage of the population was born in another country?” the average guess, 24.4%, was twice the actual figure, 13%.63 When

61 Ibid., 158.
62 Foods Food Demand Survey (Oklahoma State University Department of Agricultural Economics, 2015), http://agecon.okstate.edu/faculty/publications/4975.pdf.
asked specifically about immigrants from the E.U., the average guess was three times the actual figure. This tendency to overestimate was most extreme among those who supported leaving the E.U.—on average, “Leave” voters overestimated the extent of E.U. immigration by a factor of four. When the pollsters subsequently gave people the official government figures, 46% of the people surveyed said “I still think the proportion is much higher.” Even when presented with clear and unequivocal evidence, nearly half of people persisted in disagreeing with the facts.

1.6.2 Do people know what they are disagreeing about?

Not only do people disagree about well-established facts, but much of the time they do not have even a basic or minimally coherent understanding of the things about which they disagree. Sloman and Fernbach describe one example of this:

In 2012, just after the Supreme Court ruled to uphold key provisions of [the Affordable Care Act], the Pew Research Center conducted a survey asking people whether they approved or disapproved of the ruling. Not surprisingly, responses were strongly divided: 36 percent in favor, 40 percent opposed, and 24 percent expressing no opinion. Pew also asked what the court had ruled. Only 55 percent

65 Some research suggests that giving people clear and unequivocal evidence corrective information can actually reinforce, rather than mitigate, false beliefs about politically relevant facts. See Brendan Nyhan and Jason Reifler, "When Corrections Fail: The Persistence of Political Misperceptions," Political Behavior 32, no. 2 (2010).
responded correctly. Fifteen percent said that the court struck down the law; 30 percent had no idea. So 76 percent expressed an opinion, but only 55 percent knew what they were expressing an opinion about.\footnote{Sloman and Fernbach, 172. For the Pew poll, see Division, Uncertainty over Court’s Health Care Ruling (Pew Research Center, 2012), http://www.people-press.org/2012/07/02/division-uncertainty-over-courts-health-care-ruling/.


Ilya Somin describes another example:

One of the key policy positions staked out by President Obama in his successful 2012 reelection campaign was his plan to raise income taxes for persons earning more than $250,000 per year, an idea much discussed during the campaign and supported by a large majority of the public—69 percent in a December 2012 poll. A February 2012 survey conducted for the political newspaper The Hill asked respondents what tax rates people with different income levels should pay. It found that 75% of likely voters wanted the highest-income earners to pay taxes lower than 30 percent of income, the top rate at the time of the 2012 election.\footnote{For The Hill’s poll, see Peter Schroeder, "Likely Voters Prefer Lower Tax Rates for Individuals and Business,” The Hill, accessed September 14, 2017. http://thehill.com/polls/212643-hill-poll-likely-voters-prefer-lower-tax-rates-for-individuals-business.}

In other words, most people supported raising the top tax rate, but most people also thought the top tax rate should be lower than it already was. This suggests that people’s beliefs about whether the top tax rate ought to be raised were formed without any awareness of what the top tax rate actually was at the time.

Even people’s opinions about the merits of various parties and politicians are frequently unconnected to a basic understanding of what those parties or politicians stand for. For example, Jason Brennan notes that:
In the 2000 U.S. presidential election, significantly less than half of all Americans knew that Gore was more supportive of abortion rights, more supportive of welfare programs, favored a higher degree of aid for blacks, or was more supportive of environmental regulation than Bush.68

Similarly, Gabriel Lenz reports that, “across [more than 15] different policy issues examined… I typically find that only about 50% of voters could correctly place the major candidates or parties to the right of left of each other.”69 So roughly half of all voters seem to have beliefs about the merits of the major candidates and parties that are not based on even a very basic understanding of where those candidates or parties stand on major policy issues.

These are not cherry-picked examples. People typically do not know much about even the most prominent and widely debated political issues of the day, or even about the positions of the parties and politicians they vote for. Indeed, just how little most people know about politics has been the subject of decades of research in political science. Summarizing this research, Michael Delli Carpini writes:

Over fifty years of survey research on Americans’ knowledge of politics leads to several consistent conclusions. The most powerful and influential of these conclusions is that the “average” citizen is woefully uninformed about political

institutions and processes, substantive policies and socioeconomic conditions, and important political actors such as elected officials and political parties.\textsuperscript{70}

Similarly, Benjamin Page and Robert Shapiro write:

…it is undeniable that most Americans are, at best, fuzzy about the details of government structure and policy. They do not know with any precision how much money is being spent on the military, foreign aid, education, or food stamps. They have only a dim idea what is going on in foreign countries or (in many cases) even where those countries are. They do not know much about monetary policy or economic regulation or labor relations or civil rights. Thus it would be unrealistic to expect the average American to hold well-worked-out, firmly based preferences about a wide range of public policies.\textsuperscript{71}

1.6.3 The Illusion of Explanatory Depth

Why are people so willing to express opinions about things they know almost nothing about? Part of the answer seems to be that people consistently overestimate their own understanding, a phenomenon that psychologist Frank Keil named the illusion of explanatory depth. Keil and his colleague Leon Rosenblit developed a method of studying people’s ignorance about their own ignorance by asking them to explain how things work.\textsuperscript{72} In 2012, Sloman and Fernbach applied this method to people’s political views.


about six policy issues that were prominent at the time: (1) whether the U.S. should impose unilateral sanctions on Iran; (2) whether the U.S. should raise the retirement age for social security; (3) whether the U.S. should transition to a single-payer healthcare system; (4) whether the U.S. should institute a cap-and-trade system for carbon emissions; (5) whether the U.S. should institute a national flat tax; and (6) whether the U.S. should institute merit-based pay for teachers. Participants were asked to indicate their position on each issue on a 7-point scale ranging from “strongly against” to “strongly in favor”. Next, they were asked to rate their understanding of the effects that each policy would have. For example, they were asked “How well do you understand the impact of imposing a cap and trade system on carbon emissions?” Participants were then asked to demonstrate this understanding. More specifically, they were asked to explain, in writing, the effects that the policy was likely to have and how the policy would produce those effects.73 The result? “As in most experiments of this kind, participants were pretty bad at generating explanations. With very few exceptions, they simply had very little to say when we asked them to explain how a policy worked.”74 When subjects were asked to rate their understanding of the issue again at the end of the experiment, they reduced their ratings

---

suggesting that they realized that they did not understand the policies as well as they previously thought.

1.6.4 Political Ignorance as a Problem for the Divergent Judgments Account

While people typically believe that their political views are based on an assessment of the relevant evidence and facts, the empirical evidence suggests that this often not the case. People disagree about well-established facts even when the evidence for them is clear, unequivocal, and easy to access. And people are often unaware of even the most basic facts about the issues, policies, and politicians about which they express opinions. This is problematic for the Divergent Judgments Account in two ways. First, since the Divergent Judgments Account attributes disagreements about non-moral facts to complex and conflicting evidence, it fails to explain why people disagree about such facts even when the evidence is clear and unequivocal. Indeed, we saw in the Brexit case that people persist in disputing such facts even when presented with unequivocal and easy-to-understand evidence that ought to be decisive. Second, contrary to the general thrust of the Divergent Judgments Account, many people’s beliefs about the merits of policies, parties, and politicians do not seem to be based on an assessment of the relevant non-moral facts about how those policies would work or about the positions of the parties and candidates. They can’t be, because many people don’t know those facts. Thus, much
political disagreement cannot be explained by people first coming to divergent understandings of the facts and then subsequently disagreeing about the merits of policies or politicians.

1.7 Party Before Policy

Now we come to what I take to be the most significant problem for the Divergent Judgments Account: in many cases (perhaps even most cases), it gets the direction of causation backwards. According to the Divergent Judgments Account, disagreements about the merits of particular parties, politicians, or policies are consequences of underlying disagreements about deeper moral issues or about relevant non-moral facts. If this is true, then we should find that most people’s political views are based on their assessments of these deeper moral issues and non-moral facts. The Divergent Judgments Account gets things backwards insofar as we find the opposite, that is, insofar as people’s beliefs about the deeper moral issues and the relevant non-moral facts are consequences, rather than causes, of their beliefs about the merits of politicians, parties, and policies.

To investigate this, let us begin where the Divergent Judgments Account ends, that is, with disagreements about the merits of particular politicians and parties. The Divergent Judgments Account offers a straightforward explanation of such disagreements: Different politicians and parties espouse different policy platforms.
People decide which politicians and parties to support by determining which of the policy platforms on offer most closely aligns with their own views about various policy issues. Since people have different views about policy issues, they end up supporting different politicians and parties.

Political scientists call this issue voting and it lies at the heart of one of the most influential models of elections, the Downsian model, named for its chief architect, economist Anthony Downs. In its classic form, the Downsian model assumes that possible political policies are arranged on an ideological spectrum from left to right. Individual voters are assumed to enter the political arena with pre-existing policy preferences represented as points on the left-right spectrum. Political parties are also represented by points on the spectrum corresponding to the policies they advocate (i.e. by their political platforms). According to the model, in a two-party system like the U.S., voters will decide which party to vote for by determining which of the parties is closest to them on the ideological spectrum, that is, which party best represents their own policy preferences.

Much of the edifice of modern democracy is based on the idea of issue voting. Politicians and parties release official platforms outlining their policy proposals.

Typically, they highlight one or two distinctive policy proposals to campaign on. Candidates give speeches and run advertisements criticizing their opponents’ policy proposals and singing the praises of their own. During debates, candidates are pressed to explain and defend their positions on various issues. All of this is extensively reported on by the media and analyzed by political pundits. These practices all seem to assume that voters are paying attention to policy issues, taking note of the policy positions of the candidates, and casting their vote based largely on who they believe has the best platform.

But do people actually decide which politicians and parties to support on the basis of their policy preferences? In the vast majority of cases the answer seems to be “no.” This might come as a surprise, especially given the extent to which people’s political views and their partisan allegiances tend to line up. As we saw in Section 1.5.7, most partisans share their party’s view on most issues, and the most committed partisans share their party’s view on almost all issues. It is tempting to interpret this as evidence that people choose which party to support by determining which party most aligns with their own political views. But correlation does not imply causation. Just because people’s political views tend to line up with the views of their party does not imply that their political views cause their partisan allegiance. It could be the other way around; people’s partisan allegiances could drive their political views. Looking only at whether people’s political views tend to match their partisan allegiances does not help us decide between these two hypotheses because
both hypotheses predict a correlation. As political scientist Gabriel Lenz puts it, the two hypotheses are “observationally equivalent.”

To cut through this observational equivalence, Lenz exploited data from panel surveys: longitudinal surveys that ask the same sample of people the same questions over a period of months or years. Lenz examined whether three kinds of shifts—media priming, persuasion, and learning—created corresponding shifts in support for politicians and parties. Media priming occurs when an issue becomes more salient as a result of increased media coverage (which may in turn be a result of deliberate efforts on the part of a political campaign to raise the salience of some issue). Here the question is whether the increased salience of an issue causes people to shift their evaluations of politicians or parties to be more in line with their views about that issue. For example, if social security reform becomes a more salient issue during the campaign, do people then rate politicians who share their position on social security more favorably, and politicians who oppose their position more negatively. Persuasion occurs when voters change their views about policy issues. Here the question is whether people update their evaluations of politicians and parties following a change in their own political views. For example, if someone goes from being pro-union to anti-union, does this cause them to support pro-

76 Lenz, Follow the Leader, 3.
union politicians more and anti-union politicians less. Learning occurs when voters learn the policy positions of politicians or parties. Here the question is whether learning that a party or politician has a particular policy position changes people’s evaluation of that party or politician. For example, does learning that a particular politician supports a single-payer healthcare system increase support for that politician among people who favor a single-payer healthcare system and reduce support for that politician among people who oppose such a system.

Surprisingly, the answer to all three questions is mostly “no”. In fact, rather than adjusting their evaluations of parties and candidates to align with their policy preferences, most people most of the time do the opposite: they change their policy preferences to align with the views of their preferred party or candidate. Lenz writes:

I find surprisingly little evidence that voters judge politicians on their policy stances. They rarely shift their votes to politicians who agree with them—even when a policy issue has just become highly prominent, even when politicians take clear and distinct stances on the issue, and even when voters know these stances. Instead, I usually find the reverse: voters first decide they like a politician for other reasons, then adopt his or her policy views.\(^7\)

One might suggest that this is the result of voters using a heuristic: they know that a particular party agrees with them on many issues or shares their general ideological outlook and they use this as a way to determine what they would think about other issues

\[^7\] Lenz, *Follow the Leader*, 3.
if only they took the time to work through them. Lenz tested this hypothesis and found little support for it. He writes:

…a considerable share of the public exhibits a rather crude form of following. They did not just follow politicians on policy. They also appeared to follow politicians on basic predispositions such as ideology. Furthermore, they followed the views of politicians even when they were ignorant of those politicians’ broader policy outlooks, they followed even when they did not know much about politics in general, and they followed candidates as well as parties.  

The finding that people tend to follow even on basic ideology is particularly striking. Lenz writes:

When citizens who supported a Republican president but whose own ideologies were to the left learned that the Republican Party was on the ideological right, they didn’t shift against the Republican president—they shifted their own reported ideology to the right.

In other words, even something as basic as whether one is liberal or conservative is often a consequence, rather than a cause, of political allegiances.

Matthew Levendusky reached the same conclusion when he looked closely at how voters adjusted their partisan allegiances and ideological outlooks as the Democratic and Republican parties pulled away from one another on many major policy issues beginning in the 1970s. He writes that as the parties “pull apart to the ideological poles, they clarify what it means to be a Democrat or a Republican. Ordinary voters use these

---

78 Lenz, *Follow the Leader*, 211.
79 Ibid., 18.

How have voters responded to these cues? Levendusky explains:

...an individual voter can make this adjustment in one of two ways. He can move his partisanship into alignment with his ideology (e.g., a liberal Republican becomes a liberal Democrat), or he can adjust his ideology to fit with his partisanship (e.g., a liberal Republican becomes a conservative Republican). I find that voters typically shift their ideology to fit their party identification; ideology-driven party exit (changing one’s party to fit one’s ideology) occurs only in a narrow set of circumstances.

This helps to explain why political polarization has gone largely unchecked: parties rarely suffer any consequences of adopting more extreme positions because most voters simply update their political views to accord with their party’s stand. As the parties move toward opposite poles, their partisans simply follow them.

A series of studies by Geoffrey Cohen provide further evidence of partisan conformity. In one experiment, Cohen had participants read two proposals for a new welfare plan. One of the proposals was very generous; it would give recipients more money for more months than any existing welfare plan. The other was very stringent; it

81 Ibid.
82 As Lenz puts it, “…democracy seems to be inverted. Instead of politicians following voters on policy, voters appear to follow politicians. Moreover, they appear to do so blindly, following politicians on particular policies even when they do not know politicians’ ideologies. Politicians may therefore have considerable freedom in the policies they choose.” Lenz, *Follow the Leader*, 235.]
would give recipients less money for fewer months than any existing welfare plan. Subjects were asked the extent to which they favored each proposal. In the control condition, self-identified liberals strongly preferred the generous policy while self-identified conservatives strongly preferred the stringent one. However, when subjects in the experimental condition were told that the Democratic Party supported the stringent plan while the Republican Party supported the generous plan, participants’ preferences reversed; self-identified liberals overwhelmingly said they favored the stringent plan, while self-identified conservatives overwhelmingly said they favored the generous one. This effect held even among participants who rated themselves as highly knowledgeable about welfare policies. And Cohen found evidence that people were shifting not just their policy preferences, but also how they weighed underlying values.\(^3\) I will look at this evidence more carefully in Section 1.9.

What about deeply moralized issues like abortion? It may seem obvious that in such cases people have positions rooted in their own values and they evaluate political parties accordingly. But as it turns out, even here this seems to be only partially true. Scott Achen and Larry Bartels examined how people’s partisan allegiances and abortion views evolved as the parties took increasingly opposed stands on the issue throughout the 1980s.

---

and 1990s. More specifically, they looked at data from a national panel survey that allowed them to compare people’s abortion views and partisan allegiances in 1982 to the same people’s abortion views and partisan allegiances in 1997. Their conclusions are fascinating:

A substantial number of women gravitated to the party sharing their view of abortion... Men, on the other hand, more often changed their view on abortion to comport with their partisanship—in effect, letting their party tell them what to think about one of the most contentious moral issues in contemporary American politics.84

Interpreting this result, Achen and Bartels write:

Women’s gender identity, whichever way it brought them out on the issue of abortion, was likely to invest their views with special significance, helping them resist the cross-cutting pressures created (for some) by their partisan loyalties. Hence, when the parties diverged, women were relatively more likely to change their party rather than their abortion view. By comparison, men were more susceptible to partisan persuasion on the issue because their gender identities were generally less relevant. When their abortion views and party loyalties conflicted, they were more likely to change their abortion views.85

While women were less susceptible than men to partisan conformity, it is worth noting that partisanship still mediated women’s abortion views: women who were pro-life Democrats in 1982 were two times more likely to have become pro-choice by 1997 than their Republican counterparts (40% versus 20%).86 So, while it is certainly not the case that

85 Ibid., 264.
86 Ibid., 263.
everyone blindly follows their party on every issue, there is some evidence that conformity to party lines drives many people’s political views even on issues that are the strongest candidates for explanation in terms of deeper disagreements about fundamental values.

Contrary to the Divergent Judgments Account, people generally do not evaluate policies on the basis of their own principles and values and then decide which party or politicians to support on the basis of these policy preferences. Quite the opposite. Summarizing the results of their own work together with an extensive survey of the literature, Achen and Bartels write, “Even among unusually well-informed and politically engaged people, the political preferences and judgments that look and feel like the bases of partisanship and voting behavior are, in reality, often consequences of party and group loyalties.”\textsuperscript{87} “Even the more attentive citizens,” they say, “mostly adopt the policy positions of the parties as their own: they are mirrors of the parties, not masters.”\textsuperscript{88} This suggests that political disagreement frequently (maybe even usually) \textit{begins} with different partisan loyalties and \textit{ends} with disagreements about fundamental values and non-moral facts. The Divergent Judgments Account gets this causal chain backwards.

\textsuperscript{87} Achen and Bartels, \textit{Democracy for Realists}, 268.
\textsuperscript{88} Ibid., 299.
1.8 Political Tribalism

What’s going on here? Why does partisanship play such a dominant role in shaping people’s political views? To understand, we need to introduce two features of human psychology that are absent from the Divergent Judgments Account: tribalism and motivated reasoning. This section focuses on the former, while the next examines the latter.

1.8.1 Tribalism and Social Identities

Human beings are a tribal species; we are innately predisposed to organize ourselves into competing groups united by distinctive norms, beliefs, and ways of life. Evolutionary anthropologists Peter Richardson and Robert Boyd write:

The existence of a tribal level of organization is the most striking derived feature of human social organization. ... It is fundamental to our adaptations to the environments we have lived in. We make our livings in a staggeringly diversity of ways, but a common thread running across the gamut is the use of symbolically marked groups as foci of cooperation, coordination, and the division of labor. ... We are adapted to living in tribes, and the social institutions of tribes elicit strong—sometimes fanatical—commitment.\(^89\)

It is our tribal instincts that make social organization and cooperation possible at scales larger than a single family or kin group. But they also make us prone to divide ourselves into hostile factions. Indeed, hostility and mistrust toward out-groups is often part of what enables and sustains in-group solidarity and cooperation; we often become “we” by banding together against “them”.

In *The Federalist*, James Madison observed that our instinct to divide ourselves into competing groups is so strong that even trivial differences can become the bases of factions. He writes:

The latent causes of faction are thus sown in the nature of man; and we see them everywhere brought into different degrees of activity, according to the different circumstances of civil society. A zeal for different opinions concerning religion, concerning government, and many other points, as well of speculation as of practice; an attachment to different leaders ambitiously contending for pre-eminence and power; or to persons of other descriptions whose fortunes have been interesting to the human passions, have, in turn, divided mankind into parties, inflamed them with mutual animosity, and rendered them much more disposed to vex and oppress each other than to co-operate for their common good. So strong is this propensity of mankind to fall into mutual animosities, that where no substantial occasion presents itself, the most frivolous and fanciful distinctions have been sufficient to kindle their unfriendly passions and excite their most violent conflicts.

---

90 As Jonathan Haidt puts it, “our tribal minds make it easy to divide us, but without our long period of tribal living there would be nothing to divide in the first place. They’d be only small families of foragers...” Haidt, *The Righteous Mind*, 213.
Madison’s observation is at least partially vindicated by the work of psychologist Henri Tajfel. In a famous series of experiments, Tajfel arbitrarily divided people into groups based on made-up shared traits. For example, he told one group that they all shared a preference for Klee paintings while the other group shared a preference for Kandinsky. The groups were actually randomly assigned. He then had participants play a resource allocation game in which they could choose how much money to give to members of their in-group versus their out-group. He found that subjects heavily favored giving money to members of their in-group. Moreover, subjects distributed the money so as to maximize the difference in outcome between their in-group and their out-group, even when this meant foregoing some gains for their in-group. In other words, people prioritized creating a disparity between their in-group and their out-group over gaining additional resources for their in-group.93

These results lead Tajfel, along with his student John Turner, to develop what they called social identity theory. According to social identity theory, people’s sense of who they are as people and how they fit into the world is largely constituted by their membership in various social groups and their corresponding attitudes toward out-groups. People

---

identify themselves by identifying with groups such as families, religions, races, professions, classes, nationalities, sports teams, and political parties. In short, personal identity is largely tribal identity.

What binds a group together and determines its identity? There is no universal answer to this question, but typically groups are bound together by shared interests, norms, narratives, symbols, sacred texts, ideas, and attitudes toward other groups. To identify with a group is to take these interests, norms, narratives, and so on as one’s own. As Hadley Cantril puts it in an oft-quoted passage:

When a person’s national flag is torn down, he is insulted; when disparaging remarks are made of his parents, he is involved; when his football team or political party loses a contest, he has been defeated. Certain standards, frames and attitudes the individual feels are a part of him.

Of course, not every group to which a person belongs is central to their identity. As Achen and Bartels note, “Different people in the same group may differ. Thus some Catholics have a purely nominal attachment to the denomination. Others think of it as central to who they are, so that a disparaging remark about Catholics is an attack on them.” The key point is that group memberships are central to most people’s sense of who they are.

---


96 Achen and Bartels, *Democracy for Realists*, 228.
as people, though which memberships people take as central to their identity can vary widely.

1.8.2 Tribal Politics

How do tribalism and social identity theory help us understand politics and political disagreement? For one thing, our tribal instincts help to explain why political disagreement is so often antagonistic. As Jason Brennan explains:

> We tend to develop animosity toward other groups, even when there is no basis for animosity. We are biased to assume our group is good and just, and that members of other groups are bad, stupid, and unjust. We are biased to forgive most transgressions from our own group, and damn minor errors from other ones.\(^97\)

More importantly, tribalism and social identity theory help us make sense of the fact that political views tend to follow partisanship, rather than the reverse. This is so for two reasons: First, political parties themselves are identity-defining social groups for many people. For many people “being a Democrat” or “being a Republican” is part of their identity, and this identification with the party has little to do with the party’s ideology or policy platform.\(^98\) Second, political parties represent coalitions of other social groups with


which people identify such as ethnic groups, religious groups, labor unions, and social classes. People often align themselves with a politician or party in order to align themselves with these other social groups. Let us briefly look at each of these.

How do people come to identify with a party if not by judging that it best represents their own values, ideology, or policy preferences? For most people, the answer is simply that they were born into it. Children typically adopt the party allegiances of their parents, and these allegiances typically do not change even once children grow up, leave home, and start making their own way through the world.⁹⁹ As Achen and Bartels put it, “Partisanship, like religious identification, tends to be inherited, durable, and not about ideology or theology.”¹⁰⁰ If you’re a Democrat, odds are it is because your parents were Democrats, and their parents were Democrats, and their parents were Democrats, and so on down the line. Likewise, if you’re a Republican, it is probably because your parents were Republicans, and their parents, and so on. When people do break with their parents’

---


¹⁰⁰ Achen and Bartels, *Democracy for Realists*, 234.
partisan allegiances, it is often a result of moving to a new location (such as college) where they are surrounded by peers whose partisan allegiance is different from that of their parents.\textsuperscript{101}

Once a person has a partisan allegiance, the rest of their political thinking is typically organized around it. They adopt the norms, narratives, ideology, and policy preferences of the party as their own. In other words, people identify as Democrats or Republicans first, and then for the most part they let the party tell them what that means.

As Achen and Bartels write:

A party constructs a conceptual viewpoint by which its voters can make sense of the political world. Sympathetic newspapers, magazines, websites, and television channels convey the framework to partisans. That framework identifies friends and enemies, it supplies talking points, and it tells people how to think and what to believe. …

Once inside the conceptual framework, the voter finds herself inhabiting a relatively coherent universe. Her preferred candidates, her political opinions, and even her view of the facts will all tend to go together nicely. The arguments of the “other side” if they get any attention at all, will seem obviously dismissible.\textsuperscript{102}

To put it bluntly, people are socialized into partisan echo chambers, and then they go on repeating the echoes.

\textsuperscript{101} Sears, “The Persistence of Early Political Dispositions.”
\textsuperscript{102} Achen and Bartels, Democracy for Realists, 268.
Of course, not everyone identifies so strongly with a political party. Indeed, a record high of 42% of Americans now identify as independents.\footnote{Jeffrey M. Jones, “Democratic, Republican Identification near Historical Lows,” Gallup, accessed September 14, 2017. http://www.gallup.com/poll/188096/democratic-republican-identification-near-historical-lows.aspx.} (That said, there is evidence that the vast majority of self-identified independents consistently vote for the same party and mostly hold matching partisan political views.\footnote{Bruce E. Keith et al., \textit{The Myth of the Independent Voter} (Berkley, CA: University of California Press, 1992).}) Moreover, even within parties there are factions, some people do switch parties, new immigrants must decide which party to support, and there are occasional seismic shifts in partisanship as new parties come on to the scene, or as old parties reorient themselves within the political landscape. Understanding these cases requires broadening our view from political parties to the myriad of other social groups to which people belong.

This brings us to the second way in which tribalism and social identities orient people’s political thinking: people latch on to parties and politicians that align with and validate their memberships in the other social groups that (partially) constitute their social identities. In other words, we align with the politicians and parties that we feel represents “us” rather than “them”. Sometimes this is as simple as picking a candidate of the same race or religion. As Kinder and Kalmoe note, “In 1960, virtually every voter knew that John F. Kennedy was Catholic, and Catholics—especially “good” Catholics whose lives...
were organized around their faith—voted overwhelmingly for him.”105 Similarly, Achen and Bartels observe that, “In the 2014-2015 U.S. House of Representatives... 98% of the districts with black majorities elected black representatives. Conversely, just 5% of the districts with white majorities elected African-Americans.”106

In many cases, aligning with a politician, party, or policy is as much about opposing “them” as it is about being part of “us”. Kinder and Kalmoe write:

Scores of studies show that public opinion on matters of politics is group centered in this way: shaped in powerful ways by what they see as the principle beneficiaries or victims in play. Just as Catholics voted in great and unusual numbers for Kennedy in 1960, Protestants, driven by distrust of Catholics, voted in great and unusual numbers against him. Support for tightening welfare benefits derives, in part, from hostility towards the poor. Resistance to immigration reflects, in part, suspicions that the new immigrants are un-American. Opposition to integration of public schools and to affirmative action at work is motivated by racial hostility. And so on. Attitude toward out-groups is not the only force driving opinion in these cases, but it always seems to be present, and of all the forces that shape opinion, it is often the most powerful.107

105 Donald Kinder and Nathan Kalmoe, Neither Liberal nor Conservative (Chicago: University of Chicago Press, 2017), 137. Indeed, Catholic support for Kennedy remains a distant outlier in terms of Catholic support for any presidential candidate. Whereas in most years Catholics lean Democrat by 5 to 10 percentage points, 83% of Catholics voted for Kennedy in the 1960 election. See Achen and Bartels, Democracy for Realists, 243.

106 Achen and Bartels, Democracy for Realists, 313.

Indeed, merely associating a policy with an out-group can be enough to decrease support for it. For example, simply telling people that a policy is endorsed by evangelical religious leaders makes the policy less popular among self-identified Democrats, regardless of the content of the policy. And, I suggest, if you want to understand how Donald Trump was elected President of the United States, looking at his policy proposals or moral values is not helpful; one needs to look instead at the way in which he mobilized groups of people united by hostility and mistrust towards racial minorities, immigrants, feminists, and “political elites”.

---


108 Todd Adkins et al., “Religious Group Cues and Citizen Policy Attitudes in the United States,” *Politics and Religion* 6 (2013). The Affordable Care Act seems to be another example along these lines. Though it was heavily based on a model implanted by Republican Mitt Romney while he was Governor of Massachusetts, Republicans came to oppose it once it became linked to Obama and the Democrats.

109 Because the election was so recent (at least by academic timescales), peer reviewed evidence on the factors leading to Trump’s election is not yet readily available. However, a widely circulated conference paper by political scientist Brian Schaffner and colleagues found that racism and sexism were stronger predictors of support for Trump than economic dissatisfaction. They write, “The 2016 campaign witnessed a dramatic polarization in the vote choices of whites based on education. In this paper, we have demonstrated that very little of this gap can be explained by the economic difficulties faced by less educated whites. Rather, most of the divide appears to be the result of racism and sexism in the electorate, especially among whites without college degrees. Sexism and racism were powerful forces in structuring the 2016 presidential vote, even after controlling for partisanship and ideology.” Brian F. Schaffner, Matthew MacWilliams, and Tatishe Nteta, “Explaining White Polarization in the 2016 Vote for President: The Sobering Role of
Ultimately, the empirical evidence suggests that allegiances to political parties and other social groups with which people identify are paramount in structuring and orienting their political views. Contrary to the Divergent Judgments Account, people’s views about the merits of parties, politicians, and policies are seldom driven by a prior evaluation of their moral or practical merits. More often they are reflections of allegiances and animosities between social groups. As Achen and Bartels put it:

…voters choose political parties, first and foremost, in order to align themselves with the appropriate social groups. Most citizens support a party not because they have carefully calculated that its policy positions are closest to their own, but rather because “their kind” of person belongs to that party.\footnote{Achen and Bartels, Democracy for Realists, 307.}

And once people have a party allegiance, “often enough, they let the party tell them what to think of the issues of the day.”\footnote{Ibid., 266.}

Most political disagreements between ordinary people are thus not, first and foremost, the result of people reaching different conclusions about politics by reasoning from their fundamental values and their understandings of the non-moral facts. Rather, most people reach different conclusions about politics because they belong to different social groups and identify with different political parties that supply different ready-made conclusions. It is little wonder then that political disagreements often break down

\footnote{Racism and Sexism” (paper presented at the Conference of the U.S. Elections of 2016: Domestic and International Aspects, IDC Herzliya Campus 2016).}
along partisan and demographic lines; in many cases, allegiances and animosities toward the underlying partisan and demographic groups are what drive the disagreement in the first place.

Of course, there are exceptions. Some people do evaluate parties and candidates by carefully evaluating their policy platforms in light of their moral values and their understanding of the facts. But such people are relatively rare and almost exclusively confined to the upper echelons of society. It is easy not to see this when we spend our days surrounded by academics and when we attend mostly to political disagreement as it plays out among the intelligentsia. As Achen and Bartels note:

In the political realm, the professionals dominate political discourse and media coverage. Thus, we slip into thinking that ideology and policy preferences drive partisanship more broadly. However, the story is quite different when one looks closely at the rank and file. There, group loyalties matter a great deal, and the details of policy positions not very much.\(^{112}\)

Still, tribalism and social identities are not the whole story. As I said before, political disagreement is a massively complex and heterogeneous phenomenon and no single explanation or factor explains all of it. My point is just that tribalism and social identities are a huge part of the story of political disagreement in contemporary societies; a part that finds little, if any, place in the narrative of the Divergent Judgments Account.

\(^{112}\) Achen and Bartels, *Democracy for Realists*, 233.
1.9 Motivated Reasoning

What about the moral reasons that people give for their political views? When you ask someone why they support more generous welfare benefits, raising corporate taxes, or increasing restrictions on immigration, they rarely answer, “Because that is what my social group told me to think!” Instead, they usually talk about how we need to do more to help the poor, or how it’s unfair that corporations pay so little in taxes, or how immigrants are taking jobs from hard working Americans. How can we make sense of these justifications given the evidence that much of the time people are just parroting the positions of their party or social group? Here it is helpful to return to the study by Geoffrey Cohen described in section 1.7 (the one that found that people preferred whatever welfare policy they were told their party supports, even if it was at the opposite end of the spectrum from the policy they preferred in the absence of party information).

In addition to asking which policy they preferred, Cohen also asked participants to rate how much the following factors influenced their judgment: “the specific details of the proposal;” “your own personal philosophy of the role of government in social issues;” “what the typical Democrat or Republican believes;” and “your own background/experience with people on welfare.” Participants overwhelmingly said that they were influenced primarily by details of the policy and their own philosophy of government and least influenced by what other members of their party believe. Yet it was
clear from the results that in the presence of information about their party’s position, the
details of the policy made almost no difference; people simply preferred whatever policy
they were told their party supported.\textsuperscript{113} So the participants in Cohen’s study seem to have
been deluded about the actual reasons for their own policy preferences.

How could people think that they were choosing on the basis of the details of the
policy and their own values when in fact they were simply towing the party line? The
answer seems to be that they interpreted the policy and their own values in whatever way
was necessary to justify their party’s position. In another experiment, Cohen showed a
sample of self-identified liberals a fabricated news article describing a proposed welfare
policy. Participants were asked whether they supported or opposed the policy, and then
asked to write an editorial justifying their position. When the article did not include any
information about political parties 76\% of the subjects wrote editorials supporting the
policy. However, when the article included an additional line indicating that the
Republican Party supported the policy and the Democratic Party opposed it, the results
were totally reversed: 71\% of subjects wrote editorials explaining why the policy ought to
be rejected.

\textsuperscript{113} Cohen, “Party Over Policy.”
When Cohen coded and analyzed the content of the editorials, he found that people highlighted different moral values depending on whether or not they were told the positions of the parties. When people were left on their own to evaluate the policy without information about the parties’ positions, they overwhelmingly supported the policy and emphasized the need to strike a balance between humanitarianism and individual responsibility. However, when told the positions of the parties, most people wrote editorials opposing the policy that downplayed the importance of individual responsibility and heavily emphasized pressing humanitarian need. In other words, people went from citing their values to justify the policy to citing their values to oppose it simply because they were told which party supported it and which opposed it. Rather than weighing the competing values and then coming to a conclusion about the merits of the policy, people adopted their party’s position on the policy and then weighed values in whatever way was necessary to justify it. This is exactly the reverse of how the Divergent Judgments Account says people form their political views.

Cohen’s study wonderfully illustrates what happens when you combine political tribalism with motivated reasoning. “Motivated reasoning” is an umbrella term for an array of cognitive and affective processes that systematically skew our evaluations of reasons and evidence in favor of conclusions that what we, in some sense, want to be true. As one review of the motivated reasoning literature puts it, “people (like attorneys) often have a
preference for reaching one conclusion over another, and these directional motivations serve to tip judgment processes in favor of whatever conclusion is preferred.” Michael Huemer calls these preferences for certain conclusions “non-epistemic belief preferences”—preferences for “things we want to believe, for reasons independent of the truth of those propositions or how well-supported they are by the evidence.” What people want to believe obviously varies a great deal, but a wealth of evidence in psychology shows that people generally prefer beliefs that advance their self-interest, cohere with their existing beliefs, affirm their conception of themselves as people, validate their prejudices, and justify their partisan loyalties.

An important aspect of motivated reasoning is that the underlying processes operate mostly outside our conscious awareness. We are generally not aware that we are motivated by anything except a desire to discover the truth, and the fact that our reasoning is systematically skewed is usually invisible to us. From our first-person point of view, it seems that we are simply following the reasons and evidence where they lead. Unbeknownst to us, how we search for, interpret, evaluate, weigh, and remember information is systematically biased. Reasons and evidence that support our preferred

conclusions are more salient, seem weightier, appear more plausible, are easier to recall, and generate positive feelings—they feel right. Reasons and evidence that conflict with our preferred conclusions are less salient, seem less weighty, appear less plausible, are more difficult to recall, and generate negative feelings of dissonance—they feel like something is off. So, when we take stock of the reasons and evidence, they often seem to favor our preferred conclusions. It seems to us that our resulting beliefs are based on an objective assessment, even when in fact they’re systematically skewed. Psychologists call this the illusion of objectivity.

One classic demonstration of motivated reasoning is a study by Charles Lord, Lee Ross, and Mark Lepper. Lord, Ross, and Lepper recruited subjects who either strongly supported the death penalty or strongly opposed it. They showed their subjects short summaries of two studies investigating the deterrent effect of the death penalty. One study suggested that the death penalty does act as a deterrent, whereas the other suggested the opposite. The subjects were then asked to assess the quality of the studies and whether the studies confirmed or disconfirmed their beliefs. Lord et al. found that people generally thought the studies that conflicted with their pre-existing views were of

---


poorer quality and that, on the whole, the evidence supported their pre-existing beliefs. Supporters of the death penalty interpreted the studies as showing that it has a deterrent effect, whereas opponents interpreted them as showing no deterrent effect. Moreover, when subjects were asked about their views on the death penalty at the end of the experiment, Lord et al. found that people had become more extreme in their views. They write, “The net effect of exposing proponents and opponents of capital punishment to identical evidence—studies ostensibly offering equivalent levels of support and disconfirmation—was to increase further the gap between their views.”

Motivated reasoning explains why the subjects in Cohen’s study were able to generate moral justifications for their policy preferences even when those preferences were simply the result of partisan conformity. People have non-epistemic belief preferences for conclusions that comport with their partisan identities—people want to believe that the party they identify with is on the right side of the issues. When people are told that their party opposes a particular policy, this causes people to want to reach the conclusion that the policy is bad. They then interpret the policy and weigh relevant reasons and evidence in whatever way is necessary to reach that conclusion. If they want

to support a generous welfare policy, they highlight humanitarian needs. If they want to oppose it, they highlight the importance of individual responsibility. Though people think they’re being objective, the way they prioritize these values is biased toward whatever conclusion they want to reach.

The empirical evidence strongly suggests that motivated reasoning pervades most of our thinking about political issues. Milton Lodge and Charles Taber write:

Looking across the experimental evidence, what we find is biased processing at every stage of the evaluative process…. Even when we ask participants to stop and think, to be even handed in their appraisal of the evidence and arguments, we find precious little evidence that they can overcome their prior attitudes…

This helps to explain why reasons and arguments typically do so little to change people’s political views, and why we often look at the views of others and think “How can they really believe that?!?” It is because each of us has a systematically biased view of the reasons and evidence, and we interpret new reasons and evidence in systematically biased ways. When presented with evidence or arguments that support or views, we accept them without scrutiny. But when confronted with arguments or evidence that would undermine our views, we pick them apart looking for any reason to dismiss them.

---


121 For an interesting epistemic analysis of our tendency to selectively scrutinize disconfirming evidence, see Thomas Kelly, "Disagreement, Dogmatism, and Belief Polarization," *Journal of Philosophy* 105, no. 10 (2008).
Lodge and Taber say, when presented with arguments or evidence against their political views, “people will routinely rationalize the facts, figures, and arguments that they cannot effortlessly discount, depreciate, denigrate, or deny.” As a result, when we survey the arguments and evidence in our own minds, it often seems to us that they are overwhelmingly on our side. The arguments and evidence in support of our views often seem more or less conclusive, and the arguments and evidence against them seem flimsy at best. We conclude that anyone who denies our views must be stupid, or uninformed, or malicious. But our own views are often not nearly as obvious as they seem to us; the appearance that they are obvious is an artifact of our own motivated reasoning.

The pervasiveness of motivated political reasoning suggests once again that the Divergent Judgments Account gets things backwards. Rather than people reasoning from their values and their understanding of the facts to reach conclusions about the merits of policies and politicians, people are often, unwittingly, reasoning in the opposite direction: they begin with conclusions about policies and politicians that they want to reach, and then they interpret and weigh values and facts in whatever way is necessary to reach those conclusions.

122 Lodge and Taber, The Rationalizing Voter, 59.
To be clear, I am not saying that *all* political reasoning is motivated reasoning, nor that no one ever reasons in the way suggested by the Divergent Judgments Account. My claim is only that in in a vast range of cases (perhaps most cases) the Divergent Judgments Account does not fit the general picture of political reasoning and judgment that emerges from decades of empirical research across multiple disciplines. To adequately understand why people disagree so incessantly about politics, we need to understand how people arrive at their political views. And the empirical literature suggests that an adequate account of how people arrive at their political views must give central roles to political tribalism and motivate reasoning.

1.10 Conclusion: The Tribal Rationalizers Account

Taking political tribalism and motivated reasoning as starting points for thinking about political disagreement yields what I call the *Tribal Rationalizers Account*. This account runs roughly as follows:

People’s political views are grounded in their partisan allegiances and social identities. Partisanship is usually inherited and typically not based on prior ideology or policy preferences. People adopt the political views that comport with their partisan allegiances, help them fit in with their social group, or reflect their prejudices toward outgroups. Since people begin with different partisan
allegiances and belong to different social groups, they adopt different and often opposing political views. People then interpret and weigh values and facts in whatever way fits their pre-existing views, often simply parroting party-supplied reasons that provide an illusion of understanding. This results in further disagreement about how to interpret and weigh both values and non-moral facts.

The Tribal Rationalizers Account is more or less the inverse of the Divergent Judgments Account. Whereas the Divergent Judgments Account begins with disagreements about facts and values and ends with disagreements about the merits of parties and politicians, the Tribal Rationalizers Account begins with disagreements about the merits of parties and politicians and ends with disagreements about facts and values.

In this final section, I briefly summarize how the Tribal Rationalizers Account succeeds where the Divergent Judgments Account fails, and then close by highlighting some limits of my arguments.

1.10.1 The Advantages of the Tribal Rationalizers Account

Let us return to the three features of disagreement that the Divergent Judgments Account has trouble explaining: antagonism, clustering, and polarization. The Tribal Rationalizers Account offers compelling, though by no means complete, explanations of these features. First, political disagreements are often antagonistic because of a
combination of tribalism and motivated reasoning. Our tribal instincts predispose us to see political conflicts in terms of “us” versus “them” and to develop positive attitudes towards “us” and negative attitudes towards “them”. Moreover, motivated reasoning causes us to have a systematically biased view of politically-relevant reasons and evidence such that our own views frequently seem obviously true even when they are not. This often leads us to infer that those who disagree with us must be close-minded, uninformed, or malicious.

Second, people’s political views about logically-independent issues are often correlated because, as we saw in section 1.7, people’s political views are largely driven by their partisan allegiances. This partly explains why, for example, Americans who oppose abortion are more likely to also oppose stricter environmental regulations: it is because both of these views are associated with a single party and people tend to adopt their party’s views.

Third, the Tribal Rationalizers Account suggests a partial explanation of polarization: political parties often pay no price for taking more extreme positions because their supporters, rather than changing their evaluation of the party, simply adopt the party’s more extreme position as their own. Of course, this does not explain why the parties take more extreme positions in the first place. I suspect that part of the explanation for this is motivated reasoning. As we saw in Lord, Ross, and Lepper’s study of people on
either side of the death penalty debate, when people process information in systematically biased ways it often causes them to adopt more extreme views. This effect is exacerbated when people spend most of their time in partisan echo chambers where most of the things they see and hear are skewed towards confirming their existing views. People within these echo chambers feed off of one another, growing overconfident in their own views and adopting more extreme views over time.123

Beyond helping to explain these features of political disagreement, the Tribal Rationalizers account also better fits with the empirical literatures on political ignorance and partisan conformity. As we saw in Section 1.6, people often disagree about even well-established facts. The Tribal Rationalizers Account explains this by pointing to motivated reasoning: people tend to interpret facts in whatever way is necessary to fit their pre-existing views, even if this means denying clear and unequivocal evidence, as many Brexit voters did when told the actual proportion of immigrants. We also saw that people often know very little about issues even when they express opinions about them. This is explained by the fact that people’s political views are often rooted in partisan conformity rather than in prior assessments of the reasons and facts. As for explaining why people

tend conform to party lines, the Tribal Rationalizers account points to the fact that people’s sense of identity is often tied to their party allegiances, and identifying with a party involves largely adopting its views as one’s own.

1.10.2 The Limits of My Arguments

I will end by emphasizing once again that I am not claiming that the Divergent Judgments Account is entirely wrong, nor that the Tribal Rationalizers Account is the one true explanation of political disagreement. My primary aim has been to bring into the foreground two key factors—tribalism and motivated reasoning—that are rarely part of the scene when political philosophers talk about political disagreement. But I do not pretend that these two factors are the whole picture. I have highlighted tribalism and motivated reasoning because I believe that they are two of the most important factors missing from the Divergent Judgments Account, but I do not deny that there are many other factors that I have not discussed. Such factors include economic inequality.¹²⁴

¹²⁴ Nolan McCarthy, Keith Poole, and Howard Rosenthal argue that growing wealth inequality is one of the primary causes of political polarization in the United States. See McCarthy, Poole, and Rosenthal, Polarized America.
Moreover, I am certain that the factors highlighted by the Divergent Judgments Account are also part of a complete picture, and that the Divergent Judgments Account does explain some disagreements. I have little doubt, for example, that some people have views on abortion that are grounded in a careful assessment of the underlying moral issues, and that their views on abortion influence their evaluations of candidates and parties. My point is only that a vast amount of political disagreement does not fit this pattern. Most people’s views about most issues seem to be formed without much prior consideration of relevant reasons and facts, and much of the time people seem to be simply following their party’s lead. So, by taking the Divergent Judgments Account as their go-to account of political disagreement, political philosophers neglect much of the actual problem of political disagreement.

---

126 Sunstein, *Republic: Divided Democracy in the Age of Social Media*.
2. Utopianism and Political Irrationality

We still must discover an arrangement which engages with practical reason, whatever its character, in two ways: as an arrangement desirable in itself and as one to which we can reasonably hope to conform. Any political theory which is not utopian must bring these two types of justification together. – Thomas Nagel

We humans are not as rational and unbiased as we like to think, especially when it comes to the moral and political issues that sunder our societies. As we saw in the previous chapter, our views about such issues are often systematically biased, shaped more by partisan allegiances than by an objective assessment of arguments and evidence. While I have emphasized tribalism and motivated reasoning as the primary culprits, these are far from the only factors that distort our views about political issues. As the editors of the *Oxford Handbook of Political Psychology* write:

…both citizens and leaders exhibit distorted reasoning and a slew of cognitive and emotional biases…. Partisan resistance to new information, ethnocentric reactions to immigrants, automatic and preconscious reactions to a political candidate’s facial features, greater risk-taking in the face of losses than gains—the list goes on.

---

What should political philosophers make of these facts? Must we adjust our normative political theories in light of them? Should we be skeptical of any theory that presupposes that people are generally rational and unbiased about politics? Is such a theory utopian in the pejorative sense of the term? Answering these questions is the central aim of this chapter and the next.

In this chapter, I focus on some recent work by David Estlund in which he purports to show that the kind of empirical results surveyed in the previous chapter are “not as devastating to normative political theory as it is usually thought.” To show this, Estlund argues that even if people are very unlikely to act or believe in the way that a normative political theory requires, “the standards to which [the theory] holds people and institutions might be sound and true.” “The fact that people will not live up to them,” says Estlund, “is a defect of people, not of the theory.”

While I accept much of Estlund’s argument, I argue here that it falls short of its aim. More specifically, I argue that it fails to vindicate normative political theories that presuppose or require what I will call Enlightened Citizens:

---

4 Ibid., 264.
Enlightened Citizens: People’s political views are generally the result of more or less informed, rational, and unbiased assessments of relevant reasons and evidence.

As we will see in the next chapter, Rawls’s highly influential account of political liberalism depends on a version of this claim. Many theories of democracy also seem to require some version of it. Robert Dahl, for example, says that for democracy to realize conditions of political equality, citizens must gain “enlightened understanding” of policy issues. He does not tell us exactly what such understanding involves, but the label suggests that it precludes being ignorant and irrational. At any rate, my aim in this chapter is not, first and foremost, to criticize particular theories, but rather to explore the limits of normative political philosophy. For this reason, I focus on the more abstract question of whether a normative political theory may presuppose or require Enlightened Citizens, rather than the details of any particular theory.

---

5 Rawls says that he is “concerned with an ideal normative conception of democratic government, that is, with the conduct of its reasonable citizens and the principles they follow, assuming them to be dominant and controlling.” And Rawls says that ignorance, irrationality, and bias are incompatible with being reasonable. Rawls, Political Liberalism, 441 note 3.
2.1 The Charge of Utopianism

To say that a normative political theory is utopian, in the pejorative sense, is to say that the theory is, in some way, too unrealistic to be credible. But what makes a theory too unrealistic, and what is the relevant measure of realism? Let us begin by considering two different forms that the charge of utopianism could take. The first is that the theory is simply at odds with the facts about how politics actually works. John Horton presses a version of this objection against what he calls liberal moralism which he takes to include the work of “Rawls, Dworkin and many others, including most Habermassians.”

Horton says that such theories are descriptively deficient, by which he means, “at its simplest... the fact that one would get a rather odd idea of what politics is actually like, even in societies that came closest to meeting the conditions of liberal moralism, if one had only liberal moralism as one’s guide.”

Taking aim at much of contemporary democratic theory, Horton writes:

...conceptions of deliberative democracy favoured within contemporary liberal theory typically resemble an unusually well-conducted academic seminar rather than the hurly burly of political debate and argument. Generally, the conception of politics appears etiolated, antiseptic and impossibly high-minded. In this regard, therefore, liberal moralism seems already to have erased from its

8 Ibid., 433.
conception of politics in societies like our own what might be thought to be some of its most recognizable and familiar features.⁹

Taking Horton’s charge as a model, we might say that any normative political theory that presupposes or requires Enlightened Citizens is utopian because, as we saw in the previous chapter, Enlightened Citizens is far from an apt description of the actual world.

But this is too hasty. Merely pointing out that a normative political theory does not accurately describe the world need not threaten the theory. This is because normative theory is supposed to tell us, first and foremost, how the world should be, not necessarily how it is. To point to a mismatch between the world as it is and the world a normative theory describes may simply be to point out that the actual world is not as it should be. As Estlund rightly points out, if the kind of realism desirable in a normative theory is correspondence to the world as it is, then “the most realistic normative theory of all... would recommend or require people and institutions to be exactly as they actually are already (or exactly as, it so happens, they will come to be).”¹⁰ But a theory that tells us only to do what already is (or will be) done is obviously not a plausible normative theory, at least not if we think that our actual world falls short of moral perfection. Thus, the

---

¹⁰ Estlund, "Utopophobia," 115.
charge of utopianism, if it is to be a damaging charge, must point to more than just a mismatch between the world as it is and the world a normative theory describes.

This brings us to the second form that the charge of utopianism may take. One might point out not only that the conditions described by a normative theory do not obtain, but also that those conditions are very unlikely to ever obtain. This kind of objection is expressed when people say things like, “Sure, that would be nice, but we both know it’s never going to happen.” Adopting this approach, we might say that the problem with a theory that presupposes Enlightened Citizens is not simply that this presupposition is false of the actual world, but also that it is very unlikely to ever be true. In this case, the charge of utopianism amounts to the charge that the conditions required by the theory are too improbable.

There are two different ways to defend a theory against this second kind of charge. First, one can deny that the conditions identified by the theory are actually so improbable. In the present case, this would mean showing that although people tend to be uninformed, irrational, and biased about politics as it is, it is not too improbable that they will cease to be so at some point in the future. For now, I set this defense aside; we will return to it later. The second defense is to argue that even if the conditions identified by a theory are highly improbable, the improbability of these conditions is not necessarily problem for the theory. This is the approach that Estlund takes.
Estlund does not deny that people are often ignorant, biased, and irrational about politics. He writes:

There are many serious charges against voters, and I want to grant many, though not all of them, for the sake of argument. Rather than challenge the literature that explores the deficiency of voters, I want to concede much of it in order to explain how it is not as devastating to normative political theory as it is usually thought to be.\textsuperscript{11}

So, let us assume for now that Enlightened Citizens is not only false, but also unlikely to ever be true. The question is what, if anything, follows from this about the merits of a theory that presupposes or requires Enlightened Citizens.

### 2.2 Probability, Possibility, and Hopeless Aspirational Theory

Estlund argues that the mere fact that some action is very unlikely to be done or that some state of affairs is very unlikely to come about does not in any way undermine a theory that claims that the action ought to be done or that the state of affairs ought to be brought about. He writes:

...moral theories of social justice, political authority, political legitimacy, and many other moral-political concepts are not shown to have any defect in virtue of the fact, if it is one, that the alleged requirements or preconditions of these things are not likely ever to be met. If a theory of social justice is offered, and it is objected “But you and I both know that people will never do that,” I believe the right response is (as a starter), “I never said they would.”\textsuperscript{12}

\textsuperscript{11} Estlund, \textit{Democratic Authority}, 259.
\textsuperscript{12} Estlund, "Utopophobia," 113-14.
On this point, Estlund and I agree. Indeed, I think the point is obviously true. To see this, suppose that in 1977 it was very unlikely that Pol Pot’s Khmer Rouge would, of their own volition, stop murdering ethnic and religious minorities in Cambodia. This fact clearly does nothing to undermine a theory that says (or implies) that the Khmer Rouge ought to have stopped committing such murders. “Ought” does not imply “likely”.

Estlund suggests that charges of utopianism often rest on conflating improbability with impossibility. He grants, “for the sake of argument, that ‘ought’ entails ‘can’—that if something is not within someone’s abilities, then it is not required.” But, he insists, we must be careful to distinguish cannot from almost definitely won’t. The fact that a person or a society is very unlikely to do something does not entail that they cannot do it. Indeed, Estlund points out, it does not even entail that it would be difficult for them to do.

Illustrating this point, Estlund writes:

Consider this case: what is the chance that I will dance like a chicken while giving a lecture? It is very, very close to zero, trust me. So should we accept that it is nearly impossible to do? To say that an action is “nearly impossible to do” suggests that it is extremely difficult, that would be a non sequitur. It does not get more difficult as it gets less likely.

14 Ibid., 118.
This demonstrates that, “an action’s being unlikely does not entail that it is beyond the agent’s ability. So even if moral standards of justice entail ability, they would not thereby be shown to be refuted by very low probability.”\textsuperscript{15}

Given that “ought” does not imply “likely,” and that “very unlikely” is not the same as “impossible”, there is nothing defective, says Estlund, about a normative political theory that prescribes standards that are very unlikely to be met. He writes:

Consider, then, a theory that holds individuals and institutions to standards that it is within their ability to meet, but that there is good reason to believe they will never meet. So far, I contend, the theory has no defect. It might be a false theory if it claimed that the standards would someday be met, but it does not say that. And we can suppose that it would be false if the standards were impossible to meet, but again, by assumption, they are not. Many things that are within people’s abilities will never be done, of course. The imagined theory simply constructs a vision of how society should and could arrange itself (maybe even without great difficulty…), even while acknowledging that this will not happen.\textsuperscript{16}

Estlund calls this kind of theory a \textit{hopeless aspirational theory}. He emphasizes that the term “hopeless” is not intended to signal a defect. He says:

The name “hopeless” might suggest that I aim to criticize this kind of theory, but in fact, I want to defend it. I keep the sad name in order to avoid any suggestion that my point is maybe the standards will, after all, be met… I want to defend political theory that defends standards even though they will \textit{not} be met, and even if we knew this for sure.\textsuperscript{17}

\textsuperscript{15} Estlund, “Utopophobia,” 117.
\textsuperscript{16} Ibid.
\textsuperscript{17} Ibid., 118.
I do not believe that there is anything inherently wrong with hopeless aspirational theory as Estlund defines it. Certainly, we should not limit our normative political theories to proposing only things that are likely to happen. And, as I said, I think Estlund is obviously right that “ought” does not imply “likely”. Of course, sometimes there are good reasons to work within such limits, especially if our aim is to propose immediately practicable political solutions, but Estlund does not deny this. He acknowledges that there is an important role for what he calls hopeful realistic normative theories; theories that propose “appropriate standards that are not only possible for people and institutions to meet (that is, within their abilities), but that it is also reasonable to believe they might meet.”18 However, Estlund insists, the constraints of such a hopeful theory should not constrain all normative political philosophy.

At this point, one might object that hopeless aspirational theory is a useless enterprise. Why should we carry on developing visions of society that we know will never be realized? And don’t such visions fail to provide the kind of practical guidance that normative theories ought to provide? In response, Estlund suggests that even if a hopeless aspirational theory provides little practical guidance, it may still be the true theory about

18 Estlund, "Utopophobia," 121.
what justice, or legitimacy, or some other such value requires. And even if this truth is of little practical application, it may still be worth knowing. Estlund writes:

...the general view that only practically applicable intellectual work has value is difficult to maintain without implausibly implying that even what are widely regarded (not just by mathematicians) as the greatest mathematical achievements have, it turns out, no great value of any kind. If higher math is important, then aspirational theory is not shown to be unimportant simply for being (like much higher math) of no practical value.¹⁹

I am not sure that this analogy is a good one; it seems to me that the value of political theories is much more closely tied to their practical applicability than are truths of math. But I will not press the point here, since I grant that a theory may soundly require things that are very unlikely to happen, and thus that hopeless aspirational theory is not defective per se.²⁰

¹⁹ Estlund, "Utopophobia," 134.
²⁰ Ultimately, I think the value of a practically useless normative theory depends on whether truths about justice are mind-independent truths (as argued by realists like Russ Shafer-Landau and David Enoch) or constructed truths (as argued by constructivists like Sharon Street, David Copp, and — on some interpretations — John Rawls). If it is the former, then there is little reason to think that truths about justice must be truths we could, practically speaking, act on. However, if it is the latter, then truths about justice are likely to be bound by constraints that preclude them from being practically inert. See Russ Shafer-Landau, Moral Realism: A Defense (New York: Oxford University Press, 2003); David Enoch, Taking Morality Seriously: A Defense of Robust Realism (Oxford: Oxford University Press: 2011); Sharon Street, "Constructivism About Reasons," Oxford Studies in Metaethics 3 (2008); Sharon Street, "What Is Constructivism in Ethics and Metaethics," Philosophy Compass 5, no. 5 (2010); David Copp, "A Skeptical Challenge to Moral Non-Naturalism and a Defense of Constructivist Naturalism," Philosophical Studies 126, no. 2 (2005); John Rawls, "Kantian Constructivism in Moral Theory: The Dewey Lectures 1980," Journal of Philosophy 77, no. 9 (1980).
2.3 The Limits of Estlund’s Argument

Let us return to the question I began with: Is a normative political theory utopian if it requires or presupposes Enlightened Citizens? We have seen that Enlightened Citizens is false, and we have supposed that it is unlikely to ever be true. But does this condemn a theory that presupposes or requires it?

Estlund’s argument might seem to provide a sound case for answering “no”, and to a certain extent, I believe it does. Merely pointing to the fact that Enlightened Citizens is unlikely to be true does not show that a theory that presupposes or requires Enlightened Citizens is defective. However, I contend that Estlund’s argument nonetheless falls short of vindicating such a theory against the charge of utopianism. To see this, we need to ask why Enlightened Citizens is unlikely to ever be true.21 Why is it very unlikely that we will ever reach a point at which most people are informed, rational, and unbiased about political issues? Is it because, as in Estlund’s example of dancing like a chicken, people could easily do it but they will never choose to? Or is it more like, say, trying to run 100

21 This question, rather than a mere mismatch between the world and our theories, seems to be what Horton is ultimately driving at in making the charge of descriptive inadequacy. He writes, “The motive behind the charge of descriptive inadequacy is not... that we have uncritically to embrace every aspect of the world as it is, but that whatever we think of it, we first have to acknowledge it for what it is, seek in general terms to understand why it is so, and also accept that it is only ever to a limited extent directly malleable in accordance with the aspirations and intentions of human beings.” Horton, “Realism, Liberal Moralism and a Political Theory of Modus Vivendi,” 435.
meters in under 10 seconds? That is, not impossible, but very difficult—so difficult that even with a lot of effort most people will fail most of the time.22

If it is the former, then I grant that Estlund’s argument vindicates a theory that presupposes or requires Enlightened Citizens. However, if it is the latter, then Estlund’s argument comes up short. This is because Estlund’s argument shows only that the mere fact that something is unlikely to come about does not undermine a theory that requires it. To accept this conclusion, we need only accept that a theory may soundly require things that are easy but very unlikely. Indeed, Estlund focuses heavily on cases in which “very low likelihood simply reflects agents’ ordinary motives and choices, not obstacles of any kind.”23 Accepting that a theory may require things that are easy but unlikely is compatible with denying that a theory may soundly require things that are very unlikely because they are very difficult.24 Thus, Estlund’s argument tells us nothing, one way or the other, about whether a theory can soundly presuppose or require Enlightened Citizens if Enlightened Citizens is very unlikely to be true because realizing it is very difficult. So,

---

22 One might suggest that for most people, running 100 meters in under 10 seconds is not just very difficult, but impossible. However, what is at issue here is what is possible for humans in general, not for any particular human. It is humanly possible to run 100 meters in under 10 seconds—some humans have done it—but it is very difficult.


24 Estlund himself notes this with regard to theories that are “harsh”, by which he means theories that “purport to require more strain or sacrifice than is genuinely required” and thereby “demand too much of agents.” Ibid., 122.
although Estlund’s argument might seem to insulate normative political theories from concerns about the ignorance, bias, and irrationality of voters, it falls short of doing this if being informed, unbiased, and rational about politics is very difficult.

At this point, there are two lines of response available to Estlund. First, he could reject the empirical claim (which I have not yet argued for) that it is very difficult to be rational and unbiased about politics. I set aside this response until the end of the chapter. For now, I focus on the second possible response: Estlund could argue that normative political theories can soundly require not only things that are unlikely but easy, but also things that are unlikely because they are very difficult. In that case, even if it is true that it is very difficult to be rational and unbiased about politics, this still would not undermine a normative theory that presupposes or requires Enlightened Citizens. My aim in the next section is to show that Estlund’s own defense of democracy requires foregoing this second line of response. The point of doing so is not simply to put Estlund on pain of inconsistency, but rather to illustrate the role that constraints on normative theorizing often play in framing the very problems that our normative theories are supposed to solve—an issue that will be central to my argument in the next chapter.
2.4 Estlund Against Epistocracy

I will not attempt to explain Estlund’s entire defense of democracy here. For now, it suffices to focus on one aspect of his defense, his rejection of an alternative to democracy that he calls *epistocracy*.²⁵ Whereas democracy is rule by the many, epistocracy is a form of rule by the few, specifically by the few who are experts. Epistocracy poses a challenge to democracy because it seems likely that some form of epistocracy is likely to make good political decisions more reliably than democracy. “It is certain,” says Estlund, “that there are subsets of citizens who are wiser than the group as a whole,” and who “know better than the rest of us what should be done.”²⁶ Why shouldn’t these wiser and more knowledgeable citizens rule over the rest of us? Anyone who contends that democracy is the normatively best form of political rule must answer this question, but the question is especially pressing for philosophers like Estlund who seek to defend democracy in part because of the epistemic merits of democratic procedures. If the value of democracy lies largely in its tendency to make good decisions, then why not favor an epistocracy that would make even better decisions?

---

²⁶ Ibid., 40.
Whereas some theorists reject the presupposition that an epistocracy is likely to make better decisions than democracy,\textsuperscript{27} Estlund concedes that the challenge of epistocracy cannot be met by any appeal to epistemic considerations. Instead, he says, we must “appeal to some principle” that rules out epistocracy on non-epistemic grounds.\textsuperscript{28} Estlund invokes just such a principle, which he calls the \textit{qualified acceptability principle}. This principle asserts that any legitimate exercise of political power must “be justifiable in terms acceptable to all qualified points of view (where “qualified” will be filled in by “reasonable” or some such thing).”\textsuperscript{29} This principle rules out epistocracy as a legitimate form of political rule because “any particular person or group who might be put forward as such an expert would be subject to controversy, and qualified controversy in particular.”\textsuperscript{30} In other words, since reasonable\textsuperscript{31} people will inevitably disagree about who the experts are, no form of epistocracy can be politically legitimate.

Of interest here is Estlund’s claim that reasonable people will inevitably disagree about who the experts are. If this amounts to the claim that agreement about who the

\textsuperscript{27} See, for example, Landemore, \textit{Democratic Reason}.
\textsuperscript{28} Estlund, \textit{Democratic Authority}, 40.
\textsuperscript{29} Ibid., 41.
\textsuperscript{30} Ibid., 36.
\textsuperscript{31} Estlund eschews the word “reasonable” in favor of “qualified” in order to avoid committing to any particular interpretation of the qualified acceptability requirement. Since it makes no difference to the arguments that follow, I opt to use “reasonable” instead of “qualified” for ease of exposition.
experts are is very unlikely, then Estlund’s defense of hopeless aspirational theory seems to open the door to a defense of epistocracy. To see this, imagine that a proponent of epistocracy makes the following response to Estlund:

**The Hopeless Epistocrat**: I defend what you might call *hopeless epistocracy*. According to this theory, people *ought* to agree about who the experts are, and the agreed-upon experts ought to rule over the rest. Of course, I grant that this vision will probably never be realized because people are very unlikely to ever agree about who the experts are. But as you’ve argued, the fact that the requirements of a theory are unlikely to be met does not undermine the theory. Indeed, hopeless epistocracy is exactly the kind of hopeless aspirational theory you champion. So, what argument do you have against hopeless epistocracy?

Notice that the Hopeless Epistocrat’s version of epistocracy does not run afoul of the qualified acceptability requirement. It does not say that experts ought to rule even if there is disagreement about who the experts are. Rather, it says that in an ideal society people would agree about who the experts are and those experts would rule. In other words, hopeless epistocracy presupposes that people do not disagree about who the experts are. The Hopeless Epistocrat grants that this presupposition is very unlikely to ever obtain, but why should this undermine hopeless epistocracy as a normative theory? The theory seems to do exactly what Estlund says a hopeless aspirational theory should do; it
“constructs a vision of how society should and could arrange itself... even while acknowledging that this will not happen.” So, Estlund’s argument against epistocracy seems to fail against hopeless epistocracy, and thus fail to show that democracy is uniquely justified as a normative ideal.

How might Estlund respond to the Hopeless Epistocrat? Perhaps he could argue that the vision of society proposed by hopeless epistocracy is not just improbable, but impossible. But, so far as I can tell, it isn’t impossible. It is not logically impossible. Nor does it violate any laws of nature. Moreover, we can imagine scenarios in which the Hopeless Epistocrat’s vision of society is realized. Suppose we have a very small society of 10 people who must decide who ought to govern them. We can easily imagine that they might all agree that one or two among them know better than the rest how to govern. The same is true of a slightly larger society of 15, or of 20. As we continue scaling up the size of the society, it may become increasingly unlikely that an agreement will be reached about who the experts are, but, as Estlund himself insists, increasingly unlikely is not impossible. So why can’t a normative theory require or presuppose that people agree about who the experts are?

__________________

32 Estlund, "Utopophobia," 117.
Insofar as Estlund has a viable response, I think it comes when he highlights “the difficulty of identifying, in a way acceptable to the broad range of qualified points of view, a set of experts who could be expected to perform better than the best democratic arrangement.”33 This suggests that the problem with presupposing that everyone will agree about who the experts are is not simply that such agreement is unlikely, but rather that it is exceedingly difficult—indeed, nearly impossible—to achieve such agreement, at least in large and heterogeneous societies. And this remains so even if we require only agreement among reasonable people. The problem with hopeless epistocracy, then, is not merely that its preconditions are unlikely, but rather that they are practically unachievable in anything resembling a modern society. Requiring such agreement is more like requiring everyone to run 100 meters in under 10 seconds than like requiring everyone to dance like a chicken. A few exceptional people may succeed, but most will fail.34

However, to make this move, Estlund must grant that even hopeless aspirational theories cannot require things that are very unlikely if the reason they are very unlikely is that they are very difficult. But what counts as “very difficult” in the relevant sense? Abolishing the Atlantic slave trade was very difficult, but it would be perverse to take this

33 Estlund, Democratic Authority, 262.
34 Note that the reason for their failure, at least according to Estlund, is not that there are no experts that could be agreed on. As we saw, Estlund concedes that “it is certain” that there are such experts. Ibid., 40.
as a reason to reject a normative theory that forbids slavery. But if normative theories can soundly require very difficult things like abolishing the slave trade, why can’t they also require that everyone come to some agreement about who the political experts are? I think the answer—explored more in the next chapter—lies in the idea that a normative vision of society should be one that ordinary humans have a reasonable chance of realizing if they try, at least under favorable, yet still realistic, conditions. This need not mean that we are likely to realize that society—we might not even try. But it must be a society that, in Rawls’s words, “is feasible and might actually exist, if not now then at some future time under happier circumstances.” A world without slavery is such a world, whereas a world without reasonable disagreement about who counts as a political expert is not.

My purpose in pointing out that Estlund’s own defense of democracy requires invoking some such constraints on normative political theorizing is not simply to back Estlund into a corner. I think there is a more general lesson to be learned. The lesson is that some constraints on normative theories are necessary in order to frame the very problems that those theories are supposed to solve. Before we can begin to explain why

---

35 For a gripping account of the trials and tribulations of the abolitionist movement, see Adam Hochschild, *Bury the Chains: Prophets and Rebels in the Fight to Free the Empire’s Slaves* (New York: Mariner Books, 2005).
democracy is the best form of political rule, we must grant several facts about people and about the world, one of which is that people disagree about who ought to rule. If we allow political theories to go unconstrained by such facts—if we allow them to presuppose anything that is within the limits of even very remote possibility—then we are, as Sidgwick beautifully puts it, cast into “an illimitable cloudland surrounding us on all sides, in which we may construct any variety of pattern states.”\textsuperscript{37} In the next chapter, I argue that some constraints are necessary in order to frame one of the central problems of liberal political philosophy, what Rawls called the fact of reasonable pluralism.\textsuperscript{38}

2.5 Thinking Straight is not like Dancing like a Chicken

Let us now return to the question we began with: Is a normative political theory utopian, in the pejorative sense, if it requires or presupposes Enlightened Citizens, that is, that people’s political views are generally the result of more or less informed, rational, and unbiased assessments of relevant reasons and evidence? While Estlund takes himself to have shown that the answer is “no”, I have argued that his argument falls short of establishing this conclusion. More specifically, I have argued that it falls short if being rational and unbiased about politics is not just something people are unlikely to do, but

\textsuperscript{38} Rawls, \textit{Political Liberalism}, 36.
also very difficult to do. I have also suggested, using Estlund’s own defense of democracy as an example, that normative theories must be bound by some constraints that rule out normative requirements that are too difficult to satisfy. The next chapter explores the importance and implications of such constraints in much greater detail, but for now, I turn to the question I put off earlier: Is it very difficult to be rational and unbiased about politics? Is it more like dancing like a chicken or like running 100 meters in under 10 seconds? I submit that it is the latter.

The basic reason that it is very difficult to avoid biased and irrational thinking about politics is that most of the processes that cause such thinking operate outside our conscious awareness and in ways that are difficult for us to detect and correct. Motivated reasoning, partisan conformity, and the many cognitive and affective biases that afflict political reasoning are, as Samuel Bagg puts it, “unintentional and even invisible to us; unconsciously woven into the fabric of our moral experience.”\(^\text{39}\) When we introspectively examine our own reasoning, it typically seems to us that our reasoning is free of distorting influences even when it’s not. We sincerely believe that we are guided only by the weight of the evidence and a concern for the truth, when in fact our reasoning is biased towards

conclusions that reaffirm our existing beliefs, cohere with our partisan loyalties, or validate our prejudices. We suffer from an illusion of objectivity.\footnote{Kunda, “The Case for Motivated Reasoning,” 482-83.}

The illusion of objectivity makes it very difficult to detect and correct our own biased reasoning. Moreover, this illusion is most robust under exactly the conditions within which we make most political judgments: conditions where there is complex and conflicting evidence, where we must interpret and apply abstract normative concepts, and where there are competing normative considerations on both sides of the issue. Such conditions “leave a motivated moral judge considerable flexibility to construct plausible justifications for preferred moral conclusions without offending their sense of their own objectivity.”\footnote{Ditto, Pizarro, and Tannenbaum, “Motivated Moral Reasoning,” 314.} Under such conditions, we are especially prone to what Elizabeth Pronin calls our \textit{bias blind spot}: our inability detect our own biased reasoning, even when we are aware of the dangers of biased reasoning and can readily detect it in others.\footnote{Emily Pronin, Daniel Lin, and Lee Ross, “The Bias Blind Spot: Perceptions of Bias in Self Versus Others,” \textit{Personality and Social Psychology Bulletin} 28, no. 3 (2002).}

One might respond that even if we cannot always determine whether our reasoning is biased, we could still avoid biased reasoning by carefully reflecting on the relevant reasons and evidence and making a concerted effort to form an objective view.
However, the empirical evidence suggests that this is not an effective strategy, at least when it comes to divisive moral and political issues. Lodge and Taber, for example, write:

…in our studies we promoted central-route processing [i.e. conscious, deliberative thought] by choosing issues which our participants thought important, motivated them to think seriously about the arguments, encouraged them to be evenhanded, and in some studies upped the ante by informing them that they would have to publically defend their attitudes.43

Summarizing the results, they write:

Across multiple experiments we find no evidence of an even-handed integration of new-found information into evaluations, finding instead affect-driven systematic confirmation and disconfirmation biases in judgment leadings to attitude polarization.44

Indeed, not only is careful reflection not a reliable antidote to biased reasoning, but some evidence suggests that intelligent people who engage in more critical reflection are actually more prone to both biased reasoning and to the bias blind spot.45 The same is true of greater political knowledge, as Danielle Shani reports:

…political knowledge does not correct for partisan bias in perception of “objective” conditions, nor does it mitigate the bias. Instead, and unfortunately, it

43 Lodge and Taber, The Rationalizing Voter, 229.
44 Ibid., 227.
enhances the bias; party identification colors the perceptions of the most politically informed citizens far more than the relatively less informed citizens.\textsuperscript{46}

The fact that biased political reasoning has been shown to resist conscious efforts to avoid it and to be \textit{positively} correlated with both increased critical reflectiveness and increased knowledge provides reason to be pessimistic about our prospects of overcoming it.

To be clear, I am not arguing that humans are totally incapable of being rational or unbiased about politics. Nor am I arguing that people could not be much more rational and unbiased than they are now; I am not saying the status quo is the best we can do. My point is just that there are good reasons to think that being rational and unbiased about politics is not like Estlund’s case of dancing like a chicken; it is not something we could easily do but simply choose not to. Quite the opposite; the empirical literature suggests that even when we try to do it we will often fail. It is more like running 100 meters in under 10 seconds—some people can do it some of the time, but most of us fail most of the time.

Of course, one may dispute this empirical claim. And I cannot rule out the possibility that we will eventually discover techniques or interventions that effectively eliminate most political bias and irrationality. But it is clear, I think, that we face

\textsuperscript{46} Danielle Shani, "Knowing Your Colors: Can Knowledge Correct for Partisan Bias in Political Perceptions?" (paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago 2006), 31.
formidable, and possibly insurmountable, obstacles to creating a world where Enlightened Citizens is true. In the next chapter, I argue that these obstacles give us reason to think that *unreasonable* political disagreement is a fact of life that our normative political theories must be designed to cope with.
3. The Fact of Unreasonable Pluralism

“This reasonable society is neither a society of saints, nor a society of the self-centered. It is very much a part of our ordinary human world, not a world we think of much value until we find ourselves without it.” – John Rawls

Any successful normative theory of democracy must contend with the fact of reasonable pluralism—the fact that reasonable disagreement about morality and religion is inevitable in a free society. Because they subscribe to incompatible doctrines, citizens will inevitably disagree about what is good, sacred, and otherwise important in life. A theory of democracy that denied this, or that assumed such reasonable disagreement to be nonexistent, or even just rare, would rightly be criticized for neglecting a central problem of democratic politics. This is all well and good, but why is it only reasonable pluralism that commands such attention? What about disagreements rooted in ignorance, bias, irrationality, and prejudice? Aren’t these just as common—perhaps more common—than reasonable disagreements? Why isn’t unreasonable disagreement also a central problem of democratic politics?

---

2 Ibid., 36.
For answers to these questions, we must look to the work of John Rawls to whom the fact of reasonable pluralism owes both its name and most of its currency in contemporary political philosophy. Rawls grants that bias and irrationality explain many of the disagreements found in actual liberal societies, “but,” he says, “while such explanations explain much, they are too easy and not the kind we want... for we always work at first within ideal theory.” As I understand it, this answer consists of two key claims:

(1) The primary project of political philosophy is (or ought to be) ideal theory, that is, the project of sketching the normative outlines of an ideal (or well-ordered) society.  

(2) For the purposes of ideal theory, we may assume that the overwhelming majority of citizens of an ideal society would be overwhelmingly reasonable and rational, so ideal theory need not concern itself with unreasonable disagreements rooted in bias and irrationality.

---

3 Rawls, Political Liberalism, 55.

4 Rawls argues that ideal theory must take priority because “until the ideal is identified, at least in outline... non-ideal theory lacks an objective, an aim, by reference to which its queries can be answered.” Rawls, The Law of Peoples, 90.

5 Rawls assumes that in an ideal (or “well ordered”) liberal society, reasonable citizens are “dominant and controlling” and unreasonable doctrines “do not gain enough currency to undermine society’s essential justice.” Rawls, Political Liberalism, 441 note 3, 39.
Many philosophers reject (1). They allege that ideal theory is unnecessary, misguided, or even pernicious, and thus that we ought to stop doing it or at least stop affording it such priority.⁶ I think that many of these criticisms are somewhat misplaced, and that the problem is not ideal theory per se, but rather the way in which it is typically done. At any rate, I shall grant (1) here in order to focus on (2).

My thesis in this chapter is that (2) is false; ideal theory cannot assume that the overwhelming majority of citizens of an ideal society would be overwhelmingly reasonable and rational (at least not about divisive moral and political issues). I do not defend this thesis in unqualified form, but rather as the consequent of a conditional. The antecedent of this conditional consists of some standard Rawlsian assumptions about the limits of ideal theory. The assumptions I have in mind stem from the idea that ideal theory is supposed to be “realistically utopian.”⁷ To be such, the ideal society described by ideal theory must be “an achievable social world”⁸ that could be realized, at least “under the

---


⁸ Ibid.
best foreseeable conditions.”⁹ Thus, ideal theory cannot assume that an ideal society (or its citizens) transcend “the fixed constraints of human life.”¹⁰ These constraints include both facts about human nature and about the world around us. Exactly which facts these are is open to debate, but Rawls assumes that they include the fact that humans are partly selfish (i.e. not perfectly altruistic), and the fact that resources are moderately scarce.¹¹

The limits of ideal theory play a crucial, and I think underappreciated, role in grounding the philosophical significance of the fact of reasonable pluralism. Rawls argues that because of various obstacles that beset human reason—the so-called burdens of judgment—we cannot plausibly expect even intelligent and conscientious people to reach agreement about what is good, right, sacred, and otherwise important in life.¹² Thus, ideal theory must assume that reasonable disagreement about such matters is one of the fixed facts of life in a free society. This assumption limits the range of societies that ideal theory

---

⁹ Rawls, Political Liberalism, xvii.
¹¹ Rawls, drawing on Hume, calls these facts the circumstances of justice. See ibid., 109-12. This aspect of Rawls’s view is challenged by David Estlund. He contends that normative political theories need not accommodate the limits of human motivational capacities (i.e. the fact that people are partly selfish). Estlund argues that if this were so, it would (potentially) commit ideal theory to adjusting its conclusions to accommodate many unsavory aspects of human nature that (intuitively, at least) ought not to affect our vision of an ideally just society. In Section VII, I suggest that my main argument here could be interpreted as warranting an even more radical rejection of Rawls’s constraints. See David Estlund, "Human Nature and the Limits (If Any) of Political Philosophy," Philosophy & Public Affairs 39, no. 3 (2011): especially pages 225-27.
¹² Rawls, Political Liberalism, 55.
may recommend. For example, it rules out any vision of a free society that is premised on unanimous (or even nearly unanimous) assent to a single moral or religious worldview.

I argue that if one accepts this standard Rawlsian story about the fact of reasonable pluralism, then one must accept that unreasonable moral and political disagreement rooted in bias and irrationality is also an inevitable fact of life in a free society, and that this fact also limits the range of societies that ideal theory may recommend. My argument parallels Rawls’s. I argue that we face obstacles, like the burdens of judgment, that make it very difficult for even intelligent and conscientious people to reason about divisive moral and political issues in rational and unbiased ways. These obstacles, which I call the sources of unreason, stem from deep-rooted features of human psychology including tribalism, motivated reasoning, and an array of other cognitive and affective biases. I contend that just as we cannot expect even intelligent and conscientious people to reliably overcome the burdens of judgment, we also cannot plausibly expect such people to reliably overcome the sources of unreason. Thus, we must expect that even intelligent and conscientious people will often reason about moral and political issues in biased and irrational ways. Since bias and irrationality undermine reasonableness, we must expect that such people will often hold unreasonable moral and political views. The upshot is that, as with reasonable disagreement, much (though certainly not all) of the unreasonable disagreement found in contemporary societies “is not a mere historical condition that may
soon pass away; it is a permanent feature of the public culture of democracy.”\textsuperscript{13} I call this the fact of unreasonable pluralism.

Before proceeding, it may be helpful to note explicitly two theses that I am not arguing for here. First, I am not arguing for the thesis, familiar from Plato and others, that ordinary citizens are too ignorant, irrational, or otherwise incompetent to rule themselves democratically.\textsuperscript{14} Even if, as I argue, intelligent and conscientious citizens are liable to reason about moral and political issues in biased and irrational ways, it is a separate question to what extent this fact spells doom for any form of democracy or speaks in favor of some non-democratic alternative. Second, I am not arguing that reasonable and unreasonable disagreement are morally equivalent, nor that we must respond to them in the same way. A normative theory may rightly treat the problems posed by reasonable disagreement as different from the problems posed by unreasonable disagreement. My point is only that if (for more or less the reasons that Rawls gives) reasonable disagreement is a fact of life that our normative theories must adjust to, then so too is unreasonable disagreement. I leave open the question of what form these adjustments must take,

\textsuperscript{13} Rawls, Political Liberalism, 36.
\textsuperscript{14} For Plato’s classic argument, see book VIII of Plato, Republic, trans. C.D.C Reeve (Indianapolis, IN: Hackett, 2004). For a more recent defense of this view, see Brennan, Against Democracy.
including the question of whether our principles of justice must be adjusted so that they can be endorsed by citizens who hold unreasonable views.\textsuperscript{15}

To set the stage for my argument, I begin by briefly rehearsing the standard Rawlsian account of the fact of reasonable pluralism.

### 3.1 The Fact of Reasonable Pluralism

The fact of reasonable pluralism is not just the familiar fact that reasonable disagreement is, as it stands, a pervasive feature of contemporary liberal societies. Rather, it is a fact (or prediction) about “the inevitable long-run result of the powers of human reason at work within the background of enduring free institutions.”\textsuperscript{16} As Charles Larmore puts it, “it is... the idea that reasonable people, precisely in virtue of exercising their reason in good faith and to the best of their abilities, tend to come to contrary opinions when they consider, especially in some detail, what it is to live well.”\textsuperscript{17} This idea

\textsuperscript{15} My aim is just to show that ideal theory must deal with the problems posed by unreasonable disagreement rather than setting those problems aside by means of idealizations. This does not entail that dealing with unreasonable disagreement requires adjusting our principles of justice so that they can be endorsed by people who hold unreasonable views.


is a radical departure from the view, common in the history of philosophy, that reason reliably leads to truth and thereby to agreement (at least in the long run). The basis for this departure is not that there is no truth to be found, but rather that, whatever the truth may be, we cannot plausibly expect that even intelligent and conscientious people will converge on it. It is thus not a claim about the limits or nature of morality itself, but rather about the limits of human reasoning. Because of these limits, we must expect that people will inevitably hold a diverse assortment of diverging, yet reasonable, moral and religious views.

The fact of reasonable pluralism constrains what our normative political theories may propose. Our theories must recognize that reasonable disagreement is an inevitable feature of any free society, and the principles, institutions, and practices our theories recommend must be compatible with this fact. This forestalls any theory that requires or presupposes that all citizens, or even just the majority, freely converge on a shared moral or religious doctrine. A theory cannot, for example, require that all citizens freely endorse utilitarianism, and this is so even if utilitarianism is the true moral theory. Rawls argues

18 As Burton Dreben puts it, that the fact of reasonable pluralism is “really an attack on the traditional view of reason: an attack on the idea that reasonable people can all (or at least sufficient numbers of them) be brought to agree solely through the use of reason on the same philosophical doctrine.” See Burton Dreben, “On Rawls and Political Liberalism,” in The Cambridge Companion to Rawls, ed. Samuel Freeman (Cambridge,UK: Cambridge University Press, 2003), 319.
that, given the fact of reasonable pluralism, a society premised on such agreement with a single moral doctrine “can be maintained only by the oppressive use of state power.”

Thus, no such moral doctrine can serve as the publically shared moral basis of a free society.

Why should we accept these claims? Why can we not hold out hope that, at least in the long run, reasonable people will freely converge on the truth about how we ought to live, or, if the case may be, on the conclusion that there is no such truth to be had? Rawls answers by appealing to “the many hazards involved in the correct (and conscientious) exercise of our powers of reason and judgment in the ordinary course of political life.”

These hazards, which Rawls refers to collectively as the burdens of judgment, are obstacles to discerning the truth and reaching agreement with others. They include:

a) The relevant empirical evidence is often complex, conflicting, and difficult to evaluate.

b) We may disagree about the weight that various evidence and considerations ought to receive.

c) Our concepts are vague, so we must rely on judgments and interpretations to decide hard cases.

---

20 Ibid., 56.
d) Our differing life experiences lead us to assess evidence and weigh values in different ways.

e) It is difficult to make all-things-considered judgments when there are different and competing normative considerations on either side of an issue.

f) Social institutions can typically only realize a subset of values, forcing us to make difficult choices as to how values should be prioritized.21

These obstacles make reasonable disagreement an ineliminable feature of any free society. As Jonathan Quong explains:

Put together, the burdens of judgment inevitably create reasonable disagreement over philosophical, moral, and religious issues. In a free society, it will simply prove impossible, Rawls argues, for rational people to overcome the burdens of judgment and all arrive at some common moral, religious, or philosophical perspective. Political philosophy thus needs to accommodate the fact of reasonable pluralism because it is the unavoidable product of rationality and freedom.22

My aim here is to show that a parallel case can be made for the inevitability of unreasonable disagreement. I begin in the next section by identifying some analogues of the burdens of judgment.

21 Rawls, Political Liberalism, 56-57.
22 Quong, Liberalism without Perfection, 194-195.
3.2 The Sources of Unreason

To explain the fact of reasonable pluralism, it is crucial that the burdens of judgment include only sources of disagreement that are “fully compatible with, and so [do] not impugn, the reasonableness of those who disagree.” Bias, irrationality, and other forms of distorted reasoning are thus not among the burdens of judgment. Rawls is explicit about this. He writes:

If we say [that the cause of a disagreement] is the presence of prejudice and bias, of self- and group-interest, of blindness and willfulness—not to mention irrationality and stupidity (often main causes of the decline and fall of nations)—we impugn the reasonableness of at least some of those who disagree.

For ease of reference, I will refer to the causes of disagreement that are inimical to reasonableness as the sources of unreason. I take them to include (among other things):

- g) our prejudices and biases, both explicit and implicit, that skew our beliefs and perceptions in unjustifiable ways
- h) the tendency of self- and group-interest to distort our judgments, especially those about justice and fairness

---

i)  denial of facts that are relevant, publically accessible, and well-established, such as well-confirmed scientific conclusions

j)  stubbornness and dogmatism that prevent us from changing our minds even when presented with sufficient reason to do so

k)  irrationality and fallacious reasoning

Like the burdens of judgment, the sources of unreason are obstacles to discerning the truth and reaching agreement. Unlike the burdens of judgment, the sources of unreason are also obstacles to being reasonable.

3.2.1 How the Sources of Unreason Make Us Unreasonable

How do the sources of unreason undermine reasonableness? One answer is that reasonableness includes an epistemic component: to be reasonable, one’s beliefs must respond to reasons and evidence in rational and unbiased ways. This seems to be Rawls’s view. As Samuel Freeman notes, “Rawls defines ‘reasonable comprehensive doctrines’

________________________

25 Rawls says that public reason may appeal to “the methods and conclusions of science when these are not controversial”. Rawls, Political Liberalism, 224. I omit the qualifier “when not controversial” in my formulation of i) because there are cases, such the reality of climate change, where the conclusions of science are controversial (among the general public, not among scientists) but about which I do not think there is scope for reasonable disagreement.

26 I read Rawls as requiring that reasonable doctrines meet some epistemic standards of coherence and reasons-responsiveness, but it is unclear how stringent the standards. See Rawls, Political Liberalism, 59.
epistemically, as doctrines that are responsive to evidence and possess certain other theoretical features.”  

Since the sources of unreason undermine such reason-responsiveness, they thereby undermine reasonableness as well. For example, the belief that climate change is “the greatest hoax ever perpetrated against the American people” is unreasonable because it depends on denying well-established facts and reflects a biased and irrational assessment of the available evidence.

However, many philosophers (including many Rawlsians) object to the epistemic component of reasonableness. They contend that views that are epistemically unreasonable—that is, views that are not sufficiently supported by reasons and evidence—nonetheless ought to be regarded as politically reasonable so long as they are consistent with the fundamental values of a liberal political regime. Martha Nussbaum, for example, argues that a wide range of popular doctrines, from astrology to Christianity, would wrongly be counted as unreasonable if assessed on their epistemic credentials. She

---

29 For a brief overview of the evidence, see *Climate Change: Evidence & Causes* (US National Academies of Sciences and The Royal Society, 2014).
30 I am indebted to Philip Shadd for pushing me on this point in his comments at the Canadian Philosophical Society’s 2016 Annual Congress.
writes, “So long as people are reasonable in the ethical sense [of respecting persons], why should the political conception denigrate them because they believe in astrology, or crystals, or the Trinity? Why not let them, and their beliefs, alone?”31 In a similar vein, Erin Kelly and Lionel McPherson argue that even “views that have little or no rational support or seem plainly irrational” ought to count as politically reasonable so long as they are consistent with accepting fair terms of social cooperation among free and equal people.32

It is not necessary to take a stand on this debate here. This is because the sources of unreason also undermine reasonableness in another way that both sides of the debate should recognize: The sources of unreason are prone to corrupt our sense of justice causing us to affirm distorted, and often erroneous, views about what justice requires. And when our beliefs about what is fair, just, and right are shaped by bias, self- or group-interest, ignorance, dogmatism, and irrationality, those beliefs are liable to be not only false, but pernicious. Misjudging what justice requires is not usually as politically harmless as misjudging whether astrology is true or whether crystals give off positive energy. If we misjudge what justice requires—and especially if our misjudgment is rooted in bias, self-interest, or the like—then we are likely to support (or propose) unjust laws,

policies, and institutions because we mistakenly believe them to be just. In such cases, our beliefs are unreasonable even by the strictly moral (or political) standards of Nussbaum, Kelly, and McPherson. This is because beliefs that support unjust laws, policies, and institutions cannot be consistent with accepting fair terms of social cooperation, or with respecting others as free and equal persons.

3.2.2 Two Types of Unreasonableness

It is important to recognize that one can support unjust laws, policies, and institutions while at the same time sincerely desiring that society be a fair system of cooperation among free and equal persons. It is useful here to distinguish two different ways that a person can be unreasonable. First, a person can be bereft of the basic moral commitments and motivations that are partially constitutive of reasonableness. Samuel Freeman describes unreasonable people in this way:

Politically unreasonable people are not willing to cooperate with others on fair terms; they insist on politically enforcing what they believe to be the “whole truth” even though they know that others reasonably disagree; they do not have a sense of justice or other moral dispositions. Anyone who has all these characteristics and all they entail is deeply flawed morally: For he or she has no respect for others with different values or for their rights, and thus is hardly fit for social life, at least not among people who do not think as he or she does.33

I call this *dispositional* unreasonableness to mark the fact that people who are unreasonable in this way lack the normal moral dispositions that (Rawls says) are possessed by all reasonable people.34

When I say that the sources of unreason undermine reasonableness, I do not mean that they are apt to turn us into dispositionally unreasonable people of the sort that Freeman describes. My point is that one need not be such a person in order to hold unreasonable moral or political views. As Jonathan Quong notes, “We must be aware… that it is possible (indeed very likely) that citizens who are generally reasonable may sometimes make unreasonable demands.”35 One can have the basic moral commitments and motivations of a reasonable person—respect for others, a concern for justice, and so on—but nonetheless make unreasonable judgments or hold unreasonable beliefs about some issue because one’s reasoning is distorted by bias, self-interest, or irrationality. One may genuinely desire to live on fair terms with others, but nonetheless hold self-serving and prejudicial beliefs about what constitutes fair terms. It is this kind of unreasonableness, which I call *judgment* unreasonableness, that I am primarily concerned with here.

35 Quong, *Liberalism without Perfection*, 291.
Judgement unreasonableness can manifest at many different levels, from judgments about abstract principles to judgments about particular laws, policies, institutions, decisions, or actions. One might believe that the abstract values of freedom, equality, and fairness are of paramount importance, but nonetheless interpret or weigh those values in unreasonable ways. For example, one might interpret equality such that it is consistent with state-sanctioned racial segregation, as the Supreme Court of the United States did in the case of *Plessy v. Ferguson*. Or, one might value tolerance in the abstract, but erroneously believe that certain actions are so wrong that they should not be tolerated. For example, as of 2016, more than a quarter of Americans believe that “gay and lesbian relations between consenting adults” should be forbidden by law. Or, one might affirm reasonable principles of justice, but hold unreasonable beliefs about their practical implications or about whether particular laws, policies, or institutions satisfy them. Often (but not always) this kind of judgment unreasonableness will be mediated by false non-

36 My thinking on this point owes much to the comments of David McCabe at the 2017 American Philosophical Association Central Division Meeting.
37 The court ruled that segregation laws were consistent with equal protection of political equality because such laws “do not necessarily imply the inferiority of either race to the other.” See *Plessy V. Ferguson*, vol. 163, US (Supreme Court of the United States, 1896). The judgment remained in effect for almost sixty years until overturned in *Brown V. Board of Education*, vol. 347, US (Supreme Court of the United States, 1954).
38 “Gay and Lesbian Rights.”
moral beliefs. For example, one might accept a principle of “just savings” that requires that each generation preserve the conditions needed to maintain a just society over time, but deny that this requires taking any action to address climate change because one is among the twenty percent of Americans who believe that there is “no solid evidence” that climate change is occurring.

Later, I will say more about how the sources of unreason afflict the reasoning of otherwise reasonable people. But first, I turn to a different question: Are the sources of unreason—and the disagreements that result from them—the kind of thing that ideal theory must take into account?

3.3 The Limits of Ideal Theory

Rawls does not deny that bias and irrationality explain much of the moral and political disagreement that exists in actual societies. However, he suggests that such disagreements are not the proper concern of ideal theory. He writes:

39 As Allen Buchanan argues in a different context, “…wrong behavior often occurs not because people are utterly lacking in the moral virtues or because their most basic moral principles are defective, but because they… systematically misapply unexceptionable moral principles because they hold certain false empirical beliefs.” Allen Buchanan, "Social Moral Epistemology," Social Philosophy & Policy 19, no. 2 (2002): 150.
41 Funk and Kennedy, The Politics of Climate.
We might suppose, say, that most people hold views that advance their own more narrow interests; and since their interests are different, so are their views. Or perhaps people are often irrational and not very bright, and this mixed with logical errors leads to conflicting opinions. But while such explanations explain much, they are too easy and not the kind we want. We want to know how reasonable disagreement is possible, for we always work at first within ideal theory.42

What is ideal theory, and how might it license setting aside disagreements rooted in the sources of unreason?

In its Rawlsian form, ideal theory aims to tell us “what a perfectly just society would be like.”43 To this end, ideal theory assumes that certain favorable conditions obtain and then asks what would constitute a perfectly just (or “well-ordered”) society under those conditions. For Rawls, such favorable conditions include that all citizens fully comply with the requirements of justice, and that social and economic conditions are sufficiently good to sustain a just and stable liberal society.44 Of course, full compliance

42 Rawls, Political Liberalism, 55.
43 Rawls, A Theory of Justice, 8. Rawls’s version of the distinction between ideal and non-ideal theory, while probably the most popular, is not the only version. Different philosophers draw the line between ideal and non-ideal theory in different places and along different dimensions. Some deny that there is a clean border to be drawn; only degrees of idealness to be evaluated along multiple dimensions. While I focus on Rawlsian ideal theory here, I return to some of the competing versions, especially that due to G.A. Cohen, in Section 3.5.1. For a general overview of the ideal/non-ideal distinction and related debates, see Zofia Stemplowska and Adam Swift, "Ideal and Nonideal Theory," in The Oxford Handbook of Political Philosophy, ed. David Estlund (Oxford: Oxford University Press, 2012). For useful taxonomies of the different ways in which the distinction is drawn, see Alan Hamlin and Zofia Stemplowska, "Theory, Ideal Theory and the Theory of Ideals," Political Studies Review 10 (2012); Laura Valentini, "Ideal Vs. Non-Ideal Theory: A Conceptual Map," Philosophy Compass 7 (2012).
44 Rawls, A Theory of Justice, 8, 216; Rawls, Justice as Fairness, 47.
does not exist in any actual society, and many societies face social or economic conditions that undermine their justice and stability. However, assuming favorable conditions is not intended to accurately represent the world as it is. Rather, the purpose of such assumptions is to abstract away from the contingent features of actual societies that we recognize are inimical to perfect justice. John Simmons encapsulates this idea nicely. He writes:

Rawls’s ultimate target—the ideal of his ideal theory—is the most just institutional structure that can be achieved within the constraints set by the more or less intractable aspects of human nature and the natural world. Fleshing out and arguing for such an ideal requires the use of counterfactual assumptions, since the current state of our institutions—and the resulting social and moral condition of the persons living under them (not to mention the ways in which the world is being used)—is so distant from the best we can realistically hope to achieve.45

The use of counterfactual assumptions frees the ideal theorist to develop a picture of a perfectly just society that is unafflicted by the contingent obstacles to justice found in the real world. Rawls insists that only once we have worked out such an idealized vision of society can we go on to do non-ideal theory, which tells us what to do “under less happy conditions.”46

3.3.1 Ideal Theory and Unreasonable Pluralism

In actual liberal democracies, unreasonable moral and political views are sufficiently common, both among the general public and among their elected officials, that they undermine the justice of society. Rawls assumes that this is one of the unfavorable conditions, like imperfect compliance and poor economic conditions, that ideal theory may idealize away. He assumes that in a well-ordered society, “unreasonable comprehensive doctrines do not gain enough currency to undermine society’s essential justice.”47 In a later footnote he writes:

Of course, every society also includes numerous unreasonable doctrines. Yet... I am concerned with an ideal normative conception of democratic government, that is, with the conduct of its reasonable citizens and the principles they follow assuming them to be dominant and controlling.48

Jonathan Quong explains the justification for these assumptions. He says that political liberalism, as an ideal theory, need only concern itself with “the kind of citizens who would be raised in a society well-ordered by a liberal conception of justice.”49 Such citizens, he says, constitute “an idealized constituency of reasonable persons.”50 The underlying thought here seems to be this: While actual societies contain many biased, irrational, and ignorant people, such people would be rare (maybe even non-existent) in

47 Rawls, Political Liberalism, 39.
48 Ibid., 441 note 3. My emphasis.
49 Quong, Liberalism without Perfection, 158.
50 Ibid., 159.
an ideal liberal society. So, when we are developing a normative vision of the principles and institutions of an ideal liberal society, it is perfectly acceptable to assume that the people of that society are overwhelmingly (or even entirely) reasonable.

However, if ideal theory allows us to idealize away most of the unreasonable disagreement found in actual societies, why can’t we idealize away most of the reasonable disagreement as well? Recall that the fact of reasonable pluralism is supposed to constrain the range of viable political ideals. It is supposed to rule out, for example, “a society united on a reasonable form of utilitarianism, or on the reasonable liberalisms of Kant or Mill” because, given the inevitability of reasonable moral disagreement, such a society could only be maintained by the “oppressive use of state power.” But why can’t a proponent of, for example, utilitarianism simply say of anti-utilitarian views exactly what Rawls and Quong say of unreasonable views? That is, why can’t a utilitarian grant that anti-utilitarian views are pervasive in the real world, but assume that such views would be rare or non-existent in an ideal society? To answer these questions, one must do more than simply repeat that reasonable moral disagreement is inevitable because of the burdens of judgment. What we need to know is why ideal theory must accommodate this fact. If ideal theory can posit an idealized citizenry that is overwhelmingly reasonable, then why can’t

____________________

51 Rawls, Political Liberalism, 37.
it posit an idealized citizenry that overcomes the burdens of judgment and converges on the moral truth, whatever it may be?

3.3.2 The Constraints of a Realistic Utopia

Rawls’s answer, simply put, is that to do so would be too unrealistic.\textsuperscript{52} The task of ideal theory, at least as Rawls sees it, is to depict a realistic utopia: “a reasonable and just society”\textsuperscript{53} that “is feasible and might actually exist, if not now then at some future time under happier circumstances.”\textsuperscript{54} A realistic utopia may depart from the actual world in significant ways, but it must be “an achievable social world.”\textsuperscript{55} So, while ideal theory can—indeed, must—employ idealizing assumptions, it cannot assume conditions that are better (or more idealized) than “the best foreseeable conditions.”\textsuperscript{56} Rawls does not tell us exactly what he means by “feasible”, “achievable”, or “foreseeable”, but it is clear that a society that satisfies these terms cannot require us to transcend “the fixed constraints of human life”\textsuperscript{57} which include “the actual laws of nature”\textsuperscript{58} as well as “general facts of moral
psychology.” Borrowing from Rousseau, Rawls says that ideal theory must “[take] people as they are… and constitutional and civil laws as they might be.”

The requirement that ideal theory be realistically utopian explains why it must face up to the fact of reasonable pluralism. Since we are incapable of reliably overcoming the burdens of judgment, a world without widespread reasonable disagreement about moral and religious issues is not an achievable social world. As Rawls puts it, “the fact of reasonable pluralism limits what is practically possible.” Thus, a vision of society that presupposes or requires that all citizens freely converge on a shared moral or religious doctrine is not sufficiently realistic. This idea was central to the evolution of Rawls’s own views. He came to regard the notion of a well-ordered society employed in *A Theory of Justice* as objectionably “unrealistic” precisely because it assumed that all citizens would stably affirm a shared moral doctrine. In his later work, Rawls introduced several key changes “to make the idea of a well-ordered society more realistic and to adjust it to the historical and social conditions of democratic societies, which include the fact of reasonable pluralism.”

---

61 Ibid., 12.
Whether political philosophy must be realistic in the way that Rawls’s requires is controversial. As we saw in the previous chapter, David Estlund contends that normative theories may soundly require things that we know for certain will never happen.64 G.A. Cohen goes even further, arguing that ultimate principles of justice cannot depend on any non-moral facts, and that Rawls errs in adjusting his conception of justice to account for human nature.65 But as we saw in Estlund’s case, it is difficult to do away with these constraints without at the same time doing away with the very problems that our normative theories are supposed to solve. The fact of reasonable pluralism is a prime example. Without constraints on ideal theory, there would be nothing to prevent ideal theory from idealizing away reasonable disagreement in the same way that it idealizes away many other conditions found in actual societies. In other words, considerations of realism are essential to framing the problem of reasonable pluralism that has been central to liberal political philosophy for at least three decades. Absent such considerations, one cannot explain why ideal theory must face up to the fact of reasonable pluralism. So, rejecting the “realistic” part of “realistically utopian” is not an option for anyone who believes that ideal theory must take account of the fact of reasonable pluralism.

64 Estlund, “Utopophobia.”
3.3.3 Applying the Constraints

Once one accepts some constraints on ideal theory, one must apply them consistently. One cannot invoke the constraints in the course of explaining why normative theories must account for reasonable pluralism while at the same time building one’s own normative theory on the basis of assumptions that violate the constraints. If our theories must accommodate the fact of reasonable pluralism because it is ineliminable even in the best foreseeable conditions, then our theories must also accommodate any other facts that are similarly ineliminable in the best foreseeable conditions. More specifically, if widespread unreasonable disagreement rooted in the sources of unreason is ineliminable in the best foreseeable conditions, then ideal theory cannot assume that such disagreements would be rare or non-existent in an ideal society.

Of course, it is difficult to know what is achievable under the best foreseeable conditions. As Rawls notes, “we have to rely on conjecture and speculation.”66 In the next section, I conjecture that a world without widespread unreasonable disagreement is not an achievable social world, at least not under any conditions that could plausibly count as foreseeable. I argue that the available evidence gives us good reason to believe that even intelligent and conscientious citizens cannot reliably overcome the sources of

unreason. At the very least, I contend, the sources of unreason are not significantly easier to overcome than the burdens of judgment, so if we cannot foresee overcoming the burdens of judgment, then we also should not foresee overcoming the sources of unreason. Thus, just as we must expect reasonable disagreement to persist in an ideal society because of the burdens of judgment, we must also expect unreasonable disagreement to persist because of the sources of unreason.

3.4 Can we Overcome the Sources of Unreason?

The sources of unreason include prejudices and biases, the distorting effects of self- and group-interest, denial of well-established facts, stubbornness, dogmatism, irrationality, and fallacious reasoning. In section 3.2, I claimed that these are liable to distort our sense of justice. I now need to explain why reliably overcoming the sources of unreason is not something we can plausibly expect of even intelligent and conscientious people. But first, let me grant that people could be more informed, rational, and unbiased than they are now. I am not trying to show that we cannot improve on the status quo. Rather, my aim is to show that, even under the best foreseeable conditions, there are limits to how much improvement we can plausibly expect.

I am also not trying to show that the sources of unreason are so insurmountable that nobody can ever overcome them in any particular case. All that my argument requires
is that people are unable to overcome them reliably enough to eliminate the vast majority of unreasonable disagreements. The relevant comparison here is the burdens of judgment. The burdens of judgment are not so insurmountable that nobody can ever overcome them in any particular case. The existence of complex and conflicting evidence does not make it impossible for us to sometimes converge on the truth. Nor do our differing life experiences preclude us from ever agreeing. We often converge on truths despite the burdens of judgment. The reason that the burdens of judgment inevitably create widespread reasonable disagreement is not that there are no cases in which people are capable of overcoming them. Rather, it is because people cannot overcome them reliably enough to settle many disagreements about how we ought to live. My aim is to show that the same is true of the sources of unreason: they inevitably create widespread unreasonable disagreement, not because there are no cases in which people can overcome them, but rather because people cannot overcome them with sufficient reliability.

There are two lines of evidence that suggest that we cannot expect even intelligent and conscientious people to reliably overcome the burdens of judgment. The first is the historical record and the second is the empirical literature on tribalism, cognitive biases, and motivated reasoning. Previous chapters have already discussed the latter, so I will only briefly summarize and add to that evidence here. But first, let us consider what we can learn from history.
3.4.1 The Historical Record

To justify his claim that reasonable pluralism is a permanent fact of life, Rawls points largely to the historical record. In the introduction to Political Liberalism, he recounts how the Reformation “fragmented the religious unity of the Middle Ages and led to religious pluralism, with all its consequences for later centuries.”67 He notes that despite principles and efforts designed to settle the ensuing religious disagreements, “the fact of religious divisions remains.”68 Immediately after noting this, he writes, “For this reason, political liberalism assumes the fact of reasonable pluralism as a pluralism of comprehensive doctrines, including both religious and non-religious doctrines.”69 So although the fact of reasonable pluralism is a claim about “the inevitable long-run result of the powers of human reason,” much of the evidence for it consists of looking at the outcomes that human reason has produced in the past.70

Does the historical record give us reason to be optimistic about the prospects of intelligent and conscientious people reliably overcoming the sources of unreason? I do not think it does. History is marred by countless examples of masses of ordinary, intelligent, well-meaning people supporting morally abhorrent laws, institutions, and

67 Rawls, Political Liberalism, xxii.
68 Ibid., xxiv.
69 Ibid.
70 Ibid., 4.
regimes. Indeed, even the most educated and intelligent members of society—the ones we are inclined to think ought to have known better—have endorsed all manner of moral and political views inimical to justice. Achen and Bartels succinctly recount just a few of the most egregious cases:

In the antebellum era, prominent southern professors and university administrators often defended slavery. Brilliant 19th-century German professors helped give shape to German nationalism and the racial identity theories that led to Nazism, and German university students in the 1930s were often enthusiastic supporters of Hitler. Protestant and secular professors backed Otto von Bismark’s campaign to suppress the civil liberties of Catholics in 19th-century Germany. Crude prejudice against Catholics, Jews, and others was common among American intellectuals until recent decades, too.

More recently, 20th-century communism attracted many highly educated people around the world. Numerous French intellectuals supported Russian communism well after its crimes had been exposed. Radical Chinese intellectuals backed Mao Zedong’s campaign to establish his regime and keep it in power—a regime that eventually became, not just a relentless oppressor of individuals, but the most murderous government in the history of the world. And further examples are not hard to find.

71 Achen and Bartels, Democracy for Realists, 311.
72 For example, Allen Buchanan points out that throughout the first half of the twentieth century (and even into the second half) “coercive negative eugenics flourished in liberal societies as well as deeply illiberal societies. Large-scale forced eugenic sterilization occurred in the U.S., Sweden, Denmark, Finland, and Canada, not just in Nazi Germany.” And it was often doctors and scientists leading the charge, armed with false factual beliefs about natural differences between people and the importance of preventing the degeneration of society. Allen Buchanan, "Institutions, Beliefs and Ethics: Eugenics as a Case Study," The Journal of Political Philosophy 15, no. 1 (2007): 38.
One might suggest that all these people were dispositionally unreasonable—that they lacked a sense of justice or the moral motivation to live with others on just terms. But this is implausible. Indeed, one of the most disturbing aspects of these dark parts of history is that the participants often were, in a perverse way, motivated by morality—they were doing what they firmly (though erroneously) believed was morally right and just.\(^73\) The problem in such cases is not that people are utterly lacking in moral scruples, but rather that their sense of right and wrong is distorted by prejudice, misplaced deference to authorities, and false factual beliefs.\(^74\) Nor it is plausible to think the participants would all have seen the wrongness of their beliefs and actions if only they had reasoned a little harder. These cases are thus primarily what I have called cases of judgment unreasonableness—they are cases of otherwise reasonable people going astray in their moral judgments due to distorted reasoning.

Of course, the historical record also shows that moral progress is possible, and I certainly do not deny this. My point is not that we cannot make any progress, nor that we can never overcome whatever abhorrent moral and political views happen to be in

---


\(^74\) This is a point that Allen Buchanan has emphasized repeatedly. See Buchanan, "Social Moral Epistemology,"; Buchanan, "Institutions, Beliefs and Ethics: Eugenics as a Case Study,"; and Allen Buchanan, "Social Moral Epistemology and the Tasks of Ethics," in *Ethics and Humanity: Themes from the Philosophy of Jonathan Glover*, ed. N. Ann Davis, Richard Keshen, and Jeff McMahan (Oxford: Oxford University Press, 2010).
currency at any given time. Rather, my point is that the historical record gives us strong evidence that even intelligent and conscientious people who apply their powers of reason are not bound to exclusively affirm reasonable moral and religious views. Just as the exercise of reason does not reliably lead to truth and agreement about moral and religious matters, it also does not reliably lead to being reasonable.

3.4.2 The Pitfalls of Human Reason

To understand the prevalence of unreasonable views, both throughout history and today, I think we need to recognize that the sources of unreason—bias, prejudice, dogmatism, irrationality—are very difficult to avoid, even for even intelligent and conscientious people. And, I submit, the reason that they are very difficult to avoid is that they are deeply ingrained in human psychology. (In fact, it may not even just be human psychology; recent experiments with capuchin monkeys, bonobos, and chimpanzees have shown that they are prone to some of the same cognitive biases we are, suggesting that at least some of our cognitive biases are evolutionarily very old.75) In Chapter 1 we saw that tribalism and motivated reasoning play prominent roles in shaping people’s political

---

views. And, as I noted in Chapter 2, these are far from the only factors that distort our reasoning and judgment. Decades of empirical research on human reasoning and judgment has uncovered a myriad of ways in which our cognitive capacities are prone to systematic errors. We are, in behavioral economist Dan Ariely’s words, “predictably irrational.”76 This is not to say that we are utterly incapable of rational thought; only that our capacity for such thought is susceptible to being led astray by our interests, prejudices, partisan loyalties, and by the many mental short-cuts or heuristics that our brain relies on to process and make sense of the world around us.77 While we can sometimes overcome these obstacles, it is difficult to overcome them reliably because they usually operate outside our awareness.

Even brilliant individuals who subscribe to sound moral principles are prone to go awry when it suits their prejudices or their self-interest. Consider Thomas Jefferson, the author of the words “all men are created equal,” yet also a slaver-owner and a defender of slavery. After examining Jefferson’s writings about slavery, Hugo Mercier and Dan Sperber remark on how such a brilliant and in many ways progressive figure came to justify the continuance of such a heinous practice. They write, “Jefferson, armed

with a brilliant intellect, all the knowledge of his time, and the noblest of ideals, should have reasoned his way to the right creed and just behavior. Instead, reason provided him with convenient rationalizations, allowing him to keep his slaves and his wealth.”

Why is biased reasoning such a robust phenomenon? One answer may lie in the function of human reasoning. According to a compelling theory developed by Mercier and Sperber, reasoning did not evolve to track the truth, but rather to “devise and evaluate arguments intended to persuade.” To perform this task, reasoning is “designed” to generate reasons and arguments in favor of our own views, and to detect the flaws in arguments for opposing views. It is not “designed” to detect the flaws in our own arguments, or to generate arguments for opposing views. While this often leads to folly when people reason alone, it can be adaptive at the group level. Mercier and Sperber write:

When a group has to solve a problem, it is much more efficient if each individual looks mostly for arguments supporting a given solution. They can then present these arguments to the group, to be tested by the other members. This method will work as long as people can be swayed by good arguments…

---

80 Ibid., 65.
And Mercier and Sperber present evidence that, in many cases, people are often swayed by good arguments.

The trouble is that this evidence comes overwhelmingly from experiments in which groups were tasked with solving logic or math puzzles. In such cases, there are clear criteria defining what count as correct answers, and everyone in the group shares the goal of arriving at such answers. Such settings are very unlike most moral or political debates. Few people have partisan allegiances or prejudices about the answers to math problems. And unlike math problems, in morality and politics, what counts as a correct answer is frequently part of what is in dispute. Because of this, political reasoning tends to exhibit the negative results of biased reasoning rather than the positive results observed in such studies. Indeed, Mercier and Sperber write:

Many experiments, and before them, countless personal and historical observations have rendered the intellectualist view of moral reason implausible. Moral judgments and decisions are quite commonly dominated by intuitions and emotions with reason providing, at best, inert rationalizations and, at worst, excuses that allow the reasoner to engage in morally dubious behavior... Reason does what it is expected to do as a biased and lazy producer of justifications.81

This does not mean that moral reasoning never leads to truth, nor that moral argument and debate is a futile endeavor. But it does mean we should be leery of the idea that even

the careful exercise of human reason will reliably lead to *reasonable* moral and political views.

Ultimately, I think that history together with the empirical literature I have surveyed beginning in Chapter 1 give us compelling reasons to believe that we are very unlikely to reliably overcome the sources of unreason, at least under any conditions that could plausibly count as foreseeable. At the very least, the distorting influence of our biases, prejudices, interests, and partisan loyalties does not seem to be much easier to overcome than the burdens of judgment that Rawls cites to explain the inevitability of reasonable disagreement, such as complex and conflicting evidence and differing life experiences. So, if citizens cannot be expected to reliably overcome the burdens of judgment even in the best foreseeable conditions, then we should not expect them to reliably overcome the sources of unreason either. And if we cannot expect people to reliably overcome the sources of unreason in even the best foreseeable conditions, then ideal theory cannot assume that the citizens of an ideal society will be overwhelmingly reasonable.

3.4.3 *Am I an evoconservative?*

In a recent article, Allen Buchanan and Russell Powell take aim at theorists they call *evoconservatives* who “[appeal] to the prevailing evolutionary explanation of morality
to show that it is unrealistic to think that cosmopolitan and other ‘inclusivist’ moral ideals can be meaningfully realized.”

Buchanan and Powell point out that contemporary morality is already far more inclusive than we would expect if it is tightly constrained by its evolutionary roots in the way that such evoconservatives suggest. Among the examples they point to are growing concern for the well-being of non-human animals and the establishment of a global human rights culture. To understand these inclusivist features of morality, Buchanan and Powell argue, we must recognize the capacity of humans to critically reflect on and revise their moral standards and to update their behavior and institutions accordingly. Buchanan and Powell conclude that “human beings have not slipped the ‘leash’ of evolution, but … the leash is far longer than evoconservatives and even many evolutionary anthropologists have acknowledged—and no one is in a position at present to know just how elastic it will turn out to be.”

It might be thought that my appeal to history and empirical research to show that we cannot overcome the sources of unreason is subject to a similar critique. However, there is an important difference between my argument and the argument of evoconservatives. The evoconservatives Buchanan and Powell criticize claim that “facts

---

83 Ibid., 48-51.
84 Ibid., 67.
about human evolutionary history significantly constrain the shape of plausible moralities and the scope of other-regarding concern” and that this shows that “cosmopolitan and other inclusivist moral principles are not appropriate or realistic for beings like us.”85 In other words, such evoconservatives contend that evolution places firm limits on the content of morality itself. I am not proposing any such hard limits on the content of morality. My point is only that we face obstacles to being reasonable that are very difficult to reliably surmount, and that ideal theory cannot simply assume that intelligent and conscientious people will reliably surmount these obstacles and arrive exclusively at reasonable moral and religious views. This point is quite compatible with recognizing the human capacity to revise our moral standards in ways that take human morality well beyond its evolved functions.

3.5. Conclusion and Implications

I set out to demonstrate a conditional: If one accepts the standard story about how and why the fact of reasonable pluralism is an inevitable fact of life that constrains ideal theory, then one must accept that unreasonable pluralism is also an inevitable fact of life that constrains ideal theory. I have argued that we face obstacles, the sources of unreason,

85 Buchanan and Powell, “The Limits of Evolutionary Explanations of Morality and Their Implications for Moral Progress,” 44.
that make even intelligent and conscientious people prone to hold unreasonable moral and political views. And I argued that these obstacles are not significantly easier to surmount than the causes of reasonable disagreement, the burdens of judgment. So, if we cannot expect citizens to reliably overcome the burdens of judgment, we also should not expect them to reliably overcome the sources of unreason. Thus, if the foreseeability constraint precludes us from idealizing away reasonable disagreements caused by the burdens of judgment, it must also preclude us from idealizing away unreasonable disagreements caused by the sources of unreason. Therefore, if reasonable disagreement is “a permanent feature of the public culture of democracy”86 that constrains the range of viable normative theories, so too is unreasonable disagreement. I call this the fact of unreasonable pluralism. This conclusion raises several interesting challenges and questions for liberal political philosophy and for ideal theory. I will end by briefly discussing just two of them.

3.5.1 Implications for Ideal Theory

First, what does this conclusion mean for ideal theory? I have argued that the constraints on ideal theory that are typically invoked to explain why even ideal theories

86 Rawls, Political Liberalism, 36.
must face up to the challenge of reasonable pluralism turn out, when applied consistently, to also require that our ideal theories face up to unreasonable pluralism rooted in bias and irrationality. To some, this might seem to be a reductio of ideal theory: If ideal theory must grapple with something as seemingly non-ideal as unreasonable pluralism, is it still ideal theory? Does it even make sense to call a society in which unreasonable disagreement is pervasive an ideal society?

These questions become all the more pressing if my basic argumentative strategy generalizes to many other standard assumptions of ideal theory. Consider, for example, Rawls’s assumption of full compliance with the demands of justice by all citizens. Rawls believes that full compliance is sufficiently achievable because:

…men’s propensity to injustice is not a permanent aspect of community life; it is greater or lesser depending in large part on social institutions, and in particular on whether these are just or unjust. A well-ordered society tends to eliminate or at least to control men’s inclinations to injustice.”

However, one could argue that a world where everyone complies with the demands of justice is no more foreseeable than a world where everyone overcomes the burdens of judgment and converges on the truth about morality and religion. In fact, the sources of unreason may provide grounds for such an argument. To comply with the demands of justice, one must have an accurate and unbiased view of what justice requires, and that is

---

exactly what the sources of unreason make it difficult for us to do. And, even when we do have a relatively clear view of what justice or morality demands, we are very good at rationalizing our unjust and immoral behavior so that, from our own perspective, it does not seem unjust and immoral. Given these facts, assuming full compliance may also run afoul of Rawls’s own constraints on ideal theory.

Zooming out a bit, the issue here is that the idealizations of ideal theory must be made in a principled way. If we insist that reasonable pluralism cannot be idealized away because to do so would violate certain constraints, then we must apply those constraints consistently when determining which other conditions we can idealize away. It may turn out that the constraints necessary to force ideal theorists to confront the fact of reasonable pluralism, when applied consistently, also force ideal theorists to confront much of what was previously thought to be the domain of non-ideal theory. And if it turns out that ideal theory must deal with partial compliance, unreasonable disagreement, and a host of other seemingly “non-ideal” problems, then we can legitimately begin to wonder whether the distinction between ideal theory and non-ideal theory remains worth making.

---

89 That said, there may be special reasons for assuming full compliance which justify making an exemption to constraints. John Simmons provides an argument to this effect in Simmons, “Ideal and Nonideal Theory,” 8-9.
However, one could also go the other direction and suggest that what my argument shows is that Rawlsian ideal theory is not ideal enough. The mistake, one might say, was introducing constraints of realism in the first place. G.A. Cohen argues for a version of this view, insisting that principles of justice must be “fact-insensitive.”\(^90\) Accepting Cohen’s view keeps the fact of unreasonable pluralism out of the domain of facts that constrain our ideal theories, but it keeps all other (non-moral) facts out as well, including the fact of reasonable pluralism. I will not attempt to settle this issue here. For now, I simply note that the line between ideal and non-ideal theory cannot be drawn where Rawls tried to draw it.\(^91\)

### 3.5.2 The Stability of a Liberal Society

A second, and I think deeper issue, concerns the question of stability. In A Theory of Justice, Rawls tells us that “it is an important feature of a conception of justice that it should generate its own support,” and that it is “a consideration against a conception of justice that in view of the laws of moral psychology, men would not acquire a desire to

---

\(^90\) Cohen, Rescuing Justice and Equality, 232.

\(^91\) My own, still nascent, view is that what we need is a more complex picture that allows for more levels and dimensions of “idealness”. For proposals along these lines, see Hamlin and Stemplowska, Theory, Ideal Theory and the Theory of Ideals; Joseph Heath, Morality, Competition, and the Firm (New York: Oxford University Press, 2014), 175-182.
act upon it even when the institutions of their society satisfied it.” The idea here is the following: Suppose we have a society that, at time $t$, perfectly satisfies a liberal conception of justice. The question of stability is, “What happens when we roll the clock forward?” Are there forces within the society that ensure it remains at a just equilibrium, continuing to satisfy the conception of justice? Or are there inevitable forces within the society that will cause it to deviate from justice? 

This is what makes the fact of reasonable pluralism a challenge for liberal theory. The challenge is this: If citizens of a liberal society will inevitably affirm a diverse array of reasonable, but incompatible, moral and religious views, there may not be a conception of justice that all citizens can freely affirm. And if not all citizens can affirm the same conception of justice, then it may not be possible for a society to stably realize any such conception. Rawls’s later work tries to meet this challenge by showing how a liberal conception of justice could be affirmed by all reasonable citizens despite their disagreements about morality and religion.

---

92 Rawls, A Theory of Justice, 119.
94 The task Rawls sets himself in Political Liberalism is to show “how it is possible for there to exist over time a stable and just society of free and equal citizens, who remain profoundly divided by reasonable religious, philosophical, and moral doctrines.” Rawls, Political Liberalism, 4.
Rawls insists that a stable society cannot rest on a mere modus vivendi – a compromise or “equilibrium point” that is “founded on self- or group-interests, or on the outcome of political bargaining.”95 He argues that even if such a modus vivendi could temporarily secure justice and equality, it would remain an inadequate basis for a liberal democratic society because “its stability is contingent on circumstances remaining such as to not upset the fortunate convergence of interests.”96 To be sufficiently stable, and stable for the right reasons, Rawls argues that an agreement on the fundamental norms, principles, or values of a liberal society must be what he calls an overlapping consensus of reasonable doctrines. In such a consensus, reasonable citizens endorse the overlapping consensus for reasons grounded in, or at least consistent with, their own moral, religious, and philosophical doctrines. Because they endorse it on the basis of such reasons, those who accept the overlapping consensus “will not withdraw their support of it should the relative strength of their view in society increase and eventually become dominant.”97 This, says Rawls, makes an overlapping consensus “quite different from a modus vivendi”98 because it is not hostage to “a balance of forces in contingent and possibly fluctuating circumstances.”99

95 Rawls, Political Liberalism, 147.
96 Ibid.
97 Ibid., 148.
98 Ibid., 147.
99 Ibid., 392.
However, the stability of an overlapping consensus depends on keeping unreasonable disagreement out of the picture; it depends on a balance of power that heavily favors reasonable citizens. Rawls simply assumes that such a balance of power will exist. But, as I have argued, this assumption is inconsistent with the reasoning behind the fact of reasonable pluralism. If we accept this reasoning, then we must also accept that unreasonable doctrines will be pervasive in any free society. And in the absence of a reason to believe that reasonable citizens will always hold the balance of power, the stability of an overlapping consensus must be taken to depend on exactly the sort of “contingent and possibly fluctuating” circumstances that a modus vivendi does. In other words, once we accept the fact of unreasonable pluralism, it is unclear whether the balance of power needed for a stable overlapping consensus can reliably be maintained. This calls into question whether an overlapping consensus can provide the strong kind of stability Rawls sought – stability that is robust against “changes in the distribution of power among views.”

More generally, insofar as unreasonable pluralism undermines the justice and stability of a liberal democratic regime, a liberal conception of justice seems to generate internal threats to its own justice and stability. This does not entail that we must reject

---

101 Ibid., 148.
liberalism, but it does show that liberalism is inherently fragile. The very rights and freedoms that are central to liberalism also create conditions under which threats to those rights and freedoms can flourish. This, no less than the fact of reasonable pluralism, is a central problem that theorists of liberal democracy ought to be grappling with.
4. Making Dirty Compromises with Clean Hands

I had hoped the day of compromise with wrong had gone forever. Ample experience shows that it is the least practical mode of settling questions involving moral principle. A moral principle cannot be compromised. – Charles Sumner

Believing then, that this is the best proposition that can be made effectual, I accept it. I shall not be driven by clamor or denunciation to throw away a great good because it is not perfect. I will take all I can get in the cause of humanity and leave it to be perfected by better men in better times. – Thaddeus Stevens

Accomplishing things democratically often requires compromise. Given the persistence and pervasiveness of political disagreement, it is frequently impossible to secure the support of a majority without making concessions. But making concessions can be morally perilous, especially when the justice of society is at stake. In this chapter, I examine several ethical dimensions of what I call dirty compromises: compromises that involve making serious concessions to injustice.

My first task, which I take up in the next section, is set out more carefully what counts as a dirty compromise. Before moving on to some questions about the ethics of dirty compromise, I then describe a real-world example: Vermont’s 2000 compromise on

---

same-sex marriage. With both a definition and an example in hand, I then turn to three questions: (1) Are dirty compromises ever morally acceptable? (2) Can a person who opposes injustice make dirty compromises without sacrificing their integrity? And, (3) do those who make dirty compromises necessarily end up with “dirty hands”? Using the Vermont case as an example, I lay out a framework that allows us to see what is morally wrong with dirty compromises, while at the same time showing that it is sometimes morally acceptable to agree to them. I also show that agreeing to them need not cost advocates of justice their integrity or the cleanliness of their hands. At the end of the chapter, I raise some further issues involving hard cases.

4.1 What is a dirty compromise?

A dirty compromise is one in which at least one party makes a concession to injustice by agreeing to a law, policy, or course of action that it considers to be seriously unjust. In this section, I explain this definition in greater detail, and distinguish dirty compromise from moral compromise more generally and from what Avashai Margalit calls rotten compromise.
The word “compromise” can be used in at least three different ways. First, it can function as a noun referring to an agreement as in “The District of Maine was admitted to the Union as part of a compromise that permitted slavery in Missouri.” Second, it can function as a verb referring to the act or process of reaching such an agreement as in “Reagan and O’Neill compromised to reform social security.” Third, it can function as a verb indicating that a person or thing has been undermined or degraded as in “Trump has compromised U.S. relations with important allies.” All three uses of “compromise” are of interest here, but for now I will focus on the first, compromise as an agreement.

As I understand it here, a compromise is an agreement between two or more parties. I use the word “parties” to refer to the individuals or groups that constitute the different sides of a dispute, rather than just to formal political parties. What makes an agreement a compromise, as opposed to some other kind of agreement, is that each of the parties makes some concession from their most preferred alternative in order to reach an agreement with the other parties. This distinguishes a compromise from a consensus on

---

3 For a discussion of these three different uses of “compromise”, see Martin Benjamin, *Splitting the Difference: Compromise and Integrity in Ethics and Politics* (Lawrence, KS: Kansas University Press, 1990), 4-10.

4 The parties to a compromise may be formal political parties, but they may also be individuals, interest groups, states, or any other kind of agent that can enter into agreements.

one hand, and from a capitulation on the other. In a consensus, all of the parties agree about which alternative is best, and no party makes any concessions to the others. The parties may disagree about why their preferred alternative is best, but they nonetheless converge on the same alternative as their most preferred alternative. No party to a consensus gives up anything by being part of the consensus. This is not so in a compromise. In a compromise, the parties continue to disagree about what the best alternative is, and each side gives up something in order to secure the agreement of the other parties. As a result of these concessions, the outcome of the compromise typically lies somewhere between what each party believes would be best. The fact that all parties give up something is what distinguishes a compromise from a capitulation. In a capitulation, one party gets its most preferred alternative because the other parties, despite disagreeing about what the best alternative is, are unwilling or unable to extract any concessions.

The difference between compromise, consensus, and capitulation can be illustrated by a simple example. Suppose that Party A and Party B disagree about whether, and by how much, the minimum wage ought to be raised. Party A believes that the minimum wage should be raised to $12, while Party B believes that the minimum
wage should remain at $8. If, after discussion and debate, both Party A and Party B alter their views and both parties come to believe that the minimum wage should be raised to $10, then the result is a consensus. Suppose, however, that despite significant discussion and debate, the two parties’ views remain unchanged: Party A still believes the minimum wage should be $12, and Party B still believes it should remain at $8. If, at this point, the two parties agree to “split the difference” and raise the minimum wage to $10, then the agreement is a compromise. Note that the parties need not literally split the difference in order for an agreement to count as a compromise. An agreement in which one side concedes more than the other, or in which there is no way of determining a clear midpoint, can still count as a compromise. For example, if Party A and Party B agreed to raise the minimum wage to $9, then the agreement would still be a compromise, even though Party A concedes more than Party B. That said, if the concessions are vastly unequal, then an agreement may be better counted as a capitulation than as a compromise. For example, if Party A and Party B agree to raise the minimum wage to $8.01, then Party A can rightly be said to have capitulated to Party B. It is also case of capitulation if Party A gives up trying to raise the minimum wage because Party B refuses to cooperate, thereby allowing Party B to get their most preferred option.

Compromise is not unique to politics. The need for it can arise in any setting where some end cannot be achieved without the agreement of two or more parties that disagree
about what the best alternative is. A group of friends may compromise about where to meet for lunch. Two firms may compromise on the terms of a business deal. A mother may compromise with her child about the child’s bedtime. Though what I say may extend to these non-political cases, I focus here only on political compromises understood as compromises made by political actors in order to secure political ends.

A dirty compromise is a kind of moral compromise. By “moral compromise” I mean a compromise in which at least one party gives up something it considers morally important. Not all political compromises are moral compromises. Suppose, for example, that a city council is trying to determine how frequently to host public concerts in the city’s central park, and that the population of the city is divided: Half the city loves public concerts and so prefers to have concerts in the park every week. The other half prefers that the park be a place of quiet refuge, and so prefers that concerts take place no more than once a month. The city council might implement a compromise by agreeing to host concerts, say, every two weeks. This would be a political compromise, but not a moral compromise, at least on the plausible assumption that neither side sees the frequency of concerts in the park as a moral issue.

One might even speak of compromising with oneself in cases where one is of two minds about what to do. In this case, we can say that each of one’s minds is a party to the compromise. I set aside the issue of whether we can say this literally or only as metaphor.
In a moral compromise, at least one side (usually all sides) see the issue on which they are compromising as a moral issue. In making concessions, they accept an outcome that falls short of what they believe would be morally best. In this regard, each side sees the compromise as, at the very least, morally suboptimal. Not all moral compromises are dirty compromises because an outcome can be morally suboptimal without being seriously unjust. Many things that we think would be morally best fall under the heading of the supererogatory; things that it would be morally good to do, but that are not morally obligatory. For example, suppose that Party A believes that it would be morally best to eliminate all fees for the use of public recreation facilities such as swimming pools, gyms, and ice rinks to encourage people to live healthy lifestyles and to improve access to those facilities among low income families. Party B opposes this policy because they believe that providing free access to those facilities would require too much tax money. If the two sides agree on a compromise, say, cutting fees to half their current level, then the compromise is a moral compromise because Party A settles for less than they believe would be morally best. However, they need not see this outcome as seriously unjust, at least on the plausible assumption that providing free access to public recreation facilities, though perhaps morally desirable, is not required by justice.

To qualify as a dirty compromise, at least one side must see the outcome as not merely morally suboptimal, but seriously unjust. I use the qualifier “seriously” to exclude
cases of relatively benign compromises between parties who disagree about the precise demands of justice. Suppose, for example, that Party A and Party B disagree about whether progressive income taxes or consumption taxes are better from the standpoint of justice. If the two parties agree to some system that employs a combination of progressive income tax and consumption taxes, then each side may see the resulting tax policy as less than perfectly just, but they need not see it as seriously unjust. In a dirty compromise, at least one side makes a concession that it sees as a serious concession to injustice, that is, as one that makes the outcome of the compromise seriously unjust. For example, the many compromises that abolitionists made with slave holders during the formative years of the United States are dirty compromises. In making concessions that allowed slavery to continue in particular states (as in the Missouri Compromise), or that required that escaped slaves be returned to their masters (as in the Fugitive Slave Act), abolitionists agreed to laws and courses of action that they rightly considered to be seriously unjust.

Dirty compromises are not necessarily instances of what Avishai Margalit calls rotten compromises. As Margalit defines it, a rotten compromise is “an agreement to establish or maintain an inhuman regime, a regime of cruelty and humiliation, that is, a regime that does not treat humans as humans.” He argues that such compromises should

---

7 Margalit, On Compromise, 2.
be “prohibited in all circumstances.”

Margalit’s examples of rotten compromises include the concessions to slave-holders made during the formation of the United States, the Munich agreement permitting Nazi Germany to annex Czechoslovakia, and the Yalta agreement, in which the allies agreed to aid in forcibly returning refugees to Stalin’s regime where they typically faced imprisonment or execution. All rotten compromises are dirty compromises because any compromise that involves aiding the establishment of an inhuman regime is one that involves making serious concessions to injustice. However, not all dirty compromises are rotten because not all serious injustices are instances of “inhuman regimes” as Margalit defines the term, that is, regimes that constitute “extreme manifestations of not treating humans as humans.” For example, consider the “Don’t Ask Don’t Tell” compromise implemented by the Clinton administration which allowed gay and bisexual people to serve in the military on the condition that they did not reveal their sexual orientation. I take this to be a dirty compromise because it discriminated against gay and bisexual individuals in a way that is seriously unjust. However, prohibiting openly gay and bisexual individuals from serving in the military does not rise

---

8 Margalit, On Compromise, 3.
9 Ibid., 11-13, 19-23.
10 Ibid., 54-61.
11 Ibid., 95-108.
12 Ibid., 2.
to the level of maintaining an inhuman regime. So, while Don’t Ask Don’t Tell was a dirty compromise, it was not rotten in Margalit’s sense.

4.2 An Example of a Dirty Compromise

It is useful to have a detailed case to work with when considering various questions about the morality of dirty compromises. The section lays out such a case, Vermont’s 2000 compromise on same-sex marriage.

First, the backdrop to the compromise: In 1999, only two U.S. states, Hawaii and California, provided any form of legal recognition of same-sex relationships. Only about 35% of the U.S. population supported gay marriage, with 62% opposed. In fact, nearly half the U.S. population believed that all homosexual acts ought to be illegal. On December 20, 1999 the Supreme Court of Vermont issued a controversial ruling in the case of three same-sex couples that were suing the state of Vermont for denying them marriage licenses. The court unanimously decided that while same-sex couples did not have a right to marriage per se, they did have a right to many of the protections and benefits traditionally associated with marriage. The court decreed that the Vermont legislature find some way to provide those protections and benefits “within a reasonable period of

13 "Gay and Lesbian Rights.”
time.” Following the court’s ruling, the state spent months engaged in what Governor Howard Dean later called “the least civil public debate in the state in over a century.” For his part, Dean announced that he would veto any legislation that granted same-sex couples the right to marry.

To simplify things, I will treat the debate as if it had two fairly unified sides. One side, the advocates of same-sex marriage, believed that justice called for full marriage equality. They called for a law that would permit same-sex couples to marry just as heterosexual couples could. The other side, the opponents of same-sex marriage, believed that marriage is a sacred institution between a man and a woman, and that it would be deeply wrong to allow same-sex couples to marry. Many on this side called for a constitutional amendment to circumvent the court’s decision and prevent any legal recognition of same-sex couples.

With neither side able to secure enough support to have their way, Vermont’s House Judiciary Committee proposed a compromise: Same-sex couples would be allowed

---

16 In doing so, I exclude, for example, those who advocated eliminating state involvement in marriage all together, or who believed from the start allowing same-sex couples to form civil unions was the morally best option.
17 Halloran, ”’How Vermont's 'Civil' War Fueled the Gay Marriage Movement.”
to form civil unions that would afford them the legal protections and benefits afforded to married couples. Marriage, however, would remain reserved for heterosexual couples only; the proposed bill reaffirmed the “traditional” definition of marriage stating, “Marriage’ means the legally recognized union of one man and one woman.”

The compromise faced harsh criticism from both sides. Opponents of gay-marriage warned that allowing same-sex couples to enter civil unions would destabilize traditional families and send the state down a path to moral oblivion. Meanwhile, advocates of same-sex marriage, such as Representative Steven Hingtgen, were appalled by the concessions to “traditional” views of marriage. Making such concessions, Hingtgen said, “validates the bigotry. It does more than validate it. It institutionalizes the bigotry and affirmatively creates an apartheid system of family recognition in Vermont.”

4.3 Three Problems for Dirty Compromises

The Vermont compromise was a dirty compromise. By reaffirming the “traditional” definition of marriage, the resulting law unjustly discriminated against same-sex couples, marking them out as having a lesser status because of their sexual

---

19 Halloran, "How Vermont’s ‘Civil’ War Fueled the Gay Marriage Movement.”
20 Goldberg, "Vermont Panel Shies from Gay Marriage.”
orientation. While this is not as grievous a concession to injustice as some of the more extreme examples of dirty compromise, such as the compromises abolitionists made with slave holders, it is nonetheless serious enough to qualify the Vermont compromise as a dirty compromise. I now turn to three moral questions about dirty compromises: First, are they ever morally acceptable? Second, do advocates of justice sacrifice their integrity when they agree to dirty compromises? Third, does agreeing to a dirty compromise necessarily leave one with so-called dirty hands? This section develops each of these questions in greater detail; I hold off providing my own answers until Section 4.4.

4.3.1 Are dirty compromises ever morally acceptable?

If it can be shown that the Vermont compromise was morally acceptable, then it can be shown that at least some dirty compromises are morally acceptable. But was the Vermont compromise morally acceptable? In seeking an answer this question, one confronts two seemingly inconsistent lines of thought. The first focuses on fundamental values and moral principles. The second focuses on pragmatic considerations and the range of realistic alternatives. I will call the former the principles approach and the latter the pragmatic approach.

The principles approach determines whether a compromise is morally acceptable by asking whether it involves violating a fundamental moral principle. If it does, then the
compromise is morally unacceptable. Hingtgen’s remarks about the Vermont compromise suggest a version of this approach. In saying that the compromise would create “an apartheid system of family recognition in Vermont”, Hingtgen emphasized the way in which the compromise violated fundamental moral principles that forbid invidious discrimination.

Among philosophers, Martin Benjamin expresses a version of the principles approach. In his view, our judgment of a compromise should be guided by three values: individual integrity, a (roughly utilitarian) concern for social welfare, and a (roughly Kantian) concern for equal respect. Benjamin grants that these values underdetermine the range of acceptable compromise, but he contends that they rule out at least some compromises as morally unacceptable from the start. He writes:

There is no ethical reason to accommodate conduct that flagrantly and seriously violates considerations of utility or equal respect or to accommodate the world view and way of life in which it is rooted. Moral compromise with Nazis, racists, sexISTS, and rapists should, for example, be rejected out of hand, even if in some circumstances prudential considerations lead one to “compromise”—or, more accurately, to capitulate—in the face of their greater power.21

Here Benjamin suggests that when faced with views that seriously undermine integrity, social welfare, or equal respect, we should refrain from making concessions unless we find ourselves in the dire situation of simply having to surrender. At that point, the

agreement is not a compromise—it is not an agreement in which each side gives up something—but a capitulation, in which one side concedes to the other’s overwhelming power. Depending on how demanding the minimum standards of social welfare and equal respect are, Benjamin’s view may imply that most, or even all, dirty compromises are morally unacceptable. (And, as we will see in the next section, Benjamin also precludes the possibility of making dirty compromises without sacrificing one’s integrity.)

In a similar vein, Simon May suggests that a compromise can be morally acceptable so long as “it violates no fundamental moral principle.”22 He writes:

Before the first democratic elections in South Africa in 1994, a compromise between major parties led to the inclusion of the Government of National Unity provision in the interim constitution: all parties registering more than 10 percent of the vote were to be represented in the cabinet. This compromise was morally acceptable since coalition government violates no fundamental democratic principle. In contrast, the U.S. slavery compromises of 1820 and 1850 were morally unacceptable since they permitted the practice of slavery in Missouri and the New Mexico territory seized in 1848. Perhaps there was no way for abolitionists to eliminate slavery before the Civil War, but that would make the compromises unavoidable, not morally acceptable. 23

Assuming that principles of justice count as fundamental moral principles, May’s criterion implies that dirty compromises are never morally acceptable.

22 May, "Principled Compromise and the Abortion Controversy," 322.
23 Ibid., 322 note 10.
Advocates of the pragmatic approach see the principles approach as counter-productive. Rather than judge a compromise by whether it satisfies some set of fundamental moral principles, the pragmatic approach asks whether the compromise was the best that could be done under the circumstances. This line of thought is exemplified by the remarks of Thomas Little, one of the architects of the Vermont Compromise:

Leadership requires a keen sense of what ought to be done in the context of what can be done, what is achievable. Leadership untempered by careful assessment of the world we live in is not sound leadership. What is achievable in this General Assembly and this body politic this year is a broad civil rights bill – and speaking for myself, that does not cross the threshold of marriage.24

From this perspective, while the Vermont compromise was not everything that gay-marriage advocates hoped for, it was the best they could realistically achieve under the circumstances. Though the compromise fell short of affording same-sex couples equal status, it was a significant advancement over the status quo, an advancement that was only possible because the law made concessions to those who opposed gay marriage. William Lippert, the only openly gay member of the Vermont legislature, defended the compromise as “the single most forward-looking piece of legislation for gay and lesbian people in the entire United States.”25

24 Goldberg, "Vermont Panel Shies from Gay Marriage."
25 Ibid.
The pragmatic approach is forcefully defended by Amy Gutmann and Dennis Thompson. They argue that we should reject any attempt to draw principled limits on what constitutes a morally acceptable compromise. One reason they give for this stance is that “any unconditional standard (at least, any politically relevant one) will block some decent compromises that improve on the status quo.” They insist that rather than trying to set limits on the range of acceptable compromises, we ought to focus on whether a compromise is the best way of improving the status quo given the range of alternatives that are realistically achievable at the time. They write:

Achieving the best possible outcome will depend in no small measure on the nature of the negotiations and the evolving political context. Before the fact, driving a principled stake in the ground and tenaciously refusing to move—if more than a negotiating tactic—is a prescription for thwarting progress that could be mutually beneficial. Even after the fact, judging compromises by a pre-ordained set of principles is not productive. Compromises are too easy to criticize simply because the results everyone observes—which are often morally incoherent—are divorced from both the process and the alternatives that presented themselves at the time.

From this perspective, dirty compromises may be morally acceptable so long as they produce the best outcomes possible under the circumstances.

Whether dirty compromises are ever morally acceptable seems to depend on which approach to evaluating compromises—the principles approach or the pragmatic

---

26 Gutmann and Thompson, *The Spirit of Compromise*, 79.
27 Ibid., 80.
approach—is the correct one. The principles approach suggests that dirty compromises are (almost) always morally unacceptable because they violate fundamental principles, whereas the pragmatic approach suggests that dirty compromises are sometimes morally acceptable because they are the best that can be done under the circumstances. How are we to decide between these two different approaches? I do not think we have to. Though they might seem mutually exclusive, I do not believe that the principles approach and the pragmatic approach are necessarily incompatible. In section 4.4, I propose a means of partially reconciling them by suggesting that they are actually evaluating different things. But first, I want to put two other moral worries about dirty compromises on the table.

4.3.2 Do dirty compromises undermine integrity?

Despite calling the Vermont civil unions compromise a validation of bigotry and institutionalized discrimination, Steve Hingtgen voted for the bill to pass into law.\(^{28}\) This might seem hypocritical; how could Hingtgen condemn the compromise as seriously unjust and then turn around and vote for it? In voting for the bill, did Hingtgen sacrifice his integrity? More generally, is it possible to make compromises that violate our deeply

held principles of justice without sacrificing our integrity? The answer depends in part on what we mean by “integrity”. For this, I turn to the account of integrity developed by Martin Benjamin.

Benjamin observes—correctly it seems to me—that our everyday talk of integrity is “centered on lapses or deficiencies of integrity.” When we talk about integrity, it is usually (though not always) in the course of pointing out that some person (or thing) has failed to demonstrate enough of it. Even when we do talk of individuals as possessing great integrity, it is often with reference to some temptation or circumstances that might have lead them to act in a less praiseworthy way. Taking this feature of our language as a starting point, Benjamin develops an account of integrity by considering what goes wrong in characteristic cases of lapses or deficiencies in integrity such as hypocrisy, opportunism, flip flopping, and weakness of will. Benjamin suggests that what goes wrong in such cases is one of two things: the person is either revealed to be unscrupulous in the sense of simply not having any moral principles, or, though the person has principles, their words and actions fail to express or embody them. Benjamin concludes from this that integrity requires:

...an integrated triad consisting of: (1) a reasonably coherent and relatively stable set of highly cherished values and principles; (2) verbal behavior expressing these

29 Benjamin, Splitting the Difference, 47.
values and principles; and (3) conduct embodying one’s values and principles consistent with what one says.\textsuperscript{30}

There is room to quibble with this account. One might object, for example, to the fact that radically evil persons could satisfy this definition of integrity. A person who cherishes causing human suffering, says that he cherishes it, and sets about causing it, would be a person of integrity on this account. But many people are reluctant to ascribe integrity to such malevolent individuals. This objection could be met by adding a fourth condition requiring that the cherished values and principles required by (1) be defensible values and principles.\textsuperscript{31} I am ambivalent about the need for such an additional condition, but since such a condition makes no difference to the arguments that follow, I omit such a condition here and adopt Benjamin’s account as stated.

An apparent tension between compromise and integrity arises in cases where compromise involves concessions to the moral views of others. In making concessions to the moral views of others we give up achieving, at least for the present, something we believe is morally valuable. Backing such compromises thus requires us to support policies or courses of action that do not fully satisfy our own values and principles. The problem is even more acute in the case of dirty compromises because in such cases the

\textsuperscript{30} Benjamin, \textit{Splitting the Difference}, 51.
outcome we accept is, in light of our own principles, not just morally suboptimal, but seriously unjust. Advocates of same-sex marriage in Vermont like Hingtgen faced exactly this sort of situation. They were asked to help pass a law that they saw as seriously unjust because it wrongly discriminated against same-sex couples. Given their deep commitment to equality for same-sex couples, was it possible for them to support the law without sacrificing their integrity?

According to Benjamin’s own proposed reconciliation of compromise and integrity, the answer is “no”. Following Isaiah Berlin, Benjamin contends that there are many incompatible values, and that there are many equally defensible ways of prioritizing and trading off some of those values over others. Because of this, people of integrity can recognize that their own values and principles are not the only legitimate ones. It is this recognition that makes integrity-preserving moral compromises possible on Benjamin’s account:

What makes compromise possible is that we are... capable of adopting a more external perspective, one that allows us to acknowledge the contingency of our world views and ways of life and the equal legitimacy of others with which ours is, on at least some occasions, bound to conflict.32

In compromising with moral views that are just as legitimate as our own we still give up achieving some of what we believe would be morally best, but because the views we make

32 Benjamin, Splitting the Difference, 101.
concessions to are just as morally legitimate as our own, we can still recognize the end result as morally legitimate. Because we recognize both the views we make concessions to and the end result of those concessions as morally legitimate, the compromise does not threaten our integrity.

Theo van Willigenburg defends a similar reconciliation of compromise and integrity based on value pluralism and the indeterminacy of value rankings. On his view, integrity-preserving compromise is possible because value rankings are either ontically indeterminate—“there is no uniquely right answer to the question as to which value should take priority”—or epistemically indeterminate—“we lack sufficient epistemic abilities to determine what value should take priority.”33 In cases of indeterminacy, whether ontic or epistemic, “parties may be asked to acknowledge that if there is no sufficient argumentative grounding for their position, and the opposing view is supported by plausible arguments endorsed by thoughtful and intelligent people, a moral compromise might be appropriate.”34 Like Benjamin, van Willigenburg grounds integrity-preserving compromise in our ability to recognize the views of others as morally legitimate or defensible.

34 Ibid., 400.
On both Benjamin and van Willigenburg’s accounts, the Vermont compromise is not an example of integrity-preserving compromise. This is because advocates of same-sex marriage (rightly) did not see the views of their opponents as equally morally legitimate. Quite the opposite; many saw their opponents as, in Hingtgen’s words, “preaching hate in all its poor disguises.” In making concessions to their opponents, advocates of same-sex marriage like Hingtgen conceded to views that they believed to be morally illegitimate. If integrity-preserving compromise requires recognizing your opponents’ views as legitimate or morally defensible, then the Vermont compromise cost advocates of same-sex marriage their integrity. This, I will argue, is the wrong result, but first, I want to lay out one more moral worry about dirty compromises.

4.3.3 Do dirty compromises produce dirty hands?

It is often said that, as Michael Walzer puts it, “No one succeeds in politics without getting his hands dirty.” The basic idea is that to be an effective political actor one must often engage in activities that are morally dubious. Suggesting some examples, Bernard Williams writes:

35 Goldberg, "Vermont Panel Shies from Gay Marriage."
...a politician might find himself involved in, or invited to, such things as lying, or at least concealment and the making of misleading statements; breaking promises; special pleading; temporary coalition with the distasteful; sacrifice of the interests of worthy persons to those of unworthy persons; and (at least in a sufficiently important position) coercion up to blackmail.37

Similarly, Neil Levy writes:

...dirty actions are part and parcel of ordinary political life. Politicians must make deals, compromise with interests they abhor, distribute favors and neglect relationships, if they are to pursue the power to do good. If politics is not inevitably dirty, it is always somewhat grubby.38

This need to engage in morally dubious deeds is often said to leave politicians with dirty hands.

There are different ways of understanding exactly what it means to say that someone has “dirty hands,” but Michael Walzer’s interpretation has been by far the most influential.39 On his view, cases of dirty hands are cases in which “we know we have done something wrong, even if it was the best thing to do on the whole in the circumstances.”40

Drawing on the work of Walzer and others, Stephen de Wijze provides a general characterization of dirty hands cases. He writes:

40 Walzer, “Political Action,” 171.
…in all cases of dirty hands what is common is that actions involve justified betrayal of persons, values or principles due to the immoral circumstances created by other persons (or organizations of persons) within which an agent finds herself. In such circumstances, a good person finds it impossible not to be moved by moral considerations to commit moral violations.... the agent is forced to cooperate with evil, so furthering its immoral projects.41

Now we can ask, are dirty compromises necessarily cases of dirty hands? That is, even if making a dirty compromise is the all-things-considered right thing to do, does one still act wrongly by making concessions to injustice and thereby dirty one’s hands? Suppose that in the Vermont case agreeing to the compromise was, all things considered, the right thing for advocates of same-sex marriage to do. Did they nonetheless commit some moral wrong in supporting a bill that reaffirmed the “traditional” definition of marriage thereby relegating same-sex couples to an inferior status? Did advocates of same-sex marriage dirty their hands, even if they did the right thing?

This line of questioning is likely to strike many philosophers as simply incoherent. Walzer himself recognizes the seeming paradox. He writes, “One’s hands get dirty from doing what it is wrong to do. And how can it be wrong to do what is right? Or, how can we get our hands dirty doing what we ought to do?”42 Especially to philosophers with utilitarian sympathies, Walzer says, the problem of dirty hands may seem to be “a pile of

42 Walzer, “Political Action,” 164.
confusion upon confusion.” 43 Indeed, Kai Nielson argues exactly this; that the problem of dirty hands involves “a conceptual confusion with unfortunate moral residuals.” 44 He insists that, “one cannot logically do what is right by doing wrong.” 45

For present purposes, I set aside doubts about the logical and conceptual coherence of the problem of dirty hands. Though I aim to show that dirty compromises need not produce dirty hands, I want to do so in a way that does not require rejecting the very possibility of cases of dirty hands. Rather, I aim to show that even if there are genuine cases of dirty hands of the sort that Walzer and de Wijze describe, dirty compromises need not be among them. So, I grant, at least for the sake of argument, that cases of dirty hands are possible, that is, that there are cases in which someone commits a moral wrong, even though they do the right thing. I aim to show that one who agrees to a dirty compromise need not commit any moral wrong in doing so.

4.4 How to Evaluate a Compromise

I have now set out three questions about dirty compromises: First, are dirty compromises ever morally acceptable? Second, do advocates of justice sacrifice their

43 Walzer, “Political Action,” 162.
integrity when they agree to dirty compromises? Third, does agreeing to a dirty compromise necessarily leave one with so-called *dirty hands*?

Let us begin with the first question by returning to the seemingly inconsistent approaches—the principles approach and the pragmatic approach—described in section 4.3.1. Whereas the principles approach suggests that dirty compromises are morally unacceptable because they violate fundamental moral principles, the pragmatic approach suggests that dirty compromises may be morally acceptable given the range of realistically available alternatives. Which of these offers the right approach to evaluating compromises? My answer is “Both, at least to some extent.” Though these two approaches might seem incompatible, I think that they can be reconciled, at least to some extent, if we recognize that they are each suited to a different target of evaluation.

The first step of this reconciliation is to distinguish two questions that I think are often conflated when we ask whether a compromise is morally acceptable. The first question is about the content of the compromise itself—the law, policy, or course of action that the parties agree to; it asks whether this law, policy, or course of action is morally acceptable. The second question is about the act of agreeing to a compromise; it asks whether it was morally acceptable for a particular party to agree to a particular compromise. For example, when we ask whether the Vermont compromise was morally acceptable, we may be asking:
(I) Was the content of the Vermont compromise—the civil unions law—morally acceptable?

Or, we may be asking:

(II) Was it morally acceptable for advocates of same-sex marriage to agree to the civil unions law?  

It is tempting to think that our answer to (I) and (II) must be the same—that if the law was morally unacceptable, then it must have been morally unacceptable to agree to it. This kind of reasoning is common in political discourse where one often hears arguments of the following form:

1. Politician \( P \) voted for law \( L \).

2. Law \( L \) is morally unacceptable.

3. Therefore, Politician \( P \) did something morally unacceptable.

This is fallacious. The fact that the content of a compromise—a law, policy, or course of action—is morally unacceptable does not entail that it is morally unacceptable for anyone to agree to the compromise. It can be morally acceptable to agree to a morally unacceptable compromise, that is, to a compromise that has morally unacceptable content.

\[46\] Of course, we can also ask whether it was morally acceptable for opponents of same-sex marriage to agree to the law. However, here I will focus on the issue from the point of view of advocates of same-sex marriage.
To put the point a different way, the answers to questions (I) and (II) are logically independent. We can say that the content of the Vermont compromise—the civil unions law—was morally unacceptable, but that it was nonetheless morally acceptable for advocates of same-sex marriage help pass it. In fact, I think this is the right thing to say.

What enables the answers to the two questions to come apart? How can we coherently answer “no” to (I) and “yes” to (II)? The answer, I think, is that whereas the principles approach is the appropriate approach for answering questions like (I), the pragmatic approach is the appropriate approach for answering questions like (II). In other words, when judging whether the content of a compromise is morally acceptable we should ask whether it violates fundamental moral principles. But when judging whether it was (or is) morally acceptable for a party to agree to a compromise, we should ask whether agreeing to the compromise was the right thing to do given the range of realistic alternatives. So, I think both the principles approach and the pragmatic approach each get something right, but only once we recognize that they properly apply to different things: the content of a compromise on one hand, and the act of agreeing to a compromise on the other. Let us look more carefully at the proper domain of each approach.
4.4.1 Evaluating the Content of a Compromise

The content of a compromise is the law, policy, or course of action that the parties agree to. In the case of the Vermont compromise, the content was the civil unions law that afforded same-sex couples the ability to form civil unions while at the same time reaffirming the “traditional” definition of marriage as a union between one man and one woman. I use the word “content” rather than “outcome” here because there are many things that might rightly be said to be outcomes of a compromise that are not part of the content that the parties agree to. For example, in hindsight, many now regard the Vermont compromise as a watershed moment in the fight for LGBT rights. It served as a model for other states to follow and helped to normalize same-sex partnerships in both the law and public culture.\(^\text{47}\) However, whether a compromise produced good consequences in the long-run is a different question from whether the content of the compromise was itself morally acceptable.\(^\text{48}\) A morally unacceptable law may produce good consequences in the

\(^{47}\) Halloran, "How Vermont's 'Civil' War Fueled the Gay Marriage Movement."

\(^{48}\) It might seem that a consequentialist cannot accept this distinction, but I do not think this is true. A consequentialist may judge that a law is morally unacceptable because it violates rules or principles that are justified on consequentialist grounds. For example, a principle forbidding discrimination on the basis of sexual-orientation grounds can be justified on consequentialist grounds in light of the harm that such discrimination causes. Accepting this principle on consequentialist grounds is compatible with also judging that a law which violates this principle is likely to produce the best consequences in a particular political context.
long-run. Indeed, I contend that the civil unions law was morally unacceptable even though it produced good consequences in the long-run.

By what standard are we to evaluate the content of a compromise if not by its long-run effects? I submit that the answer is basic standards of justice. The content of a compromise, as with any law, policy, institution, or regime, is morally unacceptable if it violates basic standards of justice. By “basic standards of justice” I mean standards such as fundamental moral and political rights, the requirements of political legitimacy, the rule of law, moral and political equality, and principles of distributive justice. In my view, the Vermont compromise violated basic standards of justice because it invidiously discriminated against individuals based on their sexual orientation. Thus, the content of the Vermont compromise—the civils unions law—was seriously unjust, and therefore morally unacceptable.

My approach to evaluating the content of compromises is a version of the principles approach. It determines whether the content of a compromise is morally acceptable by asking whether it violates basic standards of justice. But what about the criticisms of the principles approach raised by proponents of the pragmatic approach such as Gutmann and Thompson? Let us look at their objections more carefully. Gutmann and

49 In this regard, the fact that a law is the content of a compromise is irrelevant to a moral evaluation of it.
Thompson insist that, “even after the fact, judging compromises by a pre-ordained set of principles is not productive.”\textsuperscript{50} They argue that “to judge compromises as acceptable only if they are consistent with a principle or with some coherent set of principles is to put the compromise—but not the status quo—on the defensive.”\textsuperscript{51} Such an approach, they say, wrongly treats principles as “a series of roadblocks to all those compromises that would move society in the right direction, however imperfectly or partially.”\textsuperscript{52}

Here I think that Gutmann and Thompson fail to clearly distinguish the question of whether the content of a compromise is morally acceptable from the question of whether it is morally acceptable to agree to a compromise.\textsuperscript{53} Once we separate these questions, we can see that Gutmann and Thompson’s objection applies only to the use of the principles approach to answer the second question, not the first. Evaluating the content of a compromise by asking if it is consistent with basic principles of justice need not be an obstacle to making compromises that improve on the status quo, at least not if we recognize that it is sometimes morally acceptable to agree to compromises that have morally unacceptable content. The principles approach is only an obstacle to making such

\textsuperscript{50} Gutmann and Thompson, \textit{The Spirit of Compromise}, 80.
\textsuperscript{51} Ibid., 81.
\textsuperscript{52} Ibid., 84.
\textsuperscript{53} Elsewhere Gutmann and Thompson distinguish the content, process, and domain of compromise, but they do not make this distinction when criticizing the attempt to find principled limits on the moral acceptability of compromise. For the distinction, see ibid., 41.
compromises if we use it to determine whether it is morally acceptable to agree to a compromise. Thus, my proposal—that we use the principles approach to evaluate the content of compromises, and the pragmatic approach to evaluate the act of agreeing to a compromise—avoids Gutmann and Thompson’s objection.

Still, one might wonder whether there is any good reason to favor the principles approach when evaluating the content of compromises. Why should we evaluate the content of compromises in a way that is, as Gutmann and Thompson say, “divorced from the process and the alternatives that presented themselves at the time?”

One answer is that, contra Gutmann and Thompson, doing so is often productive. Recognizing that the content of a compromise is morally unacceptable provides impetus for further moral progress. For example, in the case of the Vermont compromise, recognizing that the civil unions law was unjust drove people to carry on campaigning for marriage equality. They would have been less motivated to do this had they judged that the law was morally acceptable as it was.

Another answer is that whether a law or policy is morally acceptable simply does not depend on the range of realistically available alternatives. A law or policy can be unjust even if the distribution of power and beliefs in society make it practically

54 Gutmann and Thompson, The Spirit of Compromise, 80.
impossible to change it. For example, eliminating slavery within six months was not an alternative that was realistically available to abolitionists in 1800, but this fact does not render laws that permitted slavery morally acceptable. Those laws were morally unacceptable because they violated basic standards of justice. The fact that it was, at the time, practically impossible for anyone to do away with such laws is irrelevant. Similarly, just because it was not politically possible in Vermont in 2000 to pass a law granting same-sex couples the right to marry does not imply that the law that was passed, the civil unions law, was morally acceptable. This is so even if the civil unions law was the least unjust among all the alternatives that could have secured enough support to become law. A law or policy is unjust so long as it violates basic standards of justice even if no better alternatives are realistically available at the time.

One might object that my approach is utopian (in the pejorative sense) because it holds laws and policies to unrealistically high standards; it demands that they meet basic standards of justice even when meeting those standards is not a realistic option under the circumstances. To answer this objection, it is necessary to distinguish between realistically available alternatives and realistic ideals and standards. A realistically available alternative is one that can realistically be achieved at a particular point in time given the existing distribution of political power, public attitudes, institutional barriers, and other features of the immediate political context. In the Vermont case, passing legislation that would
legalize same-sex marriage was not a realistically available alternative. There was, practically speaking, no way of securing enough votes to get such a law through both chambers of the Vermont Legislature, and even if such a law could have secured enough votes, the governor had publically announced that he would veto it. However, this does not imply that legalizing same-sex marriage in Vermont was an unrealistic ideal for anyone to be working towards at the time.

More generally, a political ideal or standard can satisfy appropriate constraints of realism even if realizing the ideal or meeting the standard is not a realistically available alternative. Consider, for example, Allen Buchanan’s suggestion that a realistic ideal theory of justice should be feasible, accessible, and morally accessible. A theory is feasible “if and only if the effective implementation of its principles is compatible with human psychology, human capacities generally, the laws of nature, and the natural resources available to human beings.”55 A theory is accessible if and only if it is feasible and “there is a practical route from where we are now to at least a reasonable approximation of the state of affairs that satisfies its principles.”56 Finally, a theory is morally accessible if and only if it is possible to get from where we are now to the ideal state identified by the theory.

56 Ibid.
“without unacceptable moral costs.”\textsuperscript{57} What is important for my purposes is that an ideal or standard of justice can meet all three of these conditions even if there is no hope of satisfying it in the near future. Indeed, Buchanan allows that achieving the ideal may be possible “only after a laborious and extended process of change.”\textsuperscript{58}

Now consider a standard of justice that forbids institutionalized discrimination based on sexual orientation. Is this an unrealistic standard of justice to apply to Vermont in 2000? No, at least not by the plausible constraints of realism that Buchanan proposes. The standard is feasible because it is possible for humans to meet it without transcending the limits of human psychology, the laws of nature, or the supply of natural resources. The standard is accessible because there was a practical path from where Vermont was in 2000 to a state of affairs where the standard was approximately met. And the standard is morally accessible because this practical path did not involve any unacceptable moral costs. Indeed, it took just nine years for Vermont to move from allowing only same-sex civil unions to allowing same-sex marriage, and this progress was achieved largely by morally permissible activism and legislative means. So, judging the Vermont compromise by a standard of justice that forbids institutionalized discrimination based on sexual

\textsuperscript{57} Buchanan, \textit{Justice, Legitimacy, and Self-Determination}.
\textsuperscript{58} Ibid.
orientation is not utopian even though, under the circumstances, passing a law that met the standard was not a realistically available alternative.

4.4.2 Evaluating Compromises as Actions

So far, I have argued that the principles approach is the correct approach for evaluating the content of a compromise. But is it also the correct approach for evaluating the act of agreeing to a compromise? That is, should we determine whether it is morally acceptable to agree to a compromise by asking whether the compromise is consistent with basic standards of justice? If so, then a negative answer to the question “Is the content of the compromise morally acceptable?” would always imply a negative answer to the question “Is it morally acceptable to agree to the compromise?” In the Vermont case, this would lead to the conclusion that it was morally unacceptable for anyone to agree to the compromise.

Here it is useful to return to Gutmann and Thompson’s critique of the principles approach. I argued above that their critique fails to show that the principles approach is

---

59 This is a different question from what Martin Golding calls the question of “morality in compromise”, that is, the question of whether one goes about negotiating a compromise in a morally acceptable way. See Martin Golding, "The Nature of Compromise: A Preliminary Inquiry," in Compromise in Ethics, Law, and Politics: Nomos Xxi, ed. J. Ronald Pennock and John Chapman (New York: New York University Press, 1979), 5-6.
the wrong approach to evaluating whether the content of a compromise is morally acceptable. However, I think their critique does show that the principles approach is a misguided way of evaluating whether agreeing to a compromise is morally acceptable. The reason is that, used in this way, the principles approach wrongly denigrates agreeing to compromises that are the best available means of improving on the status quo.

Of course, just because a compromise would improve on the status quo does not necessarily imply that it is morally acceptable to agree to the compromise. There may be overriding moral reasons to reject a compromise even if it would improve on the status quo. For example, the short-term improvement over the status quo may be offset by longer-term costs of agreeing to the compromise if the compromise makes further moral improvement more difficult. Or, it may be morally unacceptable to agree to a compromise simply because one has promised not to agree to it. However, in at least some cases, the fact that a compromise is the best available means of improving on the status quo should morally justify agreeing to it, even if the content of the compromise itself is morally unacceptable. In short, it is sometimes morally acceptable to agree to a morally unacceptable compromise.

60 I briefly discuss cases of this sort in section 4.5.

61 It is worth noting that saying one is morally justified is different from saying that one is morally excused. If one is morally justified, then what one does is not morally wrong. However, if one is morally excused, then one still acts wrongly; one is only excused from blame.
I suggest that the reason that the moral acceptability of agreeing to a compromise does not neatly track the moral acceptability of the content of the compromise is that the act of agreeing to a compromise can only appropriately be judged against the range of realistically available alternatives, whereas the content of a compromise can appropriately be judged against basic standards of justice. When the range of realistically available alternatives does not include any alternatives that satisfy basic standards of justice, accepting a compromise that violates those standards may be the morally best thing to do, at least if the state of affairs created by the compromise is likely to be morally better than the status quo. In other words, while the principles approach is the appropriate way to evaluate the content of a compromise, when it comes to evaluating whether it is morally acceptable to agree to a compromise, we ought to adopt the pragmatic approach.

What exactly does it mean to adopt the pragmatic approach? It means that when we are evaluating whether it is morally acceptable to agree to a compromise, we should ask whether agreeing to the compromise is likely to result in an outcome that better approximates justice than any of the other realistically available alternatives. Of course, it is often difficult to determine which alternatives are realistically available or what the consequences of adopting a particular alternative will be. Even looking back at historical compromises, it is often difficult to evaluate the relevant counterfactuals. And without the benefit of hindsight, things are typically even harder. The best we can do is exercise our
judgment, taking full account of the facts available to us. Moreover, even if it is unclear exactly which alternatives are realistically available, it will often be clear that certain alternatives are not realistically available.

Returning to the Vermont case, it is unclear exactly what would have happened had advocates of same-sex marriage in the Vermont legislature rejected the civil-unions compromise. However, it is fairly clear that a bill legalizing same-sex marriage was not going to pass through both chambers of the Vermont legislature, and even if it did the governor had publically announced that he would veto it. For advocates of same-sex marriage, agreeing to the compromise ensured a significant advancement over the status quo, whereas rejecting it may not have secured any gains whatsoever. So, it seems to me, advocates of same-sex marriage had a sound moral justification for agreeing to the compromise even though the content of the compromise itself violated basic standards of justice. Agreeing to the compromise was the best realistically available means of improving on the status quo for same-sex couples.

So, in answer to the question “Are dirty compromises ever morally acceptable?” my answer is “yes and no”. Since, by definition, the content of a dirty compromise violates basic standards of justice, the content of a dirty compromise is never morally acceptable. Any law, policy, or course of action that constitutes the content of a dirty compromise can rightly be condemned as morally unacceptable. However, this does not entail that it is
never morally acceptable to agree to a dirty compromise. Quite the opposite, sometimes agreeing to a dirty compromise is the right thing to do because agreeing to the compromise is the best means that is realistically available to improve on the status quo. Thus, dirty compromises can be morally acceptable in the sense that it can be morally acceptable to agree to them.

4.4.3 Compromising with Integrity

I have argued that proponents of same-sex marriage did the right thing in agreeing to the Vermont compromise. Now I return to the question I raised in Section 4.3.2: Did same-sex marriage advocates like Hingtgen sacrifice their integrity by agreeing to a compromise that re-affirmed the “traditional” definition of marriage? Recall that, according to the accounts of integrity-preserving compromise developed by Benjamin and van Willigenberg, the answer is “yes, they sacrificed their integrity.” This is because, on those accounts, integrity-preserving compromise is made possible by recognizing the equal moral legitimacy of the opposing view, and same-sex marriage advocates did not recognize their opponent’s views as equally morally legitimate.

I believe that Benjamin and van Willigenberg err in requiring that the parties to a compromise recognize their opponents views as equally morally legitimate or defensible. I contend that one need not recognize an opponent’s views as morally legitimate in order
to compromise with that opponent in a way that preserves one’s own integrity. All that is required, on my view, is that one has a moral justification for compromising that is rooted in one’s own principles and values together with the relevant facts of the situation. So, while recognizing an opponent’s views as legitimate may be sufficient to ground integrity-preserving compromise, it is not necessary.

It is helpful here to distinguish two different kinds of justification: objective and subjective. The act of agreeing to a compromise is objectively justified if there are moral reasons that are sufficient to justify agreeing to the compromise; that is, if agreeing to the compromise is in fact morally permissible, or even morally required. The act of agreeing to a compromise is subjectively justified if, for a given agent, that agent’s own principles and values together with relevant non-moral facts provide that agent with a sufficient moral justification for agreeing to the compromise.

When we ask whether agreeing to a compromise was morally acceptable, we are asking whether agreeing to the compromise was objectively morally justified. When we ask whether agreeing to a compromise cost an agent their integrity, I think we need to ask whether the agent was subjectively justified in agreeing to the compromise. If the agent was subjectively justified, then the agent had sufficient reasons to agree to the compromise that were rooted in her own principles and values. And acting in accordance with what one’s own principles and values give one reason to do cannot be a failure of integrity.
To determine whether advocates of same-sex marriage sacrificed their integrity by agreeing to the Vermont compromise we need not ask whether they saw the views they made concessions to as morally legitimate. Rather, we should ask whether they had a sufficient moral justification for accepting the compromise that was rooted in their own principles and values. Without being inside their heads, it is difficult to say for certain who did and did not have such a justification, but it is easy to see how an advocate of same-sex marriage could have had such a justification.

To see this, put yourself in the shoes of a legislator in Vermont in 2000. Suppose that you believe that same-sex couples ought to have the same rights and benefits as heterosexual couples. However, you also know that in the current political context a bill legalizing same-sex marriage stands no chance of becoming law. You are offered the chance to accept a compromise—civil unions for same-sex couples, but marriage remains for heterosexual couples only. Supposing that you genuinely care about improving things for same-sex couples, what do your principles and values give you most reason to do? I think that if you know, or at least justifiably believe, that that the outcome of agreeing to the compromise is likely to be better for same-sex couples than the outcome of refusing the compromise, then your own principles and values—your own concern for the rights and wellbeing of same-sex couples—give you sufficient moral reason to agree to the compromise. And since agreeing to the compromise is what your own principles and
values give you sufficient reason to do, doing it cannot be a failure of integrity. In agreeing to the compromise, you are acting in accordance with your own principles and values exactly as Benjamin’s account of integrity requires.

Of course, it is lamentable that advocates of same-sex marriage could not achieve full-marriage equality in Vermont in 2000. However, their inability to achieve it is not a fault in them. Given the circumstances, they did the best they could to realize their values and principles. They helped to pave a path towards a more just future. Acting with integrity cannot require more than that. In fact, insofar as backing the compromise is what same-sex marriage advocates’ own principles and values gave them most reason to do, it would actually have been a failure of integrity for them to refuse to compromise. To do so would have been to act against their own best reasons, grounded in their own principles and values. So, contrary to Benjamin and van Willigenberg’s accounts, compromising with views you rightly regard as illegitimate need not threaten one’s integrity, and can even be required to preserve it.

Of course, anyone who acts as Hingtgen did—condemning the compromise as unjust and then voting for it—risks putting themselves in a difficult rhetorical and political position. Popular political discourse is often insensitive to nuance and important distinctions. As I noted above, such discourse often leaps from “The law is unjust and she voted for it,” to “She did something wrong!” A politician who says, “This compromise is
morally unacceptable,“ and then turns around and votes for the compromise is likely to be branded as a hypocrite or a flip-flopper, even though no genuine hypocrisy or flip-flopping is involved. Moreover, saying “This compromise is morally unacceptable but we ought to agree to it“ is seldom a good way of firing up one’s political base. It is much easier to either condemn a compromise and refuse to accept it, or to avoid criticizing a compromise that one intends to accept. However, these rhetorical and practical tensions do not imply that there is any moral or logical tension involved in judging that the content of a compromise offends one’s own principles, but that one’s own principles also give one sufficient reason to accept the compromise; it is just that the moral and logical nuance of that judgment is likely to be lost in the clamor of political discourse. One can accept dirty compromises without sacrificing one’s integrity; though one may risk sacrificing one’s political reputation in the process.

4.4.4 Keeping Clean Hands

The fact that it is possible to agree to a dirty compromise without sacrificing one’s integrity also suggests a way of avoiding the problem of dirty hands. Recall that, as de Wijze characterizes them, dirty hands cases involve a “justified betrayal of persons, values
or principles.” Now we can ask, does agreeing to a dirty compromise necessary involve such a betrayal? I contend that the answer is “no”. If, as suggested above, one has a moral justification for agreeing to a compromise that is rooted in one’s own principles and values, then agreeing to the compromise does not betray those principles and values. Nor does one betray any persons if, acting on their behalf, one agrees to a compromise that one justifiably believes is in their best interests given the range of realistically available alternatives.

In the case of the Vermont compromise, it might be alleged that advocates of same-sex marriage betrayed principles of equality or the LGTB community, but this would be a mistake. In agreeing to the compromise, advocates of same-sex marriage, we can suppose, agreed to what they believed to be the best realistically available means of securing greater equality and recognition for same-sex couples. This cannot accurately be described as a betrayal of their principles or of same-sex couples. It may be accurate to say that the system as a whole betrayed same-sex couples by affording them an inferior status, but that is not the fault of advocates of same-sex marriage who agreed to the compromise. Rather, it is the fault of those who opposed affording same-sex couples the right to marry.

---

Note that I am not denying the possibility of dirty hands cases. I leave open the possibility that, especially in the more extreme cases described in the literature such as cases of torture or false imprisonment, one may act rightly, but nonetheless betray a person or principle in the way de Wijze describes. More generally, I grant, at least for the sake of argument, that it is possible, as Walzer says, that “a particular act of government (in a political party or in the state) may be exactly the right thing to do in utilitarian terms and yet leave the man who does it guilty of a moral wrong.” My point is just that agreeing to a dirty compromise need not be a case of this type because, in doing so, one need not violate any moral principle or betray any persons. To think otherwise is to conflate the moral wrongness of the content of a compromise with the moral value of the act of agreeing to it.

4.5 Harder Cases and Further Issues

I have appealed to the Vermont compromise as my primary example of a dirty compromise, but the Vermont compromise is, in several respects, an easy case. It is easy to see it as a step in the right direction, and the civil unions law did not create an impediment to further moral progress. Moreover, the law did not make anyone worse off

---

63 Walzer, “Political Action,” 161.
compared to the status quo. Not all dirty compromises are as clear cut or easy to evaluate.

In closing, it is worth briefly discussing two types of dirty compromises that introduce additional difficulties.

The first type of more difficult case is what I will call obstructive compromise. An obstructive compromise is one that, though it may provide some advance over the status quo, makes further advances more difficult than they otherwise would have been. In deciding whether to accept an obstructive compromise, one faces the choice between foregoing greater long-term progress in the interest of short-term progress, or forgoing short-term progress in hopes of achieving greater long-term progress. For example, suppose that Vlad is an animal rights activist who believes that, with few exceptions, it is seriously morally wrong to slaughter animals for food. Vlad campaigns especially hard against factory farms, and believes that factory farms are incompatible with a just society. However, Vlad also recognizes that there is no hope of banning or eliminating factory farms within the near future. Now suppose that Vlad is offered the opportunity to help pass a law that will significantly improve conditions for animals on factory farms. Vlad recognizes that the proposed law will significantly reduce the amount of animal suffering in his country. However, he also recognizes that the law will make eradicating factory farms more difficult in the long run because the better conditions are for animals on factory farms, the more difficult it is for Vlad and his fellow animal rights activists to
convince people that such farms should be eradicated. So, Vlad faces a choice between helping to improve conditions for animals in the short-term while making his ideal more difficult to achieve, or foregoing those short-term improvements in hopes of achieving greater gains in the future.

The fact that a compromise is obstructive should not affect our assessment of whether the content of the compromise is morally acceptable. That assessment should still depend only on whether the content of the compromise—the law, policy, or course of action agreed to—violates basic standards of justice. Moreover, whether it is morally acceptable to agree to an obstructive compromise should still be evaluated in relation to the realistically available alternatives; it is just that determining which of these alternatives is morally best becomes more difficult. This judgment will depend on how large the short-term gains are, how probable or improbable it is that larger long-term gains can be made by forgoing the compromise, how much larger the potential long-term gains are, and how long it will take to realize any of those larger gains. All of these factors may be very difficult to judge, especially without the benefit of hindsight. Moreover, there may be no obvious way of determining whether the potential long-term gains morally outweigh the short-term gains, or vice versa. However, these additional complexities do not alter the primary conclusions of this chapter: Agreeing to an obstructive dirty compromise can be morally acceptable if doing so is the best realistically available
alternative. Agreeing to such a compromise need not undermine one’s integrity so long as one’s own principles and values give one sufficient reason to do so. And agreeing to such a compromise need not leave one with dirty hands so long as one does not betray any principles or people.

A second type of more difficult case is what I will call sacrificial compromise. A sacrificial compromise is one that sacrifices progress towards one morally desirable end in order to secure progress towards a different morally desirable end. For example, a compromise might involve sacrificing some gains in racial equality in order to secure gains in women’s rights. The Vermont compromise was easier to justify to same-sex couples because, though the resulting law wrongly discriminated against them, it also gave them sweeping new legal rights and benefits. Sacrificial compromises are more difficult to evaluate and justify because those who suffer from the concessions of the compromise are not also the people who benefit from whatever gains the compromise achieves. Often there will be no obvious way of determining whether the loss to those that suffer from the concessions is morally offset by the gains that the compromise makes for others.

As with obstructive compromise, the additional complexity of sacrificial compromises does not require a fundamentally different approach to evaluating them. Whether the content of a sacrificial compromise is morally acceptable still depends simply
on whether it violates basic standards of justice, and whether it is morally acceptable to agree to the compromise still depends on whether doing so is the best realistically available alternative. Of course, how to determine this is a difficult question, but not one that I can answer here.

Sacrificial compromises raise special problems for integrity and the question of dirty hands. This is because in cases of sacrificial compromise, one might aptly be said to betray one value or group in the name of another. For example, if one concedes some possible gain in racial equality in order to achieve a gain in gender equality, then one might rightly be said to have betrayed the value of racial equality or betrayed disadvantaged racial groups. If one holds racial equality to be a sacred value—one that cannot be traded off against other values—then agreeing to such a compromise may be inconsistent with one’s values, and therefore a failure of integrity. Moreover, betraying one principle or group in service to another seems to fit de Wijze’s description of a characteristic case of dirty hands; it is a case that involves “justified betrayal of persons, values or principles.” I take no stand on these issues here; they are best left for more detailed discussion elsewhere.

64 There is no such betrayal in the Vermont compromise because the concessions made were made to secure benefits for the very value and people that were the subject of the concession.
While obstructive and sacrificial compromises raise additional normative questions that merit further consideration, they do not threaten the major arguments or conclusions of this chapter. Even in these harder cases, it remains true that the content of the compromise should be evaluated relative to basic standards of justice, and the act of agreeing to the compromise should be evaluated relative to the range of realistically available alternatives. Moreover, I have not tried to show that it is morally acceptable to agree to *all* dirty compromises, or that *all* dirty compromises are compatible with preserving one’s integrity and keeping clean hands. Even if there exists a class of hard cases, such as sacrificial compromises, that turns out to be incompatible with preserving integrity or keeping clean hands, this would not threaten my conclusion that at least some dirty compromises are compatible with preserving integrity and keeping clean hands.
References


_Foods Food Demand Survey_. Oklahoma State University Department of Agricultural Economics, 2015.


Shani, Danielle, "Knowing Your Colors: Can Knowledge Correct for Partisan Bias in Political Perceptions?" Annual Meeting of the Midwest Political Science Association, Chicago, 2006.


Biography

Aaron James Ancell was born July 7, 1988 in Kelowna, British Columbia, Canada. He attended Simon Fraser University, graduating with a Bachelor of Arts with Honours in Philosophy in May 2011, and began working on his Ph.D. in Philosophy at Duke University in August 2011. His publications include:


From 2011 to 2015, Aaron’s doctoral studies were funded in part by a Doctoral Fellowship from the Social Sciences and Humanities Research Council of Canada. He is also the recipient of a Katherine Goodman Stern Fellowship, a Kenan Graduate Fellowship, two Summer Research Fellowships, and a Nannerl Keohane Distinguished Visiting Professorship Graduate Student Research Award, all from Duke University.