

THE ENLIGHTENMENT OF SYMPATHY

*Justice and the Moral Sentiments in
the Eighteenth Century and Today*

MICHAEL L. FRAZER

OXFORD
UNIVERSITY PRESS

2010

CONTENTS

- Introduction: A Tale of Two Enlightenments, 3
1. Sentimentalism before Hume, 15
 - I. The New Science of Human Nature, 17
 - II. Religious and Metaphysical Foundations, 25
 - III. Theories of Justice, 34
 2. Hume's Free-Standing Sentimentalism, 40
 - I. Sympathy and the Moral Sentiments, 41
 - II. Moral Development, 46
 - III. Hume's Normative Theory, 56
 3. Hume's Conservative Sentimentalism, 65
 - I. Hume's Theory of Justice, 66
 - II. The Sentimentalist Case against Hume's Theory, 76
 4. Adam Smith's Liberal Sentimentalism, 89
 - I. Sentimentalism and Individualism, 90
 - II. The Space between Actor and Spectator: Sympathy and Moral Judgment, 97
 - III. The Space between Actors: Justice and Natural Jurisprudence, 102
 5. Kant's Abandonment of Sentimentalism, 112
 - I. The Critical-Period Position on the Foundations of Morals, 114
 - II. The Critical-Period Normative Evaluation of Sympathy, 119
 - III. The Critical-Period Theory of Affects and Passions, 127
 - IV. A Contrasting Precritical Position, 132

6. Herder's Pluralist Sentimentalism, 139
 - I. Sentimentalism and the Problem of Diversity, 142
 - II. From Sympathy to Diversity, 147
 - III. From Diversity to Empathetic Understanding, 154
 - IV. From Empathetic Understanding to Justice, 160
 7. Sentimentalism Today, 168
 - I. Sentimentalism and Social Science, 169
 - II. Sentimentalism and Normative Theory, 174
 - III. Sentimentalism and Political Practice, 178
- Notes, 183
- Bibliography, 213
- Index, 229

Introduction

A Tale of Two Enlightenments

1. The Reflective Revolution

Human beings have a unique capacity for introspection, the ability to reflect upon their own thoughts and deeds. Pausing for moral reflection involves taking a step back from our habits of action as individuals, examining them, and determining whether we ought to continue as before. Political reflection involves taking such a perspective on the laws and institutions by which we govern our actions together. Moral and political philosophers pursue refined forms of reflection as a vocation, but all of us engage in reflection from time to time. Indeed, upon reflection, we often determine that we should all really take more time for reflection.

Moral reflection can lead us to revise how we are currently leading our lives. Political reflection can lead us to revise how we are governing ourselves as a society. Both involve comparing how things are currently done to standards of how they ought to be done. In the political sphere, the relevant normative standards are commonly called standards of justice. Any law, institution, or political practice is liable to rejection upon reflection if we conclude that it is unjust. Our standards of justice and morality are then themselves subject to revision upon reflection, and then further revision upon further reflection. Eventually, we may reach the conclusion that some of our standards are unlikely to be revised much further. We then treat these standards as—at least tentatively—authoritative. In the inelegant language of recent philosophy, they become our considered convictions in reflective equilibrium.¹

When we engage in reflection, we determine our own moral and political standards. When we insist on reflective freedom—on the right and responsibility of all to reflect for themselves—we insist on the importance of autonomy, or self-legislation. The political metaphor of autonomy—so common that we often forget that it is a political metaphor—is a product of the eighteenth-century. The political revolutions of that time were grounded in a notion of literal, collective self-legislation through republican governance. The intellectual revolution of the same period, known as the Enlightenment,

uses the enactment of legitimate laws by a self-governing people as a metaphor for the determination of principles of justice and morality through individual reflection.² According to this account, any law (even a legitimately enacted one) may be opposed as unjust if it conflicts with our considered convictions in reflective equilibrium. Kant's motto "*Sapere aude!* Have the courage to make use of your *own* understanding!" (WE 8:35) was the revolutionary slogan of the Enlightenment.

Revolutionaries, however, always have trouble maintaining a united front. Not all Enlightenment thinkers interpreted Kant's call for reflective autonomy as Kant himself did. The study of eighteenth-century moral and political thought reveals that there were in fact many competing Enlightenments, each with its own account of reflective autonomy. Although it is important not to oversimplify the intellectual diversity of the period, this book focuses on the contrast between two primary streams of eighteenth century analysis of moral and political reflection. The first, which I am calling *rationalist*, corresponds to our common conception of the eighteenth century as the "age of reason." The second, which I am calling *sentimentalist*, suggests an age not of reason alone, but also of reflectively refined feelings shared among individuals via the all-important faculty of sympathy.³ This is not to say that every moral and political thinker of the Enlightenment can be easily classified as exclusively "rationalist" or "sentimentalist." Many of the greatest thinkers of the period—most notably Jean-Jacques Rousseau—evade such simple categorization. But there was clearly an ongoing debate in the eighteenth century over the nature of reflective autonomy—a debate in which many took an identifiably rationalist position and many others an identifiably sentimentalist one.⁴

Most of the major philosophers of the sentimentalist Enlightenment—such as the Third Earl of Shaftesbury, Joseph Butler, Francis Hutcheson, David Hume, and Adam Smith—were British, while many of the major rationalists of the period were French or German. It is important, however, not to confuse the distinction between the rationalist and sentimentalist Enlightenments with the distinctions that have been drawn among the various "national" Enlightenments.⁵ There were many rationalists in Britain—among them Samuel Clarke, William Wollaston, and Richard Price. There were also many sentimentalists on the continent, most notably J. G. Herder and, at least for a time, his teacher the precritical Immanuel Kant. Sentimentalism and rationalism should not be seen as national worldviews, but as competing positions on a single, trans-national debate central to the intellectual life of the eighteenth century—a debate on the nature of reflective autonomy that continues to this day.

For better or worse, the heirs of the rationalist rather than the sentimentalist Enlightenment now dominate both philosophy and social science. Enlightenment sentimentalism has long been underappreciated by comparison with Enlightenment rationalism—as the very notion of the eighteenth

century as “the age of reason” will attest. Even philosophers today who are well aware of the centrality of moral sentimentalism to eighteenth-century intellectual life tend to define the Enlightenment in purely rationalist terms. John Rawls, for example, defines “Enlightenment liberalism” as a “comprehensive liberal and often secular doctrine founded on reason,” one capable of supporting political morality through a direct appeal to the rational faculties alone.⁶ Normative theorists and social scientists who are now rediscovering the importance of emotion in our moral and political lives have thus often been led to believe that they are refuting the philosophy of the Enlightenment, rather than lending support to one popular eighteenth-century view of reflective autonomy over another.⁷

This book is a reclamation and defense of the sentimentalist account of reflection. After a brief account of Shaftesbury, Butler, and Hutcheson in chapter 1, the book will focus on three of sentimentalism’s strongest defenders—David Hume, Adam Smith, and J. G. Herder—as well as its single greatest rationalist critic: the older, critical-period Immanuel Kant. Although it consists primarily of the interpretation of centuries-old texts, this is not a work of intellectual history *per se*. My aim is not to trace the development of sentimentalism in its eighteenth-century context. Instead, I seek to reclaim the sentimentalist account of reflection as a resource for enriching political science, political philosophy, and political practice today, a resource often overlooked due to the widespread influence of the opposed rationalist account.

2. Competing Reflective Regimes

Insofar as it has been treated as part of the Enlightenment at all, sentimentalism has long been seen as a purely destructive doctrine. According to this view, once the moral impotence of reason has been decisively established, one is left with no foundation for justice and morality other than mere feeling. Not only is this interpretation of sentimentalism’s negative position misleading, but it also neglects the most useful insights the sentimentalist Enlightenment has for today, all of which stem from its positive project: specifically, from its explication and defense of a particular conception of reflective autonomy. Although the rationalist and sentimentalist Enlightenments are united in their characteristically enlightened endorsement of such autonomy, they have different notions of what it means to legislate moral and political standards for oneself. They are divided on the nature of the self who is doing the legislating and the nature of the self who is obeying the standards so legislated. To use a Platonic locution, they disagree about the proper psychic regime. Rationalists separate the legislative faculties of the mind—identified as “reason”—from the faculties that obey. Sentimentalists, on the other hand, see the standards created in ethical reflection as products of the mind as a whole, and do not distinguish sovereign and subject aspects of the mind.

Admittedly, this reading of sentimentalism as a kind of democratic egalitarianism of the soul is in sharp contrast to the standard interpretation of sentimentalist moral psychology. Hume in particular is conventionally read as advocating a psychic regime as fully hierarchical as that of his rationalist opponents. Under this reading, he disagrees with them only with regard to which faculties are to be sovereign and which are to be subject. While rationalists from Plato onward maintained that reason is rightly the master and passion rightly the slave, Hume famously counters that “reason is, and ought only to be, the slave of the passions” (T 2.3.3.4). But it will be argued in chapter 2 that this memorable bit of rhetoric distorts Hume’s true view. Although philosophers may rightly distinguish the operations of reason from those of passion, Hume consistently maintains that in reality the two are “uncompounded and inseparable” (T 3.2.2.14). It is true that Hume believes reason alone is powerless to motivate action; it is in this sense that reason is and ought to be passion’s slave. Yet the sentiments that Hume describes as motivating moral action are not merely passions, but products of the mind as a whole, reason and imagination included. It is from passion alone that these sentiments get their motivational impetus, but moral sentiments are much more than mere impetus. The contrast between rationalism and sentimentalism is therefore best understood as the contrast between a hierarchical view of the moral soul on the one hand, and an egalitarian view on the other—an egalitarian view in which normatively authoritative standards are the product of an entire mind in harmony with itself.

Despite their hierarchical view of the proper psychic regime, Enlightenment-era rationalists considered their theory to be one of reflective autonomy. This was possible because they identified themselves with the sovereign, legislative faculty and not with the subject faculties that obey its legislation. Although the other features of the mind and personality are plagued by contingency, reason deals only with necessary truths. Although my emotion, imagination, and memory are all part of causal nexuses both natural and social, my reason is free. If I am to think of myself as free from natural and social contingency, Enlightenment rationalists maintained that I must think of my true self as purely rational. If my actions and my standards of action are to be truly my own, they maintained that it is this real self that must be sovereign, legislating standards in reflection and dictating behavior in practice. And since the true self is identified with a single faculty we are all held to share, the true self of all individuals is fundamentally the same. Sentimentalism, by contrast, adopts a different attitude toward contingency, and identifies the true self with the whole self, contingent social and psychological elements included.

Unlike some of the more extreme rationalists of ancient times, Kant and his Enlightenment-era allies rarely denied that social and psychological contingencies are responsible for much of our behavior. Rather than seek to

extirpate the power of contingency from human life, they instead sought to bring all contingent forces under rational control, so that these forces guide us to the very same standards and practices that reason necessarily and authoritatively demands. Even if my norms or behavior are the product of non-rational social and psychological factors, if these contingent forces have been made to comply with the dictates of my better, noncontingent self, then this behavior is rationally justified. The Enlightenment-era rationalist position is generally Platonic, not Stoic; the passions are not to be banished from the psychic regime, but are to obey their superiors, and keep to their proper place. The duties of their station involve keeping quiet during the purely rational process of proper moral and political reflection, and then obeying the rationally authoritative principles that emerge.

3. The Normative and the Descriptive

The competing reflective regimes offered by rationalism and sentimentalism can be understood to combine two distinct elements. To use Hume's most famous distinction, they both offer a theory of what "is" and a theory of what "ought to be"—a descriptive moral psychology that explains what goes on when we engage in moral and political reflection and a normative theory that explains why the standards we reach through such reflection must be treated as authoritative. While sentimentalism describes reflection as a matter of feeling and imagination as well as cognition, rationalism describes reflection as a matter of rational cognition alone. While sentimentalism understands normativity as stemming from the peace and satisfaction of a mind able to bear its own holistic survey, rationalism sees normativity as authoritative legislation by the faculty of reason—here identified with our true, autonomous self.

The disagreements between rationalists and sentimentalists within the respective fields of descriptive psychology and normative theory imply a further disagreement about the relationship between these two fields. The sentimentalist Enlightenment can therefore provide normative theorists and empirical psychologists today with a program for their fruitful collaboration that differs markedly from the model of interdisciplinarity suggested by the dominant tradition of Enlightenment rationalism. Although Hume is widely thought to have insisted on the strict separation of empirical description from normative evaluation, the empirical study of moral and political phenomena is actually central to his normative philosophical project, as it is for all Enlightenment sentimentalists.⁸ Of course, before the emergence of full academic specialization in the twentieth century, all moral and political philosophers—rationalists and sentimentalists alike—were also social scientists. The distinction between philosophy and science was unknown in the

Enlightenment; in the eighteenth century, these terms were used more or less interchangeably. Yet the relationship between empirical and normative theory takes a different form among Enlightenment-era sentimentalists than it does among their rationalist contemporaries.

The rationalists begin with normative theory—discovering valid moral principles on the basis of reason alone—and only then turn to empirical investigation to determine how imperfect, real-world creatures such as ourselves may be better brought in line with reason’s authoritative demands. Sentimentalists, however, begin where rationalists end—with the empirical investigation of what actually leads real-world human beings to follow the norms that they do. Sentimentalists describe these norms as the outcome of our moral sentiments, products of the mind as a whole. The faculty of sympathy is central to their descriptive etiology of these moral sentiments. Sympathy is the bridge between the social and the psychological, the faculty by which inner mental states are shared among individuals. So the empirical social-psychology of reflection offered by sentimentalism can be understood largely in terms of the reflective expansion and correction of our sympathetic bonds to our fellow human beings. In contrast to the picture of legislative reason offered by rationalism, sentimentalism here offers a richer account of the process by which we search for reflective equilibrium. Specifically, sentimentalism offers an empirically grounded sociology and psychology of moral and political reflection that focuses on the key social-psychological faculty of sympathy.

Yet the sentimentalist account of reflection is not merely descriptive. Of course, no account of the development of our moral psychology could ever, by itself, justify our moral commitments. To believe otherwise is to confuse an empirical explanation of the origins of a value commitment with a demonstration of its genuine normative authority, to mistake statements of what is for statements of what ought to be in precisely the way that Hume so famously warned against. Yet once we accept a sufficiently sentimentalist description of our moral psychology—one that sees our moral commitments as reflective outgrowths of basic human emotions of which we all heartily approve—a distinctively sentimentalist method for normatively justifying these commitments immediately presents itself.

John Rawls imagines a contemporary reader faulting Hume’s ethics for being nothing more than descriptive moral psychology. Yet to maintain such a position, Rawls argues, is to seriously misunderstand Hume. Focusing on the conclusion of the *Treatise*, Rawls instead interprets Hume as maintaining “that his science of human nature . . . shows that our moral sense is *reflectively stable*: that is, that when we understand the basis of our moral sense—how it is connected with sympathy and the propensities of human nature, and the rest—we confirm it.”⁹

The sentimentalists know that we not only approve and disapprove of our individual actions and our shared political practices, but also of our own sentiments of approval and disapproval. Sometimes, our moral sentiments are premised on unfounded beliefs, and can therefore be rejected as irrational. Yet even when there is no rational error in our moral judgment, we sometimes still disapprove of our own moral sentiments—not as irrational, but as immoral. The fact that we can have higher-order moral sentiments—that we can approve or disapprove of our own approval and disapproval—allows for a process of reflection in which the mind as a whole repeatedly turns on itself, winnowing out those sentiments that cannot bear reflective survey.

As the mind turns in on itself through the process of reflection, the moral sentiments must bear the test of close scrutiny, and may be revised or rejected if they are found wanting. This test is sometimes directly reflexive. Our set of moral sentiments, if internally contradictory, may lead us to disapprove morally of some or all of these very sentiments themselves. However, it can also be reflective in a broader way. For example, our moral sentiments may be rejected if they are grounded on premises found to be false upon rational inquiry, or if they make it impossible to achieve happiness.¹⁰ As the mind as a whole repeatedly turns in on itself, reflection leads us through a gradual progress of moral sentiments, as more and more of our contingently given convictions are revised on both directly reflexive and broadly reflective grounds. Only those moral sentiments that have endured when we reach reflective equilibrium can be treated as authoritative. Only they can play a role in a truly happy human life.

Hume, Smith, and their fellow sentimentalists thus stand in direct opposition to Hobbes, Mandeville, Marx, Nietzsche, Freud, and the other reductive debunkers of morality who often dominate discussions of descriptive moral psychology. These debunkers have led many modern thinkers, Rawls included, to worry that we will come to “doubt the soundness of our moral attitudes when we reflect on their psychological origins.”¹¹ If Hume’s description of our moral psychology is correct, however, a complete understanding of the origins of our proper moral commitments can only lead us to affirm these commitments. Although such a defense of human morality cannot convince a committed rationalist, who is determined to find a morality binding on any rational being as such, the sentimentalist can simply reply, citing Rawls, that “beings with a different psychology either have never existed, or must soon have disappeared in the course of evolution.”¹²

It is thus crucially important for the success of a sentimentalist normative theory that its empirical description of human psychology prove accurate—that human beings actually have the mental faculties that they describe. “As this is a question of fact,” Hume reasons, “and not of abstract science [i.e., a priori philosophy], we can only expect success by following the experimental method, and deducing general maxims from a comparison of particular

instances” (EPM 1.10). On the title page of the *Treatise of Human Nature*, Hume therefore introduces his work as “an attempt to introduce the experimental method of reasoning into moral subjects” (T Title Page).

Hume, Smith, and Herder, of course, were not experimental psychologists in the contemporary, social-scientific sense. Rather than using the term “experiment” to describe the controlled tests of today’s laboratory science, Hume instead associates “careful and exact experiments” with the simple “observation of those particular effects which result from . . . different circumstances and situations” (T Introduction:8). Since experimentation is equated with careful observation, in the case of “moral subjects” experimentation involves scrutinizing the social world around us and the psychological forces within us. The Enlightenment sentimentalists’ introspective technique in describing our moral psychology is therefore less akin to that of today’s social-scientific psychologists than to that of twentieth-century psychoanalysts or phenomenologists. Their “experimental method” might not appear as such to those today with certain set notions of what qualifies as experimental, but it is meant to describe actual human experience accurately, experience that these authors hope that readers will recognize as their own. They offer an account of reflection based on empirical observation, which could be either supported or falsified by the more methodologically rigorous empirical research being carried out in moral and political psychology today. The support that current research in moral psychology lends to sentimentalism will be discussed in the final chapter of this book.

4. In Defense of Reflective Sentimentalism: Overview of the Book

Despite its empirical and theoretical richness, reflective sentimentalism is vulnerable to a number of important criticisms. As this book moves through a chronologically organized examination of each of the Enlightenment-era philosophers under discussion, it is also structured so as to respond to each of these criticisms in turn. The explanation of how sentimentalism can rise to the challenges posed by each of these criticisms should make the attractiveness of its approach to moral and political reflection clear.

Chapter 1 will consist of an overview of the work of the three British philosophers from the first half of the eighteenth century whose work most influenced Hume and the later sentimentalists: Francis Hutcheson, Bishop Joseph Butler, and Anthony Ashley Cooper, Third Earl of Shaftesbury. Although all three made important contributions to the argument that justice and morality cannot be products of reason alone, this chapter will consider these authors primarily insofar as they presented the problems that Hume, Smith, and Herder were left to work out in their own writings. The first of

these challenges was the need for a free-standing sentimental ethics—one that does not rely on religion or metaphysics to establish the normative authority of our moral sentiments. The second challenge was to explain how our moral sentiments can lead to a sense of justice capable of being instantiated in law-governed political institutions.

The next chapter concerns Hume's answer to the first of the two challenges described in chapter 1: the challenge of constructing a free-standing sentimental theory that can provide authoritative normative standards. Perhaps the most common criticism of the sentimental understanding of moral and political reflection is that it cannot actually provide binding norms, at least not without relying on external religious or metaphysical sources of moral authority. Without a single governing faculty that is identified with the higher self—without a single sovereign in the soul bringing the rest of the soul in line—many worry that the individual is reduced to a kind of psychic chaos, doing whatever he or she feels like doing at the moment without the guidance of stable, reflectively endorsed standards. Yet to identify sentimentalism with such wantonness is to confuse psychic democracy with psychic anarchy. Just as a nation as a whole can govern itself through democratic deliberation and the rule of democratically enacted law, so too can a mind as a whole govern itself through reflectively refined and endorsed moral sentiments.

Much of the blame for this confusion lies with the Enlightenment sentimentalists themselves, most notably David Hume. In order to refute rationalism, Hume often used rhetoric that might be taken to suggest that, like his rationalist opponents, he too saw one faculty as sovereign and others as subject. In contrast to the opposing view in which reason was sovereign, Hume's rhetoric suggested a theory of reflection in which passion was the exclusive sovereign. Chapter 2, however, rejects this misleading reading of Hume's position. Our moral sentiments, Hume makes clear, are not immediate passions; they also include cognitive and imaginative elements and are subject to revision upon reflection. Hume's reflective sentimentalism should not be confused with sheer wantonness or "doing as one feels." To the contrary, Hume's moral philosophy is capable of providing a full normative justification for our commitment to genuine virtues, while also giving us grounds to reject the sham virtues advocated by others. Indeed, some of the virtues that Hume himself recommends may deserve to be rejected according to the criteria of his own sentimental ethics. I argue in chapter 3 that Hume's own theory of justice must be discarded for precisely such reasons.

This refutation of Hume's theory of justice is meant to address a second objection to reflective sentimentalism. Many argue that, even if sentimental reflection can provide standards by which to govern our actions and our politics, these are the wrong kind of standards. Specifically, since it begins from the social-psychological forces that are responsible for our accepting the norms and standards that we do, a concern arises that sentimentalism can

never successfully criticize existing institutions or practices, that it may be too complacent or conservative. Again, much of the blame for this confusion rests with Hume himself, who was sometimes unduly complacent toward existing standards and practices—although he was also sometimes an adamant advocate of moral, political, and economic reform. Nowhere is Hume's undue conservatism more evident than in his theory of justice, which describes justice as the virtue of obeying existing social conventions that promote the public interest. Yet I argue that, far from there being a direct connection between Hume's moral sentimentalism and his political conservatism, the two aspects of his philosophy are in fact incompatible.

When sentimentalists are excessively complacent, I argue, it is simply because they have failed to subject their convictions to the test of sufficient sentimentalist reflection. As our moral sentiments progress, we reflectively expand our sense of sympathy, increasing the degree to which we share specific feelings with an ever-wider array of our fellows. As our circle of sympathy expands, we will in turn demand reforms that show increasing concern for all. Hume's sentimentalist reflections were advanced in this liberal and individualist direction by his close friend Adam Smith. The most important of Smith's improvements to Hume's sentimentalism for purposes of political theory, and the subject of chapter 4, is a deemphasis of the importance of sympathy with the public interest in our commitment to justice. Smith argues that a proper commitment to justice stems primarily from sympathy with particular individuals—specifically, sympathy with the resentment felt by individual victims of injustice.

If there is any single thinker whose influence has led contemporary political theorists to be wary of sentimentalism, it is Immanuel Kant. Yet Kant actually embraced reflective sentimentalism at one point in his philosophical career, only to reject it later during his so-called critical period. Chapter 5 seeks to explain this turn in Kant's moral philosophy, and to argue that Kant's reasons for rejecting reflective sentimentalism need not lead us to do the same. Rather than accepting the conventional view of the critical-period Kant as an opponent of any form of emotionally motivated behavior in favor of a coldly rationalist sense of duty, I argue that the later Kant is willing to endorse many emotions that enable finite creatures to better conform to duty's categorical demands. He nonetheless categorizes sympathy in particular as having very limited ethical value, and argues that no form of fellow-feeling must ever be allowed to threaten the rational self-control that he identifies with autonomy. Yet once we see reflective sentimentalism as dependant on a harmony of all the faculties in a reflectively stable psyche, rather than advocating the utter slavery of reason to passion, such an approach to moral and political reflection need not be seen as a threat to autonomy. Our choice is not between the autonomy of reason and the heteronomy of feeling, but rather between two competing theories of what reflective autonomy involves.

Aside from its potential threat to autonomy, another concern is that sentimental reflection cannot provide moral and political standards that can be shared by all. While rationalist reflection can derive a single set of authoritative standards from a single universal faculty of reason, sentimentalism must assume an implausible degree of uniformity among the contingent psychological features of human beings to do the same; Hume and Smith made precisely this implausible assumption in order to do so. Another sentimentalist alternative, however, is to accept rather than reject the pluralism of human standards, while still insisting that plural sets of standards can join in a single overlapping consensus supporting basic principles of justice and reciprocity. Some factors for which sentimentalism allows entry into moral and political reflection are admittedly exclusive to particular individuals or groups. Others, however, are permanent features of human nature. While still contingent in the sense that rational creatures could always have evolved or been created without them, these features are shared by virtually all members of the human species. Among the most important features of our common human nature is an ability to sympathize with our fellows, and to share the feelings of even those whose values and worldviews are very different from our own. Such pluralist sympathy, in turn, can help form the basis of an overlapping consensus demanding fairness and reciprocity even among those who have otherwise conflicting moral sentiments. Exactly such a pluralist sentimentalism, I argue in chapter 6, can be found in the work of J. G. Herder, a student of the precritical, sentimentalist Kant and a vociferous opponent of the critical-period, rationalist Kant.

Once, with Herder, sentimentalism takes this pluralist turn, it reveals itself to be especially helpful in the diverse democracies of today. Although citizens in such societies may have opposing moral convictions, they must come to some agreement on basic principles of justice. While some might hope to build a consensus behind these principles on the basis of reason alone, this is not the only faculty that all of us share. Given that our task is to build a just society for human beings, and not for rational beings as such, there is no reason why we cannot also appeal to the many nonrational features of the human psyche that we possess in common—our emotion, our imagination, and our ability to share in the inner life of others through sympathy. The final chapter of this book will outline the many and varied improvements to normative political theory, empirical political science, and practical political life that sentimentalism can offer us today.

Contemporary scholars are increasingly coming to reject the rationalist paradigm that has long dominated a wide variety of disciplines in the humanities and social sciences. Just as psychologists and neuroscientists have been examining the central role of emotion in all mental processes, and philosophers have been refuting neo-Kantian claims that reason alone can support justice and morality, political scientists have begun acknowledging

the legitimacy of sentiment throughout all spheres of civic activity. Chapter 7 argues that greater attention to Enlightenment sentimentalism could serve to enrich not only such contemporary scholarship on ethics, politics, and psychology, but also the actual practices of our civic life.

By offering competing descriptive and normative theories of moral and political reflection, rationalism and sentimentalism can lead to different theories of how collective political activity should be managed. If reflection is a matter of rational cognition alone, then appealing to others' emotions or imagination only serves to undermine their autonomous political reflection. Although rhetoric, narrative and other emotionally loaded tools of political persuasion may be used to help others better abide by the principles that reason demands of us all, they could just as easily be used to lead us away from these principles. Sentimentalism, by contrast, sees such forces as emotion and imagination as part of the process of moral and political reflection itself. Our use of rhetoric and narrative here enhances rather than diminishes the reflective autonomy of others. In an age of widespread political apathy, a reconsideration of the work of Hume, Smith, and Herder suggests an impassioned rather than a dispassionate politics, politics in which a diverse citizenry strives to spur itself continually onward in the reflective revision of its moral sentiments. To forego the rich reflective resources of sympathy and the moral sentiments in either political theory or political practice would, I conclude, be a terrible waste.

CHAPTER 1

Sentimentalism before Hume

One problem in many recent treatments of Enlightenment sentimentalism is the tendency to discuss the work of Hume exclusively, often using “Humean” as a synonym for “sentimentalist.”¹ Doing so leads scholars astray in two ways. First, by ignoring authors who improved on Hume’s work they fail to present sentimentalism in its most compelling form. Chapter 4 of this book will argue that when Hume’s sentimentalism went wrong in its theory of justice, Smith offered a liberal sentimentalist solution. And chapter 6 will argue that while both Hume and Smith failed to appreciate the full scope of human diversity, Herder succeeded in doing so in a distinctly sentimentalist way. To use an old cliché, Smith and Herder may have been standing on the shoulders of a philosophical giant, but this allowed their theories to rise beyond his own.

It is also important to note that there were important reflective sentimentalists before Hume. Scholars have long exaggerated the originality of Hume’s ethics and political philosophy. Sheldon Wolin, for example, credits Hume for nothing less than “effecting a minor revolution in political thought” by being the first to understand ethics and politics “in psychological rather than juridical categories.”² Yet Hume knew that his appeal to reflectively stable sentiments to both explain and justify our convictions was nothing fundamentally new, and that readers familiar with previous writers on the subject would never mistake it as such. Hume cites his sentimentalist precursors Shaftesbury, Hutcheson, and Butler alongside their interlocutors Locke and Mandeville as among the “late philosophers in England who have begun to put the science of man on a new footing” (T Intro. 7). Although he observes that “they differ in many points among themselves,” Hume praises these authors for “founding their accurate disquisitions of human nature entirely upon experience” (T Abstract 2).

The purpose of this chapter is not to do full justice to the complex sentimentalist theories put forward by Shaftesbury, Butler, or Hutcheson, let alone to provide a history of the moral sentiments before Hume. My goal

here is merely to consider Hume's debts to his predecessors, setting the scene for the chapters to follow. The result is not the sort of debunking in which what were once thought to be the original insights of genius are revealed to be borrowed from previous authors. To the contrary, examining how Hume was influenced by Shaftesbury, Butler, and Hutcheson also helps us see the ways in which he was not influenced by them. After briefly discussing how many fundamental features of reflective sentimentalism were in place before Hume, this chapter will then argue that Hume departed from his sentimentalist predecessors in two all-important ways.

First, Hume successfully freed his version of sentimentalism from theological and metaphysical foundations. For Christian sentimentalists such as Butler and Hutcheson, the normative authority of our moral sentiments ultimately derives from the fact that they were built into our nature by God for the achievement of his intended ends. As for Shaftesbury, it is true that his sentimentalism seems basically free of Christian foundations—and this irreligiousness earned Shaftesbury, like Hume, considerable infamy—but these are replaced by a classical metaphysics with a strong emphasis on natural teleology. As will be clear by the end of chapter 2, however, Hume had a different theory of normativity, one that relies neither on religion nor on metaphysics. Hume's is a self-supporting account of moral and political reflection, one in which the normative authority of our considered convictions in reflective equilibrium does not depend on their correspondence to any superhuman reality, but only on the fact that human happiness requires mental stability.

Second, Hume saw that political philosophy was a weak spot for earlier sentimentalists. In the centuries before sentimentalism, as Wolin observed, Western moral and political thought was basically juridical in inspiration. Both rationalist natural law and divine command theories treated all normative phenomena on the model of a code of just laws. It was thus a relatively simple matter to evaluate the justice of existing political institutions by seeing how their positive laws measured up to either divine or natural standards. By grounding justice in human psychology, sentimentalists made this political-philosophical task considerably more difficult. Further steps were necessary to derive moral standards suitable for evaluating law-governed political practices from the nonjuridical movements of the human heart and mind. Chapter 3 will argue that Hume was not entirely successful in this endeavor, and chapter 4 will go on to elaborate how Adam Smith developed a superior sentimentalist theory of justice. As is so often the case, however, Smith's theory of justice was itself not entirely original, and took its inspiration from suggestions by Shaftesbury and Butler that our sense of justice stems from the resentment we feel when we are the victims of injustice.

I. The New Science of Human Nature

1. *Observing the Human Heart*

In his preface to Hutcheson's posthumously published *System of Moral Philosophy*, William Leechman recalls that, like so many intellectuals of the eighteenth century, Hutcheson was inspired by the success of Newtonian natural science. Leechman notes that his Glasgow colleague "had observed that it was the happiness and glory of the present age that they had . . . set themselves to make observations and experiments on the constitution of the world itself" (SMP Preface, Vol. I, p. xiii). The Enlightenment sentimentalists were united in their insistence that the modern, a posteriori or "experimental" approach be applied to moral as well as natural subjects. Leechman recalls that Hutcheson:

was convinced that . . . a true scheme of morals could not be the product of genius and invention, or of the greatest precision of thought in metaphysical reasoning, but must be drawn from proper observations upon the several powers and principles which we are conscious of in our bosoms.

(SMP Preface, Vol. I, p. xiv)

Yet this method of introspective observation was not solely an application of the techniques of modern natural science. Shaftesbury, for one, insisted that his self-contemplation was inspired by ancient philosophy, representing the best way to fulfill the classical injunction to know oneself. Religious sentimentalists could also defend their empirical-psychological method by appeal to scriptural authority. According to Paul, while the Jews have received God's law through revelation, the gentiles "are a law for themselves . . . since they show that the requirements of the law are written on their hearts, their conscience also bearing witness" (Romans 2:14–15).

The emphasis on the heart is important here. For the Enlightenment sentimentalists, our moral nature can be seen, through introspection, to be based on passionate desires for the good of others, desires often referred to by the umbrella terms "benevolence" or "beneficence."³ Benevolence in this sense is not a cold commitment to altruistic principles, but a warm concern for others' welfare. Today, we might broadly term the warmth involved as a result of "feelings" or "emotions." The former of these terms was already widely used in eighteenth-century English, and the latter was just beginning to take on its modern meaning.⁴ The most common terms for affective phenomena in eighteenth-century English were "passions" and "affections." Although the two terms were often used interchangeably or concurrently, one can distinguish between the two in that the former sometimes has a more negative connotation than the latter. In the Latin Christian tradition, *affectus* was associated with the serene love of God and the divinely commanded love of one's

neighbor, whereas the *passiones* were associated with unruly, earthly desires. Hume is deliberately departing from standard Christian usage in his distinction between calm and violent passions; the more mainstream Hutcheson called the former tranquil affections. Finally, since it carried then, as it does now, the secondary definition of opinion or belief, the term “sentiment” was often used in the eighteenth century to describe affective phenomena that also have a strong cognitive component, as with impassioned moral convictions.

Although Shaftesbury, Butler, and Hutcheson thought benevolent desires could stem from a variety of passions, affections, and sentiments, all acknowledged that one important source of our desire for the good of others is our ability to share these affective phenomena with them. Here, the available terminology is even more slippery. In English, “compassion” comes from the Latin for “suffering with” or “feeling with,” and “sympathy,” from the Greek for the same.⁵ The double meaning of both *pathos* and *passion* leads to an ambiguity concerning sympathy and compassion. In a narrow sense, they might be limited to the sharing of another’s suffering, but in a broader sense they might involve sharing any emotion, pleasant or painful. Eighteenth century authors generally use the term “sympathy” in the broader sense, reserving “compassion” and “pity” for shared suffering.⁶ In this way, British Enlightenment discussions of “sympathy” lack the narrow focus on misery sometimes found in continental discussions of *pitié* or *Mitleid*.⁷ “Humanity” and “fellow feeling” are also sometimes used by English-language writers in this period as rough synonyms for sympathy in this broad sense.⁸ Shaftesbury and Hutcheson use the Latin *sensus communis* to mean something similar, although the phrase is more often used by others to mean simply “common sense.”⁹

Each of these many terms for feeling and for shared feeling carries subtly different connotations, but in everyday speech the distinctions between them are often lost. “It is very difficult to talk of the operations of the mind with perfect propriety and exactness,” Hume observes, “because common language has seldom made any very nice distinctions among them” (T 1.3.8.15). Unlike many philosophers today, eighteenth-century British sentimentalists wrote in standard English rather than adopting a precise terminological system, so their discussions of sympathy, the affections, and the passions are often as inexact as the language itself.

2. Stoic and “Selfish” Interlocutors

Regardless of their terminological difficulties, Enlightenment sentimentalists clearly put great emphasis on powers of the soul that were previously disparaged throughout much of the western philosophical tradition. Authors from Plato and the ancient Stoics to Spinoza and the early modern neo-Stoics had shown a consistent hostility toward feeling.¹⁰ Yet despite the fact that “stoic”

and “sentimental” are now diametrically opposed in ordinary language, sentimentalism can be seen as a continuation and a revision of the modern Stoic revival rather than an outright rejection of it. First, sentimentalism represented a continuation of the Stoic conception of philosophy as a sort of therapy—the art of learning to think, act, and feel properly, and hence live well. Shaftesbury, in particular, was committed to the idea that philosophy must be devoted to this practical task, dismissing metaphysical and epistemological speculations of the kind practiced by his tutor Locke as useless distractions. Second, like most ancient ethicists, sentimentalists believed that human perfection takes the form of moral virtue. Virtue, in turn, takes the form of “the harmony of a mind,” with “knavery . . . mere dissonance and disproportion” (SAA 1.3, p. 129). This harmony is conceived, as it was by the Stoics, as a happy state because of its tranquility, because of the lack of internal conflict or emotional turmoil.

In contrast to the Stoics, however, Shaftesbury and his followers did not see this harmonious tranquility as a result of apathy, or the lack of emotion, but instead as a matter of emotional balance, of a proper “economy of the passions” (IVM 2.1.3, p. 53). The quest for Stoic apathy, Butler complains, can only lead us away from virtue and benevolence. Those who seek it “appear to have had better success in eradicating the affections of tenderness and compassion than they had with the passions of envy, pride, and resentment” (15S 5, p. 71). To be sure, some Enlightenment sentimentalists believed that the ancient Stoics never wished to destroy the benevolent alongside the selfish passions. In a note to the edition of Marcus Aurelius that he translated with James Moor, Hutcheson claims that “the Stoics always maintained that by the very constitution of our nature all men are recommended to the affectionate good-will of all” (MEM 3.5, fn., p. 42). Hutcheson, however, did acknowledge that the Stoics falsely viewed the “rational soul” as “a being distinct from both the gross body and the animal soul, in which are sensations, lower appetites and passions . . . The rational soul, they say, is the man; the seat of true perfection and happiness . . . capable of performing its proper, natural, lovely beatific offices independent of these lower parts” (MEM 5.19, fn, p. 65). Any knowledge of the realities of human psychology, the sentimentalists agree, demonstrates that this is impossible. As Butler puts it:

Reason alone, whatever anyone may wish, is not in reality a sufficient motive of virtue in such a creature as man; but this reason joined with those affections which God impressed upon his heart . . . Neither is affection itself at all a weakness, nor does it argue defect, any otherwise than as our senses and appetites do; they belong to our condition of nature, and are what we cannot do without.

(15S 5, pp. 67–68)

Rather than the Stoics or neo-Stoics, it was Hobbes and, later, Mandeville who were the sentimentalists’ primary opponents. The first edition of Hutcheson’s first book, the 1725 *Inquiry into the Original of Our Ideas of*

Beauty and Virtue, was introduced on its title page as a work “in which the principles of the late Earl of Shaftesbury are explained and defended against the author of the *Fable of the Bees*.” Hobbes, Mandeville, and other defenders of what their opponents called the “licentious” or “selfish system” (see, e.g., EPM App. 2:3 and TMS VII.ii.4), like the sentimentalists themselves, saw ethics as rightly focused on examining the empirical realities of human psychology. What the selfish system observed in the human heart, however, was very different from what sentimentalism saw. Although they were more than willing to grant the passions a central place in human life, Hobbesians thought these passions lead us to desire only our own private good. The sentimentalists thought it their primary task to defend humanity’s benevolent sociability against those who saw in our passions only natural selfishness, and only secondarily to explain this sociability through appeal to the passions against those who sought its sources in reason.

Since the debate between sentimentalism and the selfish system was over the empirical realities of human psychology, it could only be resolved by empirical means. “If any person can in earnest doubt whether there is such a thing as goodwill in one man toward another,” Butler argues, “let it be observed that whether man be thus or otherwise constituted . . . is a mere question of fact or natural history, not provable immediately by reason” (5S 1.6, fn., pp. 27–28). The problem with appeal to the observable facts of human benevolence, however, is that these facts can always be reinterpreted so as to conform to the hypotheses of the selfish system. Hobbesians, Shaftesbury complains, “would so explain all the social passions and natural affections as to denominate them of the selfish kind” (SC 3.3, p. 74). Much of the writing of Shaftesbury, Hutcheson, and Butler is therefore devoted to refuting the many and varied attempts to explain away benevolent affections as nothing more than expressions of self-interest. To give just one example, Hobbes explains compassion for another’s misfortune as nothing more than fear that the same will occur to oneself.¹¹ Butler acknowledges that such fear is a real phenomenon, but it is not what we mean when we speak of compassion. Our experiences of pity or compassion are phenomenologically very different from our experiences of fear, and can occur when there is no possible danger to ourselves whatsoever (see 15S 5, pp. 64–66, fn). Hutcheson and Shaftesbury provide a significant number of other arguments to the same effect.¹²

Rather than cataloging the many arguments and counterarguments of this sort put forward by Hobbesians and sentimentalists respectively, it seems best to conclude this section with a single, powerful argument for why human nature cannot be understood as purely selfish. Although variations on this argument appear throughout Hutcheson’s writings,¹³ its clearest statement can be found in Butler’s sermon “Upon the Love of Our Neighbor.” Selfishness cannot be the only principle in the human heart, Butler argues,

because selfishness itself only makes sense as a wish to fulfill our various desires, each of which has an independent psychological reality of its own. "If self-love wholly engrosses us, and leaves no room for any other principle," Butler explains, "there can be absolutely no such thing at all as happiness, or enjoyment of any kind whatever since happiness consists in the gratification of particular passions, which supposes the giving of them" (5S 4.9, p. 48).

To be sure, it is always possible to treat everything we desire as a mere means to the single goal of personal happiness; Butler's argument, some appearances to the contrary, does not rule this out as a conceptual possibility.¹⁴ Yet Butler observes that it is a matter of psychological fact that doing so will only make us miserable, since it will rob us of the joy that can only come from pursuing at least some of the primary objects of our desire for their own sake. As has often been observed, the happiness that comes from sources such as love or friendship is impossible if one understands one's friends and loved ones as a means to one's own happiness. Such an attitude is incompatible with the attitudes toward others that love and friendship require. The happiness of love and friendship can only be experienced if we treat our intimates as possessing an independent, intrinsic value of their own. Butler's point is that what is so widely recognized as true of love and friendship is also true of many of the other primary objects of our desire, which can only make us happy if we do not treat them as mere means to our own satisfaction. As such, "it is certainly true that even from self-love we should endeavor to get over all inordinate regard to and consideration of ourselves" (5S 4.9, p. 49).

The good of others is one of the many things that we are capable of desiring for its own sake, as is implied by the very concept of benevolence. As Shaftesbury observes, "We never call that man benevolent who is in fact useful to others, but at the same time only intends his own interest, without any desire of, or delight in, the good of others" (IBV 2.2.3, p. 103). Butler's insight is that, far from making benevolent desires unique in some way, this is something they have in common with all the other primary desires that allow for human happiness. "That benevolence is distinct from, that is, not the same thing with self-love, is no reason for its being looked upon with any peculiar suspicion," Butler argues, "because every principle whatever, by means of which self-love is gratified, is distinct from it" (5S 4.11, pp. 49–50). As a result, "benevolence and the pursuit of public good has at least as great respect to self-love and the pursuit of private good as any other particular passions and their respective pursuits" (5S 4.16, p. 53). Indeed, upon introspection, most of us see that the benevolent desire for the good of others is one of the strongest, most basic passions implanted in our hearts. The only route to happiness is therefore to treat benevolence as leading us to pursue its object for its own sake.

3. *Reason and Reflection*

Even after striving so mightily to establish the reality of benevolent desires, the sentimentalists were still concerned that such desires might grow too strong, and imbalance the economy of the passions. “It must be owned,” Butler writes, “that every affection . . . may rise too high, and beyond its just proportion” (15S 6, p. 79). Sometimes, our benevolent desires themselves may give us reason to keep these very desires under control. For example, as Shaftesbury observes, pity or compassion is sometimes “so overcoming as to destroy its own end” (IVM 2.1.3, p. 51), since “excessive pity renders us incapable of giving succor” (IVM 1.2.2, p. 16). At other times, however, our reason for limiting our benevolent desires may come from other elements in our mental economy.

The faculty of reason cannot be excluded here. The Enlightenment sentimentalists never sought to prevent reason from playing a significant role in determining our actions, only to deny that it could do so alone. When interpreting the Biblical claim that the moral law is written in our hearts, Butler insists that this refers to “the whole system, as I may speak, of affections (including rationality) which constitute the heart, as this word is used in Scripture and on moral subjects” (5S 5.11, p. 60). In order to achieve its desired ends, benevolence must make use of reason in order to figure out the means to these ends. Hume was not the first to consider reason the “slave of the passions” in this way. Butler, too, saw that reason can be “considered merely as subservient to benevolence, as assisting to produce the greatest good” (5S 5.27, pp. 64–65). If this makes reason a slave of benevolent passions, however, it is in the same way that ancient tutors were often the slaves of their charges. Lest benevolence defeat its own purposes, Butler writes, “reason must come in as its guide and director in order to attain its own end” (5S 5.27, p. 65). When, as it so often does, reason fails in this capacity, terrible harm to others can be the result. According to Hutcheson, this is why we so often call moral actions rational, and immoral actions irrational. We naturally assume all others to share our benevolent desires, and ascribe immoral actions to a failure to reason appropriately about the proper means to their benevolent ends, rather than to a rejection of the ends themselves (see ENC 2.1, p. 148). Hutcheson complains that this habit of mind then often leads into the use of the word “reasonable” in a “confused manner” to “denote whatever is approved by our moral sense” (ENC 2.1, p. 154). This is not to claim, however, that rationality and moral virtue are not strongly correlated. “[I]t is probable no person would ever do anything publicly hurtful but upon some false opinion,” Hutcheson writes. “The flowing from true opinions is indeed a tolerable character or property of virtue, and flowing from false opinion a tolerable character of vice, though neither be strictly universal” (ENC 2.3, p. 169).

Even though reason has an important role to play in keeping our passions in balance, what the Enlightenment sentimentalists called “reflection” has an even more central role. Although reason and reflection are often used in conjunction or as loose synonyms, the latter is better understood as a matter of self-observation than of reasoning *per se*. As it was used in the eighteenth century, the concept of reflection stems primarily from Locke, who famously believed all knowledge to be the result of experience. Experience, however, may take the form of either observation of the external world through the powers of “sensation” or observation of the operations of our own minds through the powers of “reflection.”¹⁵ Shaftesbury may have largely rejected his tutor’s epistemology—not so much as incorrect, but as useless—but he maintained Locke’s emphasis on the importance of reflection, often poetically calling it “soliloquy.” Unlike Locke, however, Shaftesbury placed particular importance on reflective scrutiny of our passions and affections. “In a creature capable of forming general notions of things,” he writes, “not only the outward beings which offer themselves to the sense are the objects of the affection, but the very actions themselves and the affections of pity, kindness, gratitude and their contraries, being brought into the mind by reflection, become objects.” Once our emotions become objects of reflection, moral sentiments arise as a form of reflective affection or disaffection for both our own affections and those of others. The fact that we have such reflective sentiments allows our benevolence to qualify as virtuous rather than as merely good. As Shaftesbury explains:

Let us suppose a creature who wanting reason, and being unable to reflect, has notwithstanding many good qualities and affections, as love to his kind, courage, gratitude or pity. It is certain that if you give to this creature a reflecting faculty, it will at the same instant approve of gratitude, kindness and pity, be taken with any show or representation of social passion, and think nothing more amiable than this, or more odious than the contrary. And this is to be capable of virtue, and to have a sense of right and wrong.

(IVM 1.3.3, p. 31)

Shaftesbury’s conception of moral sentiments as reflective, second-order affections for benevolent, first-order affections was then adopted and adapted by later sentimentalists in their own distinctive ways. Butler calls this reflective power “conscience.” Interpreting Romans 2:15, he distinguishes “the witness of conscience” from the law written in the gentiles’ hearts. The law of the heart is simply “the natural disposition to kindness and compassion . . . that part of the nature of man . . . which with very little reflection and of course leads him to society.” Conscience, by contrast, “is a superior principle of reflection . . . in every man which distinguishes between the internal principles of his heart as well as his external actions, which passes judgment upon himself and them . . . It is by this faculty, natural to man, that he is a moral

agent, that he is a law to himself” (5S 2.8, p. 37). While Butler insists on the importance and (as will be discussed later) the authority of this reflective faculty of conscience, he is deliberately vague as to its precise nature and mode of operation. Conscience seems to draw on all the other faculties of the human mind, although Butler never specifies precisely how. He merely asserts that it can be “considered as a sentiment of the understanding or as a perception of the heart or, which seems the truth, as including both” (5S Dissertation 1, p. 69).

Hutcheson, by contrast, gives a very particular account of the nature and operations of this reflective moral faculty, calling it “the moral sense.” While other sentimentalists occasionally use the phrase, Hutcheson was the thinker who took the idea of a moral sense most literally. In addition to the well-known five powers of Lockean sensation that give us access to the external world—seeing, hearing, smell, taste, and touch—Hutcheson insisted that we are also in possession of a number of as-yet unrecognized internal senses, powers of Lockean reflection through which we observe our own sensations and evaluate their objects. Like the powers of sensation, Hutcheson insists that the powers of reflection are literally senses. They are innate in their origin and automatic in their operation in exactly the same way as the external senses of taste or smell. In his initial presentation of this thesis, the 1725 *Inquiry into the Original of Our Ideas of Beauty and Virtue*, Hutcheson begins by arguing for an internal sense of beauty or aesthetic value. “This superior power of perception is justly called a sense,” he argues, “because of its affinity to other senses in this, that the pleasure does not arise from any knowledge . . . but strikes us at the first with the idea of beauty” (IBV 1.1.13, p. 25). Then, this sixth sense being established, he goes on to argue that “it will be no difficult matter to apprehend another superior sense, natural also to men, determining them to be pleased with actions, characters, affections” (IBV Preface, p. 10). This seventh sense is the moral sense, which according to Hutcheson, automatically approves of benevolent desires and disapproves of malicious ones.

While others might attempt to explain our reflective approval of benevolence on the basis of more basic principles of human psychology, Hutcheson insists that no further explanation of the moral sense can be given, at least none beyond the fact that it was implanted in us by God for the achievement of his own benevolent purposes. Hutcheson admits that his critics might find this moral sense an “occult quality” (IBV 2.7.3, p. 178) appeal to which is no more useful explanatorily than Molière’s satirical appeal to opium’s “dormative power” to describe how the drug puts us to sleep. Hume and Smith were both to reject as “absurd” (T 3.1.2.6) the non-explanation of the moral sense put forward by “those who resolve the sense of morals into original instincts of the human mind” (T 3.3.6.3). They instead insist that the etiology of our moral sentiments can surely be traced to more basic psychological phenomena,

sympathy foremost among them. Hume and Smith criticize Hutcheson's failure to do so on the grounds of parsimony. "Nature . . . acts here, as in all other cases, with the strictest economy," Smith writes, "and produces a multitude of effects from one and the same cause" (TMS VII.iii.3.3, p. 321).¹⁶ If there is anything in moral evaluation that cannot be traced to more basic psychological principles, Smith acknowledges, "I would be glad to know what remains, and I shall freely allow this overplus to be ascribed to a moral sense, or to any other peculiar faculty, provided anybody will ascertain precisely what this overplus is" (VII.iii.3.16, p. 326).

In his later work, Hutcheson responds that only the moral sense can explain the automaticity of much moral judgment, a phenomenon that is also responsible for the revival of the idea of an inborn, specialized moral faculty by twenty-first-century psychologists.¹⁷ Hutcheson claims nothing but a natural moral sense could "account for that immediate ardor of love and goodwill which breaks forth toward any character represented to us as eminent in moral excellence, before we have had any thoughts, or made any inquiries into his state" (SMP 1.3.5, Vol. I, p. 48). To this, one can respond with Hume's observation that, as with many mental phenomena, "however instantaneous" moral judgment may be "it proceeds from certain views and reflections which will not escape the strict scrutiny of a philosopher, though they may the person himself who makes them" (T 2.1.11.3). Arguments of the sort that Hume and Smith leveled against Hutcheson have also been used in recent years to argue against the revival of his conception of an innate moral sense.¹⁸ In order to emphasize their differences with Hutcheson, Hume and Smith used the phrase "moral sense" less often than their predecessors did, mostly speaking instead of "moral sentiments."

II. Religious and Metaphysical Foundations

1. *The Happiness of Virtue*

Unlike empirical psychologists today, Enlightenment sentimentalists moved quickly from descriptive to normative questions. Once we have properly described human moral psychology, Shaftesbury insists "it remains to inquire what obligation there is to virtue, or what reason to embrace it" (IVM 2.1.1, p. 45). As this quotation indicates, the word "obligation" had a broader meaning in the eighteenth century than it does today. An obligation to perform an action simply meant an overriding reason to do it, one that ruled out the possibility of omitting it. This determinative reason could be self-interested as easily as it could be disinterested or moral.¹⁹ Shaftesbury was particularly interested in what was then known as an "interested obligation" to virtue. "If by obligation we understand a motive from self-interest sufficient to

determine all those who duly consider it, and pursue their own advantage wisely to a certain course of actions," he writes, "we may have a sense of such an obligation . . . by considering how much superior we esteem the happiness of virtue to any other enjoyment" (IBV 2.7.1, p. 177).

The sources of the happiness of benevolence and virtue are, for Shaftesbury, twofold. First are the joys that derive from the benevolent affections themselves. These pleasures are a key feature of virtually everything we enjoy, which lose their savor if we cannot share them with others. "So insinuating are these pleasures of sympathy," Shaftesbury writes, "and so widely diffused through our whole lives, that there is hardly such a thing as satisfaction or contentment of which they make not an essential part" (IVM 2.2.1, p. 62). Even more important, benevolent pleasures are among the only ones that we can enjoy "freely and without reserve" without upsetting the larger balance of our affections. "I know no other consequence from indulging such a passion than that of growing better natured and enjoying more and more the pleasures of society," Shaftesbury writes (SAA 3.2, pp. 191–92).

Alongside the joys that stem directly from our benevolent affections are those that come from reflectively approving of them in ourselves. "Every reasoning or reflecting creature is, by his nature, forced to endure the review of his own mind and actions," Shaftesbury writes (IVM 2.2.1, p. 69). For this reason, human happiness is only possible when "together with the most delightful affection of the soul there is joined a pleasing assent and approbation of the mind to what is acted in this good disposition and honest bent" (IVM 2.2.1, p. 61). Our goal is "a mind . . . well composed, quiet, easy within itself and such as can freely bear its own inspection and review" (IVM 2.2.1, p. 66).

Given this emphasis on the happiness of a virtuous life, it may be surprising that Shaftesbury showed so much hostility toward existing arguments that virtue is in our self-interest. In addition to Hobbesians who argued that justice and morality are artificially constructed to serve the purposes of self-interested agents, Shaftesbury also excoriated theologians who saw morality as fundamentally grounded in divine reward and punishment. He thought such Christian moralists, including his own tutor Locke, "have made virtue so mercenary a thing, and have talked so much of its rewards, that one can hardly tell what there is in it, after all, which can be worth rewarding . . . For to be bribed only or terrified into an honest practice bespeaks little of real honesty or worth" (SC 2.3, p. 61). While Shaftesbury himself, like his intellectual opponents, believed virtue to be the path to happiness, he was insistent that this consideration was always far from the mind of the virtuous themselves—so long, at least, as their minds had not been corrupted by philosophy or religion. "A common honest man, while left to himself, and undisturbed by philosophy and subtle reasonings about his interest," Shaftesbury writes, "gives no other answer to the thought of villainy than that he can't

possibly find in his heart to set about it, or conquer the natural aversion he has to it" (SC 4.1, pp. 82–83). If Butler was correct in the argument against the selfish system outlined earlier, the fact that the virtuous do not see virtue as a mere means to happiness is what allows it to contribute so significantly to their happiness in the first place.

The problem with this view is that it seems to preclude the very sort of philosophical reflection that Shaftesbury clearly thought was essential to achieving human perfection. Even if the thought had never occurred to them before, Shaftesbury's readers quickly learn that the best way to promote our private interest is through virtuous, benevolent affections. Once in possession of this knowledge, it is difficult to see how we can continue to pursue the good of others without seeing their good, on some level, as a means to our own. Under Shaftesbury's view, this would destroy the moral worth of our benevolent affections. Under Butler's, it might even destroy the happiness that they could otherwise provide.

Hutcheson offers a rather elegant route out of this bind, based on the fact that "neither benevolence nor any other affection or desire can be directly raised by volition" (1738 Addition IBV 2.2.3, p. 220). Instead, they are "determinations of our nature, previous to our choice from interest, which excite us to action, as soon as we know other sensitive or rational beings, and have any apprehension of their happiness or misery" (ENC 1.4.2, p. 67). If an imaginary reflective creature without benevolent affections were to decide that only by gaining them could it achieve happiness, its attempt to do so could never succeed. As for real-world human beings who already possess such affections, learning that the possession of these affections is the key to our happiness cannot change the fact that they never were under our control and never were the kind of faculties that we could willfully deploy as means to some further end. To be sure, it is "of the highest importance to all to strengthen [these affections] as much as possible, by frequent meditation and reflection" (ENC 1.6.1, p. 111), but Hutcheson expressed real doubts about the extent to which this is actually within our power.

Shaftesbury, however, took a different view. Although it is certainly true that we cannot directly decide to feel any particular benevolent affection at any particular moment, it is nonetheless the case that, upon reflection, we may resolve to undergo the "reforming work" of "introducing into the affectionate part [of the mind] some gentle feeling of the social and friendly kind" (IVM 2.2.1, p. 67). Not only are we capable, over time, of introducing first-order benevolent affections into our soul; we are also capable of revising our reflective moral sentiments. "We may esteem and value, approve and disapprove, as we would wish," Shaftesbury insists (MRT 3.2, p. 114). The fact that we can control these affections and sentiments is very important to Shaftesbury, since he accepts the Stoic dictum that we must not bother ourselves with things that are not within our power, and seek happiness only from that which we

control (See MRT 4.1, p. 122). Since it is in our power to achieve the happiness of virtue, and yet we must not pursue virtue for the sake of this happiness, Shaftesbury believes that we must also have some other, unselfish reason to pursue virtue. For Shaftesbury, this “disinterested obligation” was provided by natural teleology.

2. *The Standard of Nature*

Though he condemns those mercenary types who pursue virtue for the sake of happiness, Shaftesbury repeatedly praises those who, in developing their character, “deliberately endeavored to frame it by the just standard of nature” (SAA 3.3, p. 218). Here again, the Stoics and other ancient ethicists were Shaftesbury’s inspiration. Hutcheson observes that, like Plato and Aristotle before them, “the Stoics define virtue to be an agreement or harmony with ‘nature’ in our affections and actions” (MEM 4.29, fn., p. 53). For Shaftesbury, as for his ancient predecessors, the universe is a well-ordered, purposive whole made up of well-ordered, purposive parts designed to work harmoniously together. In order to understand any phenomenon, natural or artificial, one must understand the purpose, or final cause, of its design. Shaftesbury here utilizes one of the watchmaker analogies so popular in the eighteenth century:

If a passenger should turn by chance into a watchmaker’s shop, and thinking to inform himself concerning watches, should inquire of what metal, or what matter, each part was composed . . . without examining what the real use was of such an instrument, or by what movements its end was best attained and its perfection acquired, it is plain that such an examiner as this would come short of any understanding in the real nature of the instrument.

(SAA 3.1, p. 181)

The explanation of any phenomenon remains incomplete without determining its purpose or final cause, moral and psychological phenomena included. Unsurprisingly, given both his insistence on the reality of benevolence and his conception of the universe as a harmonious whole, Shaftesbury believes the study of our natural affections and reflective sentiments reveals that their purpose is to lead us to pursue the good of others, and of the larger cosmic system of which we are all a part. Individual human beings, he concludes, exist not for their own sake, but for the sake of the good of the whole. Only by pursuing this naturally given end do we have any hope of achieving happiness, but our reason for doing so is the fact that it is the vocation allotted to us by our place in a purposive universe. Personal happiness is only a byproduct of pursuing our assigned task for its own sake.

The ancient Stoics reached similar conclusions without making any real use of our benevolent affections, appealing simply to the obvious subservience

of parts to wholes. The result was the Stoic commitment to natural law as discoverable on the basis of rational inquiry into the purposive nature of the universe, a commitment revived by the rationalist natural law theorists of the early modern era. Since Shaftesbury's own inquiry on the matter is little different, his ethics arguably also rests on rationalist natural law foundations. Eighteenth-century British rationalists such as Richard Price claimed Shaftesbury as one of their own (by virtue of his appeal to cosmic teleology) as surely as did sentimentalists (by virtue of his appeal to the economy of the passions).²⁰ Hume, for one, admitted that, although Shaftesbury was basically a sentimentalist, the "elegant" Lord was not "entirely free from . . . confusion" on these foundational questions (EPM 1.4).

The reliance on natural teleology plays a similar role in the work of Butler and Hutcheson. All three operated under a fundamentally premodern conception of scientific explanations as requiring appeal to final as well as to efficient causes, to the purposes of nature or its designer rather than simply to the non-purposive forces responsible for bringing a given phenomenon into being. The eighteenth century was a transitional period in the history of science, one in which Hutcheson could simultaneously maintain that "only an efficient cause is properly called a cause" (SM 1.4.5, p. 93) and that "we should not exclude the so-called final causes from physics" (SM 3.5.3, p. 183). If final causes are rightly part of natural philosophy or science, they are even more important in moral philosophy. According to Butler, "observations of final causes" in this field "show us what course of life we are made for, what is our duty, and in a peculiar manner enforce upon us the practice of it" (15S 6, p. 74).

If Shaftesbury's commitment to natural teleology brings him close to rationalist natural law theory, Hutcheson's brings him close to divine command theory. "All such as believe that this universe, and human nature in particular, was formed by the wisdom and counsel of a Deity," Hutcheson writes, "must expect to find in our structure and frame some clear evidences showing the proper business for mankind, for what course of life, what offices we are furnished by the providence and wisdom of our Creator" (PMIC I.1.1, p. 24). The main justification for empirically examining human psychology, rather than directly appealing to the intentions of the deity, is simply epistemological. "In this art, as in all others, we must proceed from the subjects more easily known, to those that are more obscure, and not follow the priority of nature," Hutcheson writes. We therefore should not "deduce our first notions of duty from the divine will, but from the constitution of our nature, which is more immediately known, that from the full knowledge of it we may discover the design, intention and will of our Creator as to our conduct" (PMIC I.1.1, p. 24).

Hume objects to this conception of nature as a harmonious and purposive whole, whether a result of divine design or immanent teleology. In the *Treatise*, Hume complains that while Shaftesbury was a "great genius," he nonetheless

fell victim to “vulgar” prejudices “concerning the uniting principle of the universe” (T 1.4.6.6, fn.). He insists in a letter to Hutcheson that ethics is best carried out without reliance on a conception of nature “founded on final causes, which is a consideration that appears to me pretty uncertain and unphilosophical.” Disputes on the natural vocation of humanity and our intended role in a purposive universe are, in Hume’s view, “endless, and quite wide of my purpose.”²¹

Yet while Hume’s version of sentimentalism has greater appeal to secular moral and political philosophers today, the fact that the theories of Hutcheson and Butler rest on unambiguously religious grounds should neither be held against them nor against Enlightenment sentimentalism more generally. To the contrary, if sentimentalism is to provide a widely shared basis for our moral and political commitments in religiously and philosophically pluralistic societies, then it must be presented in a way attractive to those who insist that their convictions have an ultimately religious or metaphysical foundation.²² Yet if sentimentalism is also to be attractive to those who reject such religious and metaphysical views, it must also be capable of being presented in a free-standing fashion, without appeal to a divine creator or a grand cosmic whole. It took the religiously and metaphysically skeptical Hume to work out this free-standing theory.

3. *The Authority of Conscience*

Like Shaftesbury, Butler made frequent appeal throughout his ethics to final causes, but he placed far greater emphasis on the special authority of conscience than he did on the vocation of humanity as such. Like Shaftesbury’s idea of cosmic harmony, Butler’s idea that a part of the mind has legitimate authority over the rest is inspired by Plato, Aristotle, and the Stoics. In Hutcheson’s translation, Marcus Aurelius says that the rational soul claims authority over the rest of the psyche “very justly, as by its own nature fitted to command, and employ all these lower powers” (MEM 7.55, p. 90). Indeed, as the introduction made clear, such a hierarchical view of the proper psychic regime can be understood as the distinguishing feature of moral rationalism. In attempting to reintroduce an element of psychic hierarchy into sentimentalism, however, Butler does not wholly abandon what I have described as its commitment to psychic democracy. For Butler, it is not an isolated faculty such as reason alone that is capable of performing authoritative moral legislation, but rather a psychologically holistic reflective power—incorporating both reason and sentiment in their capacity to observe and judge their own operations—which we earlier saw that he labeled “conscience.”

“Conscience or reflection,” Butler claims, “plainly bears its marks of authority” because it wields judgment over all the other powers of the mind; it “claims the absolute direction of them all, to allow or forbid their gratification.”

He argues that it is in the nature of our second-order reflections to overrule our first-order affections, “a disapprobation of reflection being in itself a principle manifestly superior to a mere propension” (5S Preface 24–25, pp. 16–17). This is not to say that, should conscience condemn one of our first-order affections, it will always prove the stronger psychological principle, capable of extirpating all challenges to its judgments. Even when it fails to do so, however, conscience has right on its side. The difference between first and second order mental powers, “not being a difference in strength or degree,” is, for Butler, “a difference in nature and kind” (5S 2:11, p. 38), the difference “which everybody is acquainted with between mere power and authority” (5S 2:14, p. 39). This is a normative moral distinction, not a descriptive psychological one, and Butler explains it by analogy with political morality. “As in civil government the constitution is broken in upon and violated by power and strength prevailing over authority,” he writes, “so the constitution of man is broken in upon and violated by the lower faculties or principles within prevailing over that which is in its nature supreme over them all” (5S 3.2, p. 41; 2.13, p. 39 and Preface 24–25).

Butler criticizes his sentimentalist predecessors for their failure to develop this idea. “The not taking into consideration this authority, which is implied in the idea of reflex approbation or disapprobation,” he writes, “seems a material deficiency or omission in Lord Shaftesbury’s inquiry concerning virtue” (5S Preface 26, p. 17). It is not that Shaftesbury fails to recognize the moral importance of reflective self-evaluation. As has already been demonstrated, self-examination or “soliloquy” is in fact central to Shaftesbury’s philosophy. Shaftesbury even foreshadows Butler to the extent that he agrees that reflective self-evaluation is “properly called conscience” (IVM 2:2:1, p. 69). But it is true that Shaftesbury did not think that conscience carried any special authority. For him, the importance of conscience stems not from the fact that this faculty has a right to rule the soul, but rather from the fact that reflective self-approbation is a necessary element of human happiness. And in this, as Hume would later argue, it was Butler, and not Shaftesbury, whose sentimentalist ethics suffered from a serious deficiency.

Hume expressed his objections to Butler’s conception of the authority of conscience in a letter to Hutcheson. Early in his career, Hutcheson seemed to adopt something like Shaftesbury’s view on the matter. The younger Hutcheson acknowledged that much of the happiness of virtue stems from reflective self-approbation, from our tendency “to be pleased and happy when we reflect upon our having done virtuous actions, and to be uneasy when we are conscious of having acted otherwise” (IBV 2:7:1, p. 177). In his early writings, however, there is little indication that Hutcheson believed conscience or the moral sense to carry a special authority independent of its necessary contribution to human happiness. As Butler’s influence on Hutcheson’s philosophy increased, however, this position began to change.

In his 1730 inaugural lecture as professor of moral philosophy at Glasgow, Hutcheson adopts Butler's view of conscience (which he identifies with the moral sense described in his own earlier works) as "*to hegemonikon*," Greek for "the ruling principle," the single faculty "to which all things were made subject, and rightly so, in the integral state of our nature" (NSM, p. 199). By the time of his 1742 *Philosophiae Moralis Institutio Compendiaria* (later translated as *A Short Introduction to Moral Philosophy*), the fact that our moral sense can judge all the other powers of the mind takes on central importance as the mark of its authority. "This nobler sense," Hutcheson writes, "is plainly the judge of the whole of life, of all the various powers, affections and designs, and naturally assumes a jurisdiction over them . . . That this divine sense or conscience . . . should be the governing power in man appears . . . immediately from its own nature" (PMIC 1.1.12, p. 40).

Upon reviewing Hutcheson's *Institutio*, Hume complains:

You seem here to embrace Dr Butler's opinion in his sermons on human nature, that our moral sense has an authority distinct from its force and durableness, and that because we always think it *ought* to prevail. But this is nothing but an instinct or principle, which approves of itself upon reflection, and that is common to all of them.²³

Hume knew that this argument against Butler would have a considerable hold on Hutcheson. Far from arguing that conscience or the moral sense derives a special authority from its reflective self-approbation, Hutcheson had earlier denied that the moral sense is even capable of evaluating itself. Although our moral sense can approve or disapprove of our actions and emotions when we reflect upon them, "none can apply moral attributes to the very faculty of perceiving moral qualities, or call his moral sense morally good or evil, any more than he calls the power of tasting sweet or bitter, or of seeing straight or crooked, white or black" (ENC 2.1, p. 149). Of course, there is an obvious problem with this analogy. While our sense of sight allows us to evaluate visible objects, and our sense of taste allows us to evaluate edible ones, the former sense is no more itself a visible object than the latter is an edible one. By contrast, our moral sense allows us to evaluate human sentiments and actions. Since these evaluations are themselves human sentiments, they are capable of being evaluated morally. Any sense or faculty for which human performances are part of its evaluative domain can judge its own actions similarly. Hume therefore argues that some, if not all, of our individual faculties are indeed capable of evaluating themselves, our moral faculties among them. Butler, and later Hutcheson, believed that our conscience or moral sense wields a special authority because we have a sense that it ought to do so. Since the claim that conscience has right on its side in its conflicts with the other powers of the mind is itself a moral judgment, Hume argues that it is a good example of a human faculty judging itself.

Hume then goes on to observe that, whenever a faculty is capable of such self-evaluation, the verdicts it returns will almost certainly be positive, which should come as no surprise when it is serving as a judge in its own case. The mere fact that they approve of themselves thus cannot be held to establish their authority over the rest of the mind. As a teenager, my self-approving yet sophomoric sense of humor often led me to laugh at my own asinine jokes, but this was hardly grounds for treating its determinations as authoritative. To be sure, our intellectual faculties can and should turn on themselves and judge whether the beliefs to which they lead us are true or warranted—as our moral faculties can reflect on themselves and judge whether their evaluations are good or praiseworthy. Yet why is it important, for example, for our aesthetic faculties to reflect back on themselves and judge whether their tastes are sublime or beautiful? More generally, why is a mental faculty's ability, taken in isolation, to approve of itself grounds for treating the determinations of that faculty as normatively authoritative in the first place?²⁴

Butler and Hutcheson, of course, can give a religious answer to these questions: a faculty's ability to approve of itself can be taken as a sign that it was intended by its divine designer to wield authority in its relevant domain. This is precisely the view of conscience's authority in a (possibly apocryphal) fragment from Epictetus, which Hutcheson uses as one of the epigraphs for the *Insitutio*: "God has committed men to the government of their own natural conscience. This governor we never should disobey, for it is offensive to God" (PMIC, p. 6). Perhaps this is why the heretical Hume's arguments failed to convince Hutcheson to abandon his move away from Shaftesbury toward Butler with regard to the special authority of conscience. To the contrary, Butler's influence on this point is all the more obvious in Hutcheson's last work, the *System of Moral Philosophy*, as are the unambiguous appeals to the authority of God. As Hume's arguments demonstrate, however, it is God's authority that is actually doing the normative work in Hutcheson and Butler's arguments regarding conscience, and not the faculty's authority in itself. As has already been discussed, Hume has good reasons for rejecting all such appeals to divine intensions.

Recognizing that Hume rejects the idea of a faculty's self-approval as a source of normative authority is particularly important given that prominent commentators have attributed something resembling Butler's view to Hume. Christine Korsgaard and Annette Baier maintain that Hume puts forward a theory of "normativity as direct reflexivity" under which "a faculty's verdicts are normative if the faculty meets the following test: when the faculty takes itself and its own operations for its object, it gives a positive verdict."²⁵ A fuller discussion of Hume's actual account of normativity will have to wait until chapter 2, where I will argue that Hume is closer to Shaftesbury on this matter than to Butler, albeit with an account of the happiness of virtue relying neither on the teleological metaphysics of the former nor the theology of the latter.

III. Theories of Justice

1. *Justice as Benevolence*

Given their opposition to Hobbes, it should come as no surprise that sentimentalists reject the idea that a social contract can determine principles of justice, although it may play a role in the first formation of government.²⁶ “It is ridiculous to say there is any obligation on man to act sociably, or honestly, in a formed government and not in that which is commonly called the state of nature,” Shaftesbury writes. Far from being able to create principles of justice *ex nihilo*, he claims a social contract assumes principles of justice are already operative:

[T]hat which could make a promise obligatory in the state of nature must make all other acts of humanity as much our real duty and natural part. Thus faith, justice, honesty and virtue must have been as early as the state of nature, or they could never have been at all . . . He who was free to any villainy before his contract will and ought to make as free with his contract when he thinks fit. The natural knave has the same reason to be a civil one, and may dispense with his politic capacity as oft as he sees occasion.

(SC 3:1, pp. 68–69)²⁷

By rejecting the artificiality of civil justice, sentimentalism represents a return to the ancient view of human beings as political animals. “Our political and social capacity,” Shaftesbury claims, is as “natural and essential in our species as the parental and filial kind” (MRT 3:1, p. 190). It also allies the sentimentalists with the natural law tradition, which sought to formulate universal principles according to which any given system of positive law can be judged as just or unjust. In contrast to rationalist natural lawyers, however, sentimentalists could not see legal principles as built into the universe independent of human psychology, waiting to be discovered through the rational deductions of moral metaphysicians. For sentimentalists, as Hutcheson put it, “the laws of nature are inferences we make by reflecting upon our inward constitution, and by reasoning upon human affairs, concerning that conduct which our hearts naturally must approve” (SMP 2.17.2, Vol. II, p. 119).

Of the three authors under discussion in this chapter, only Hutcheson was to develop a full system of natural jurisprudence. Although it would be going too far to describe the philosophies of Shaftesbury or Butler as wholly apolitical, they did lack Hutcheson’s emphasis on the importance of legal justice. Shaftesbury, in particular, assumes that, if only individuals achieved a benevolent balance of passions in their souls, a just polity would surely follow. “A public spirit can come only from a social feeling or sense of partnership with human kind,” he writes, “And thus morality and good government go together” (SC 3.1, p. 67). Although Hutcheson was to do much more than Shaftesbury to derive principles of natural law from our moral sentiments,

he maintained Shaftesbury's view of a smooth continuity between ethics and politics, seeing political justice as a direct outgrowth of personal benevolence.

According to Hutcheson, benevolence can determine natural jurisprudence in two distinct ways. Some principles of justice stem from our benevolence for particular individuals. "[B]y our natural sense of right and wrong, and our sympathy with others, we immediately approve any persons procuring to himself or his friends any advantages which are not hurtful to others, without any thought either about a law or the general interest of all," Hutcheson writes. "Hence everyone is conceived to have a right to act or claim whatever does no hurt to others, and naturally tends to his own advantage, or to that of persons dear to him" (PMIC 2.2.1, pp. 111–12). Such is the origin, for example, of individual property rights; Hutcheson insists that we need not "have recourse to any old conventions of all men to explain the introduction of property" (PMIC 2.6.4, p. 142). He instead points to the fact that allowing others' to enjoy the fruits of their labor is clearly an example of benevolence to them, while interfering with this enjoyment is the opposite. As a result, our moral sense cannot help but approve of the former and disapprove of the latter. "Thus therefore we should judge," Hutcheson concludes, "that it is inhuman and unjust, without the most urgent necessity, to obstruct the innocent labors others have begun and persist in" (PMIC 2.6.2, p. 138).

At the same time, however, we also have a benevolent concern for the welfare of all our fellow human beings in the aggregate. Benevolence thus leads us, as Hutcheson so famously put it, to judge "that action is best which procures the greatest happiness for the greatest numbers" (IBV 2:3:8, p. 125). Hutcheson maintains that, when our benevolence toward one comes into conflict with our benevolence toward all, it is the latter that ought to hold sway. "Our moral sense, though it approves all particular kind affection or passion," he writes, "also approves the restraint or limitation of all particular affections or passions by the calm universal benevolence. To make this desire prevalent above all particular affections is the only sure way to obtain constant self-approbation" (ENC 1.2.2, p. 33). As a result, principles of justice derived from benevolence toward all take precedence over principles derived from benevolence toward particular individuals or groups. It is the role of reason "to discern what actions really tend to the public good in the whole, that we may not do that upon a partial view of good which afterwards, upon a fuller examination, we shall condemn and abhor ourselves for" (ENC 1:4:4, p. 76).

In this respect, Hutcheson is an obvious forefather of what later became known as utilitarianism. One important difference between Hutcheson and the classical utilitarians of the following century, however, is that Hutcheson rejected what is now known as consequentialism. He insisted that "moral good and evil consists not in the external events, but in the affections and purposes of the soul" (PMIC 2.3.7, p. 125). All benevolent affections are

good, regardless of the consequences of following them, but some are better than others. The “public sense” of benevolence toward all is to be preferred to all private benevolence toward individuals or groups, not because of the greater good produced by it, but because our moral sense prefers it as the more laudable affection, as expressing a greater degree of benevolence. Political justice takes precedence over private acts of kindness or mercy for this reason; our moral sense informs us that the proper economy of our passions and affections must give priority of the former over the latter.

At the same time, however, we do not show the proper attitude of extensive, public benevolence if we do not take great care to consider all the possible consequences of our actions. Consider the case of someone considering violating an established law in a way that would, at least in the short run, do far more good than harm. Under even an ideal code of law, such a scenario is certain to arise with some frequency. “[M]any laws prohibit actions in general even when some particular instances of those actions would be very useful,” Hutcheson admits, “because an universal allowance of them, considering the mistakes men would probably fall into, would be more pernicious than an universal prohibition, nor could there be any more special boundaries fixed between the right and wrong cases” (IBV 2:3:10, p. 126). When we consider violating such a law, Hutcheson insists that “we should consider all the consequences, even of a remoter kind, which must ensue upon diminishing the deep reverence men should have for these laws” (PMIC 2.16.2, p. 207). In most cases, upon doing so we will find that “it is the duty of persons to comply with the generally useful constitution.” If disobedience is indeed required, we “must patiently resolve to undergo those penalties which the state has, for valuable ends to the whole, appointed, and this disobedience will have nothing criminal in it” (IBV 2:3:10, p. 126).

When he acknowledges the reality of cases in which we must obey a law in which this obedience is nonetheless of no direct benefit to any object of our benevolence, Hutcheson is also forced to acknowledge the existence of a very strange category of individual rights in such cases. Such a right is “but rather a shadow of right than any thing deserving that honorable name . . . in the use of which no man can be approved by God, or his own heart upon reflection.” Hutcheson calls these “external rights,” and says they arise “when doing, enjoying or demanding from others is really detrimental to the public, and contrary to the sacred obligations of humanity, gratitude, friendship or such like, and yet for some remote reasons it is for the interest of society not to deny men this faculty, but on the contrary in some instances to confirm it” (PMIC 2.2.2, p. 112). The benevolent and virtuous would never insist on these external rights, but should the selfish and vicious do so, the virtuous are prevented by their own benevolence from doing anything but respecting this immoral decision. Hutcheson gives the example of a miser’s right to hoard wealth, or even to demand repayment from a virtuous debtor.

As will be seen in chapter 3, Hume uses examples along roughly the lines of Hutcheson's "external rights" to insist that justice cannot simply be a matter of natural, benevolent sentiments, but must also be a matter of artificial conventions. In his letter on the *Institutio*, Hume chides Hutcheson for being too "afraid to derive any thing of virtue from artifice or human conventions"²⁸ Although, for most sentimentalists, Mandeville was a *bête noire*, Hume provocatively includes him among the "late philosophers in England who have begun to put the science of man on a new footing" (T Intro.7) because of his insights into the role of artifice, convention, and self-interest in our commitment rules of justice so strict as to require repaying a debt to even the most vicious miser. Hutcheson, however, concludes that as long as external rights are indeed socially useful there is nothing artificial or selfish about our obligation to respect them. "[W]hatever appears necessary for preserving an amicable society among men," he insists, "must necessarily be enjoined by the law of nature" (PMIC 2.2.2, p. 112).

2. Injustice and Resentment

Hume was not the only sentimentalist who was dissatisfied with Hutcheson's theory of justice as benevolence, or of virtue as benevolence more generally. Butler, for one, maintained "benevolence and the want of it, singly considered, are in no sort the whole of virtue and vice." Butler's arguments on this point foreshadow many of the most convincing rebuttals to classical utilitarianism, most importantly those built around the fact that while utilitarians are deeply concerned about maximizing aggregate welfare, they do not take the distribution of this welfare into account. If benevolence were all that mattered morally, Butler observes, we "would be indifferent to everything but the degrees in which benevolence prevailed, and the degrees in which it was wanting. That is, we should neither approve of benevolence to some persons rather than to others, nor disapprove injustice and falsehood upon any other account" (5S Dissertation 8, p. 73). Yet nothing could be farther from the actual judgments of our conscience or moral sense. As Butler writes:

[S]uppose one man should, by fraud or violence, take from another the fruit of his labor, with intent to give it to a third who he thought would have as much pleasure from it as would balance the pleasure which the first possessor would have had in the enjoyment, and his vexation in the loss of it; suppose also that no bad consequences would follow, yet such an action would surely be vicious. Nay further, were treachery, violence and injustice not otherwise vicious than as foreseen likely to produce an overbalance of misery to society then, if in any case a man could procure himself a great advantage by an act of injustice as the whole foreseen inconvenience likely to be brought upon others by it would amount to, such a piece of injustice would not be faulty or vicious at all, because it would be

no more than, in any other case, for a man to prefer his own satisfaction to another's in equal degrees.

(5S Dissertation 8, p. 73)

In other words, Butler recognizes justice is more a matter of distributional fairness among individuals than it is a matter of universal benevolence for all indiscriminately. In order to explain and justify this sense of fairness, sentimentalists must look first for its origins in the fabric of human psychology. They must explain why it is, in Butler's words, "that we are constituted so as to condemn falsehood, unprovoked violence, injustice and to approve of benevolence to some preferably to others, abstracted from all consideration which conduct is likeliest to produce an overbalance of happiness or misery" (5S Dissertation 8, p. 73). They must identify the passion or affections responsible for our sense of fairness and demonstrate how "every man carries about him this passion which affords him demonstration that the rules of justice and equity are to be the guide of his actions" (15S 9, p. 100).

There is a powerful, yet undeveloped, suggestion in Shaftesbury's writings that our sense of "just and unjust" finds its origin in "a natural presumption or anticipation . . . on which resentment or anger is founded" (MPR 3.2, p. 234). Butler takes up this suggestion by explaining that resentment is what a person naturally feels when he has been treated in a manner "which he thinks other than what is due to him" (15S Preface, p. xvi). This feeling, moreover, can be shared sympathetically; individuals "have this resentment in behalf of others, as well as of themselves" (AR 1.3, p. 53) To be given less than what you are due by a responsible moral agent is, in the terminology of the day, to suffer an injury at that agent's hands. Butler therefore concludes that "injury is the only natural object of settled resentment," that "men do not in fact resent deliberately anything but under this appearance of injury" (15S Preface, p. xv).²⁹ From this psychological fact, Butler thinks it is easy to deduce the final cause or purpose of resentment. "The natural object or occasion of settled resentment then being injury, as distinct from pain or loss," Butler writes, "it is easy to see, that to prevent and to remedy such injury, and the miseries arising from it, is the end for which the passion was implanted in man" (15S 8, p. 97). The fact that we feel resentment allows us to maintain our natural balance of passions when sympathy might otherwise lead us to excessive mercy for wrongdoers. "Since . . . it is necessary for the very subsistence of the world that injury, injustice and cruelty should be punished," Butler writes, "and since compassion, which is so natural to mankind, would render that execution of justice exceedingly difficult and uneasy, indignation against vice and wickedness is, and may be allowed to be, a balance to that weakness of pity, and also to anything else which would prevent the necessary methods of severity" (15S 8, p. 99).

Although Hutcheson admits that resentment of the sort that Butler describes is natural, useful, and widespread, he nonetheless argues that resentment cannot be the source of our sense of justice. “This passion,” he writes, “however wisely implanted, must be under the control of a higher principle” (SMP 2.3.2, Vol. I, p. 256). Hutcheson maintains that only general benevolence—a desire for the public good—can provide the appropriate criterion for when resentment should be indulged and when it should be condemned as excessive. “Neither anger, nor hatred of the criminal, nor even that honest indignation at moral evil which is natural to every good man, should be the sole springs of punishing,” he says, “but rather a calm regard to the common interest, and the safety of the innocent” (PMIC 3.8.9, p. 273).

Butler, however, argues that there is a way to put natural resentment “under the control of a higher principle” without appealing to the public good in this way. He begins from the observation that “vice in general consists in having an unreasonably and too great regard to ourselves in comparison of others” (15S 10, p. 117). When attempting to bring his resentment under control or when attempting “to have a due natural sense of the injury, and no more” an injured party “ought to be affected toward the injurious person in the same way any good men, uninterested in the case, would be” (15S 9, p. 108). The goal is to see one’s injury as anyone else would see it, and react as they would react. We must, as Adam Smith would later put it, adopt the perspective of an impartial spectator and judge to what degree our resentment would be warranted in his view. Butler writes:

We are in such a peculiar situation, with respect to injuries done to ourselves, that we can scarce any more see them as they really are than our eye can see itself. If we could place ourselves at a due distance, i.e., be really unprejudiced, we should frequently discern that to be in reality inadvertence and mistake in our enemy, which we now fancy we see to be malice or scorn. From this proper point of view, we should likewise in all probability see something of these latter in ourselves, and most certainly a great deal of the former. Thus the indignity or injury would almost infinitely lessen, and perhaps at last come out to be nothing at all.
(15S 9, p. 109)

Here, in a sketchy form, are all the essential elements of Smith’s theory of justice, the most sophisticated such sentimentalist theory yet devised and the basis for a complete and plausible sentimentalist system of natural jurisprudence. It would be up to Hume, Smith, and Herder collectively to develop the free-standing, liberal, and pluralist version of sentimentalism that still has much to contribute to political philosophy and political practice today. Their work would have been impossible, however, were they not building on foundations laid by Shaftesbury, Butler, and Hutcheson.