

OXFORD

FOUNDERS

ROUSSEAU

A Free Community of Equals

JOSHUA COHEN

Founders of Modern Political
and Social Thought

ROUSSEAU

FOUNDERS OF
MODERN POLITICAL AND
SOCIAL THOUGHT

SERIES EDITOR

Mark Philp

Oriel College, University of Oxford

The *Founders* series presents critical examinations of the work of major political philosophers and social theorists, assessing both their initial contribution and their continuing relevance to politics and society. Each volume provides a clear, accessible, historically informed account of a thinker's work, focusing on a reassessment of the central ideas and arguments. The series encourages scholars and students to link their study of classic texts to current debates in political philosophy and social theory.

Also available:

JOHN FINNIS: *Aquinas*

RICHARD KRAUT: *Aristotle*

GIANFRANCO POGGI: *Durkheim*

MALCOLM SCHOFIELD: *Plato*

MAURIZIO VIROLI: *Machiavelli*

CHERYL WELCH: *De Tocqueville*

ROUSSEAU

A Free Community of Equals

Joshua Cohen

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

© Joshua Cohen 2010

The moral rights of the author have been asserted
Database right Oxford University Press (maker)

First published 2010

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organization. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department,
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose the same condition on any acquirer

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Library of Congress Control Number: 2009941595

Typeset by Laserwords Private Limited, Chennai, India
Printed in Great Britain
on acid-free paper by
MPG Books Group, Bodmin and King's Lynn

ISBN 978-0-19-958149-8 (Hbk.)
978-0-19-958150-4 (Pbk.)

1 3 5 7 9 10 8 6 4 2

For Ellen, Bob, Alene, Daniel, and Isabel

This page intentionally left blank

Contents

<i>Acknowledgments</i>	ix
<i>List of Abbreviations</i>	xi
Introduction	I
1. A Free Community of Equals?	10
<i>Free and in Chains? The Society of the General Will</i>	14
<i>Realism? Natural Goodness and Democracy</i>	16
<i>Three Aims</i>	20
2. The Society of the General Will	23
<i>The Fundamental Problem</i>	24
<i>A Solution: The Society of the General Will</i>	32
3. Reflections on the General Will's Sovereignty	60
<i>Groups, Sovereignty, Consensus, Majorities, and Rights</i>	60
<i>A Solution to the Fundamental Problem?</i>	84
4. The Natural Goodness of Humanity	97
<i>Three Properties of Human Nature</i>	100
<i>Motivations</i>	106
<i>Natural Goodness</i>	110
<i>A Genealogy of Vice</i>	113
<i>Complementary Motivations</i>	122
<i>Natural Goodness and Reasonable Faith</i>	127
5. Democracy	131
<i>Some Institutions of the Society of the General Will</i>	135
<i>Principles and Institutions</i>	140
<i>Four Strategies of Argument</i>	145
<i>Popular Democracy or Executive Dominance?</i>	166
<i>A Democrat After All?</i>	175
Notes	177
Bibliography	189
Index	195

This page intentionally left blank

Acknowledgments

Chapters 2, 3, and 5 draw on my "Reflections on Rousseau: Autonomy and Democracy," *Philosophy and Public Affairs* 15/3 (Summer 1986), 275–97. Reprinted in Christopher W. Morris (ed.), *The Social Contract Theorists: Critical Essays on Hobbes, Locke, and Rousseau* (Lanham, Md. and Oxford: Rowman & Littlefield, 1999) (*Critical essays on the classics*), 197–213; in Thom Brooks (ed.), *Rousseau and Law* (Aldershot: Ashgate, 2005); in Timothy O'Hagan (ed.), *Jean-Jacques Rousseau* (Aldershot: Ashgate, 2007).

Chapter 4 is based on my essay "The Natural Goodness of Humanity," in Andrews Reath, Barbara Herman, and Christine Korsgaard (eds.), *Reclaiming the History of Ethics: Essays for John Rawls* (Cambridge: Cambridge University Press, 1997).

This page intentionally left blank

Abbreviations

- B *Letter to Beaumont*, in *Collected Writings of Rousseau*, vol. 9, trans. Christopher Kelly and Judith Bush (Hanover, NH: The University Press of New England, 2001)
- Cor. *Constitutional Project for Corsica*, in Jean-Jacques Rousseau, *Political Writings*, trans. and ed. Frederick Watkins (Madison: University of Wisconsin Press, 1986)
- C *Confessions*, in Jean-Jacques Rousseau, *Œuvres complètes*, vol. 1 (Paris: Gallimard, 1959).
- D1 *First Discourse (Discourse on the Arts and Sciences)*, in Rousseau: *The Discourses and Other Early Political Writings*, ed. and trans. Victor Gourevitch (Cambridge: Cambridge University Press, 1997)
- D2 *Second Discourse (Discourse on the Origin of Inequality)*, in Rousseau, *The Discourses and Other Early Political Writings*, ed. and trans. Victor Gourevitch (Cambridge: Cambridge University Press, 1997)
- E *Emile*, trans. Allan Bloom (New York: Basic Books, 1979)
- FP *Fragments politiques*, in Jean-Jacques Rousseau, *Œuvres complètes*, vol. 3 (Paris: Gallimard, 1964)
- GM *Geneva Manuscript*, in *On the Social Contract with Geneva Manuscript and Political Economy*, ed. Roger D. Masters, trans. Judith R. Masters (New York: St. Martin's, 1978); references to book, chapter, and paragraph (GM 2.2.4 = book 2, chapter 2, paragraph 4)
- LD *Letter to d'Alembert on the Theater*, in Jean-Jacques Rousseau, *Politics and the Arts*, trans. Allan Bloom (Ithaca, NY: Cornell University Press, 1980)

ABBREVIATIONS

- LM *Letters Written from the Mountain*, in *Collected Writings of Rousseau*, vol. 9, trans. Christopher Kelly and Judith Bush (Hanover, NH: University Press of New England, 2001)
- M1 Letter to M. de Malesherbes, in Jean-Jacques Rousseau, *Œuvres complètes*, vol. 1 (Paris: Gallimard, 1959)
- M2 Letter to Mirabeau, in Rousseau, *The Social Contract and Other Later Political Writings*, ed. and trans. Victor Gourevitch (Cambridge: Cambridge University Press, 1997)
- N *Narcissus*, preface, in Rousseau, *The Discourses and Other Early Political Writings*, ed. and trans. Victor Gourevitch (Cambridge: Cambridge University Press, 1997)
- OL *Essay on the Origin of Languages*, in *Collected Writings of Rousseau*, vol. 7, trans. and ed. John T. Scott (Hanover, NH: University Press of New England, 1998)
- P *Considerations on the Government of Poland*, in Rousseau, *The Social Contract and Other Later Political Writings*, ed. and trans. Victor Gourevitch (Cambridge: Cambridge University Press, 1997)
- PE *Political Economy*, in Rousseau, *The Social Contract and Other Later Political Writings*, ed. and trans. Victor Gourevitch (Cambridge: Cambridge University Press, 1997)
- RJ *Rousseau, Judge of Jean-Jacques*, in *Collected Writings of Rousseau*, vol. 1, trans. Christopher Kelly, Judith Bush, and Roger Masters (Hanover, NH: University Press of New England, 1990)
- SC *Social Contract*, in Rousseau, *The Social Contract and Other Later Political Writings*, ed. and trans. Victor Gourevitch (Cambridge: Cambridge University Press, 1997); references to book, chapter, and paragraph (SC 2.2.4 = book 2, chapter 2, paragraph 4)

Introduction

I first read Rousseau in 1973. A beginning PhD student in philosophy, I was taking John Rawls's course on social and political philosophy. I think we read parts of the *Social Contract* and *Discourse on Inequality*. I found Rousseau's work annoying and confusing. It seemed high on ringing phrases, self-indulgence, and portentousness, low on clarity and sustained argument. Despite Rawls's interpretive efforts, I was not getting what Rousseau was about.¹ Things improved some two years later when I was a teaching assistant for the same course. Still, I was having trouble with both trees and forest.

Despite these misgivings, I stayed with it. Rousseau's themes were so important, and his impact so large: I had to assume that the fault was mine.

A cluster of points about themes and impact seemed especially important to me. I was interested, for example, in Rousseau's ideas of direct democracy, which inspired modern ideas of participatory democracy. In this connection, I wanted to understand Marx's idea of a "withering away of the state," and could not see anything in Marx's enthusiasm for the direct democracy of the Paris Commune that was not in Rousseau's account of popular legislative assemblies. Although I was concerned about the charge that authoritarianism and terror were close cousins of these enthusiasms, I was reassured by what Kant—beyond reproach on authoritarianism and terror—said of Rousseau: Rousseau had "set me straight," and taught "[me] to respect mankind." Kant compared Rousseau and Newton: "Newton first saw order and lawfulness going hand in hand with great simplicity, where prior to him disorder and its troublesome partner, multiplicity, were encountered, and ever since the comets run in geometrical paths; Rousseau first discovered amid the manifold human forms the deeply hidden nature of man, and the secret law by which Providence is justified through his observations."² Having spent much time trying to understand Hegel's political philosophy, with its critique of individualism, I also felt the force of the appreciative (if somewhat grudging, in the context) remark in his *Lectures on the History of Philosophy*: "The principle of

freedom emerged in Rousseau, and gave man, who apprehends himself as infinite, this infinite strength. This provides the transition to the Kantian philosophy, which theoretically considered made the principle its foundation."³ And I had a growing sense of Rousseau's impact on Rawls, who once said in passing that his two principles of justice could be understood as an effort to spell out the content of the general will.

In addition to being struck by these lines of influence, I was drawn to Rousseau's identification of and claim to have solved what he calls the "fundamental problem" of the social contract: "To find a form of association that will defend and protect the person and goods of each associate with the full common force, and by means of which each, uniting with all, nevertheless obey only himself and remain as free as before" (SC 1.6.4). I was interested, too, in the effects of private property and inequality on political equality. On these subjects, I was struck both by Rousseau's critical discussion of private property and inequality in his *Discourse on Inequality*, with its concern about psychological and political effects, and its relative lack of attention to concerns about the (un)fairness of inequality. Finally, I thought that Rousseau rightly resisted the temptation to read his moral convictions into a science of history.

Rousseau, in short, had powerfully influenced the moral-political thinkers who most interested me; he had addressed the issues about democracy, civic equality, and political autonomy that seemed most fundamental; and he combined morally forceful social criticism with an understanding of the fragility of moral progress, and its costs. Here was an optimism of heart not head—a hopefulness about human possibilities, without extravagant assurances of progress or the intellectual conviction that history was on his side.

As I taught Rousseau in courses at MIT in the late 1970s and early 1980s, the pieces started falling into place. After a few years, I thought I had a more coherent account of Rousseau than was available in the English-language literature, at the least the parts of the literature of which I was aware. Moreover, I was troubled that there was not a very good treatment of Rousseau written in a more analytical style. So sometime in the early 1980s, I decided to write a book on Rousseau's political theory.

The book, as I initially conceived it, would do four things. First, it would explore Rousseau's ideas about democracy, in

particular about participation and citizen engagement with the substance of political issues, but also more generally about political institutions. Much in those ideas seemed attractive, but their attractions required that they be formulated apart from his obviously implausible picture (implausible in contemporary terms) of citizens in a small republic gathering in person in a legislative assembly, or his exaggerated expectations about social and political consensus. Second, it would explain and assess Rousseau's more abstract conviction—expressed in his statement of the fundamental problem—about the possibility of combining autonomy with political authority, his thought that legitimate political authority is a form of self-legislation, a condition of “moral freedom” in which one obeys “the law one has prescribed to oneself” (SC 1.8.3). Third, it would explain the intimate connections between Rousseau's convictions about equality and freedom. And fourth, it would provide an account of Rousseau's political views with a level of clarity and attention to argument that would distinguish it from much of the literature on Rousseau. In 1985, with an academic leave supported by an ACLS (American Council of Learned Societies) fellowship, I read more widely in Rousseau's corpus, explored lots of the (not very satisfactory) secondary literature on Rousseau, and started to write.

This book is the result, written in just the way that one should never write a book: fitfully, with many stops and starts, over too many years.

I produced about 25,000 words in 1985–6, and published a shortened version of the material as a review essay in *Philosophy and Public Affairs* (1986). Many of the leading ideas in Chapter 2—about Rousseau's problem, the nature of the general will, how the society of the general will solves Rousseau's problem, and why we should not think of Rousseau as a “self-effacing Hobbesian”—appeared in that early essay. I also sketched the ideas, developed in greater detail in Chapter 5, about the strategies of institutional argument.⁴

Because my attention was drawn to other projects, some of which involved developing a conception of deliberative democracy that was partly of Rousseauian inspiration, I found it hard to sustain the focus on the book needed to finish it. As a result, I filled out the details slowly, largely in the context of teaching political philosophy seminars at MIT. I am very grateful

to the many students in those courses for their comments and criticisms.

Not that the problems in finishing it were only matters of distraction. Two large, substantive problems stood in the way.

First, I did not have a very good grasp of Rousseau's doctrine of the natural goodness of humanity: an unfortunate limitation because this idea, Rousseau says, runs through all his work. In particular, Rousseau describes his account of inequality in his *Second Discourse* as a "genealogy" of vice (B 28). I could not see how exactly to understand that genealogy, because I did not see how to fit it together with the account of the society of the general will in the *Social Contract*. Readers of Rousseau sometimes see a conflict between a "primitivist" Rousseau of the *Discourse on Inequality*, celebrating our natural state of unreflective innocence, and a Rousseau in the *Social Contract*, who had made his peace with culture and authority. I was sure that this view was wrong, but was having trouble seeing how the pieces hung together. I assumed they did, not least because of Rousseau's own confident assertions about the unity of his work: responding to criticisms about his own inconsistencies and vacillations, he says "I have written on various subjects, but always with the same principles: always the same morality, the same belief, the same maxims, and if you will the same opinions" (B 22). He identified the doctrine of natural goodness as "the fundamental principle of all morality about which I have reasoned in all my Writings" (B 28). Although this idea is never stated in the more specifically political writings—its fullest expression is in *Emile*, and it does not appear in the *Social Contract*—it seemed clear that a confident grasp of the political theory required an understanding of this central theme.

In 1988 (I believe), I read galleys of Nicholas Dent's excellent book *Rousseau*, and found his account of Rousseau's psychological views eye-opening. Aided by Dent's interpretation (see below, Chapter 4 n. 9), I found a way to fit Rousseau's doctrine of natural goodness together with Rousseau's views about autonomy, authority, and democracy, as part of an account of how the society described in the *Social Contract* might be realized. I incorporated this material into the evolving manuscript, which I continued to work on largely in the context of teaching. In 1994–5, I extracted the account of natural goodness, and expanded it as a separate paper: my contribution to an edited collection

of papers written by students of John Rawls who had worked on issues in the history of moral and political philosophy. I folded that paper back in, thought the book might be getting close to finished, and sent the manuscript to Mark Philp for his Oxford University Press series.

In 1999, I taught a short course at Oxford on Rousseau. I am grateful to the participants, including Philp and Andrew Williams, for very helpful discussion. In preparing for the course, I was struck by a second very important limitation on my understanding of Rousseau. I reread the early chapters of his *Government of Poland*, and concluded that I was underplaying (to a fare-thee-well) the more “communitarian” strands in Rousseau’s work: not his republican focus on the importance of a vigilant citizenry animated by civic virtue, but his emphasis on social solidarity and national attachment, on the “reforms required to make love of fatherland the dominant passion” (P 188), on “distinctive practices . . . always exclusive and national” (P 181) as a basis of political solidarity, and associated suspicions about political disagreement and concerns about its destructive effects. In some of the more recent pieces of the book, I have tried to remedy this deficiency. Rousseau’s views, I believe, draw together an egalitarian-democratic ideal of a free community of equals, founded on a conception of individuals as free and animated by self-love, and owing much to the modern contractualist tradition, with a sometimes-communitarian political sociology, focused on the social solidarities that are arguably required to unite the independent members of a society of equals.

The communitarian political sociology is not the part of Rousseau I find most attractive. But it is a very powerful presence, with strong resonances in Rousseau’s important writings on language and music.⁵ No sensible interpretation can put it to the side, and not only for reasons of interpretive fidelity. Those of us who are attracted to the ideal of a free community of equals need to take seriously the fact that one of its great exponents combined it with an (unattractively) anti-political communitarianism, with a large emphasis of solidarities built on national distinctiveness, and the fear that disagreement is the canary in the coal mine, rather than a normal condition of the only kind of political life worth hoping for.

Finally, in Fall 2008, after several false starts and on a promise to the publisher, I taught a seminar at Stanford on Hobbes and

Rousseau, and finished the work in the context of the course. I am very grateful to the students in the seminar for their indulgence, their helpful comments and criticisms, and their encouragement. Assaf Sharon in particular made some very helpful suggestions, all of which I have followed. (And Marilie Coetsee provided essential assistance in completing the manuscript.)

Although I say that I have finished it, I am acutely aware of its many limitations, some of which reflect the odd writing process, stretched over too many years, managed in fits and starts. I want to call attention to one of the many substantive omissions that limits the discussion. Rousseau believed that women should be excluded from politics and he believed that the justification of that exclusion is provided by the "nature" of women.⁶ I have assumed here—assumed, but not argued—that it is possible to provide a reconstruction of important elements of Rousseau's political philosophy while simply abstracting from his view that natural sexual differences are of decisive social significance. In simply assuming this for the purposes of the discussion here, I do not mean to suggest that it is obviously true—though I do believe that it is true.

Although the book is limited in this and many other ways, I am confident that it improves on the cleaner but vastly oversimplified book I would have finished twenty years ago, had I been able to concentrate exclusively on it. I am also sure that it is, for better or worse, a less coherent book than I would have done then, or would have written now, had I started from scratch rather than adding pages and interspersing paragraphs. Despite these limitations, I am persuaded by readers of the manuscript that it makes enough of a contribution to be worth publishing. In particular, I think it presents a picture of Rousseau's distinctive contribution to the tradition of democratic thought: his ideal of a democracy as a free community of equals. I think there is much to be said for this ideal, and that Rousseau provided its initial formulation.

One last point on the writing. Because it took such an unusual path, I have not been very attentive to the more recent literature on Rousseau. In particular, I regret that I have not been able to engage in the text with Frederick Neuhouser's wonderful book *Rousseau's Theodicy of Self-Love*, which appeared in Fall

2008.⁷ Nor have I discussed the treatment of Rousseau in Rawls's *Lectures on the History of Political Philosophy* (see n. 1, above). The published material on Rousseau is different from what I heard in 1973 and 1975, but I never would have been able to write this book without the initial direction provided by Rawls's lectures, and the continuing inspiration provided by his model.

Rousseau's Corpus

This book, as I have said, focuses on Rousseau's political theory. It thus omits large areas of Rousseau's work that are of extraordinary interest, or touches on them only insofar as they bear on the issues of political theory. In his remarkable biography of Rousseau, Leo Damrosch says: "In a series of amazingly original books, of which the *Social Contract* is the best known, he developed a political theory that deeply influenced the American Founding Fathers and the French revolutionaries, helped to invent modern anthropology, and advanced a concept of education that remains challenging and inspiring to this day. His *Confessions* virtually created the genre of autobiography as we know it, tracing lifelong patterns of feeling to formative experiences and finding a deep unity of the self beneath apparent contradictions; modern psychology owes him an immense debt."⁸ All that, without having attended school for a single day. And there is much else: *Le Devin du Village*, a comic opera admired by Gluck and the very young Mozart, performed 400 times (including at Fontainebleau and the Paris Opera, the first performance after the fall of the Bastille); a less successful play, *Narcissus, or the Self-Lover*, which was performed by the Comédie-Française in 1752; and *Julie, or the New Héloïse*, one of the most popular novels of the eighteenth century.

I follow Rawls in separating Rousseau's writings into three broad groups.⁹ In his early and more "critical" writings, including his *Discourse on the Sciences and Arts* and *Discourse on Inequality*, as well as his *Letter to d'Alembert on the Theater* (in which he objects to a proposal that a theater be built in

Geneva), Rousseau challenges the dominant Enlightenment view that the advance of science and understanding has improved the human condition, making human life freer, happier, and more virtuous. As an alternative, Rousseau argues for a connection between enlightenment and the evolution of social and political constraint, unhappiness, and vice. A central part of his story is the emergence of a rage to distinguish ourselves, and, more fundamentally, a destructive preoccupation with how we fare relative to others.

Then, second, we have Rousseau's more positive writings: *Social Contract*, *Emile*, and *Julie*, as well as the constitutional writings on Poland and Corsica, and his important account of the Genevan constitution and political system, written in response to the condemnation there of *Emile* and the *Social Contract*. In these works, Rousseau offers an account of political institutions and education, designed to show how we might repair our corrupt conditions, return to a free, happy, and virtuous life while benefiting from the development of human powers that occurred under corrupt conditions¹⁰, and maintain legitimate political institutions in the face of the inevitable pressures to degenerate that come from, inter alia, concentrated executive power.

Finally, in his more personal writings, including the *Confessions*, *Dialogues*, *Rousseau*, *Judge of Jean-Jacques*, and his beautiful *Reveries of the Solitary Walker*, Rousseau explains and justifies himself, affirms through detailed self-revelation his own singularity and authenticity, claims that he has not been trapped in the elaborate web of deception, hypocrisy, manipulation, and pathological preoccupation with status and reputation that we have woven for ourselves, and (perhaps) suggests that we, too, may be able to extricate ourselves from it. How, if Rousseau's own earlier depiction of our corrupt state is correct, could anyone have freed him- or herself from it in sufficient measure to have written Rousseau's books? Marx faced a similar kind of question: how, if what Marx said about the pervasiveness of ideology is correct, could Marx himself have seen through the mystical veil covering society's life process, and grasped the laws of motion of capitalism? Marx's answer was tied to an account of the evolution of capitalism and the experience of the working class in that evolution. Rousseau's answer to the comparable question about

INTRODUCTION

ideology and understanding points to the distinctiveness of his own life as an outsider.

In this book, I concentrate principally on the concerns in the second set of writings. Although I draw freely on the others, I address the issues they raise only insofar as they contribute to addressing the issues in Rousseau's political theory—fundamentally, the ideal of a free community of equals—that provide the book's central focus.

I

A Free Community of Equals?

Organized societies are marked by profound differences of power and advantage. Some people make decisions—about war and peace, taxes and public projects, health and education, public security and family life, rules of exchange and permissible forms of worship—with fateful, life-shaping, and life-shattering impact on the lives of less powerful others. Some lives are blessed by economic, social, and cultural advantages that others lack. Those differences of power and advantage result from some mix of sheer fortuity and human decision. To the extent that they reflect human decisions, or could be addressed and ameliorated through such decisions, what could possibly justify those differences of power and advantage? Can they be justified at all?

Jean-Jacques Rousseau offers a distinctive answer to these great questions. In strikingly spare, intense prose, he gives us a picture of a *free community of equals*, a social-political world in which individuals realize their nature as free by living together as equals, giving the laws to themselves, guided in those lawgiving judgments by a conception of their common good. Moreover, a free community of equals, Rousseau tells us, is not an unrealistic utopia beyond human reach, but a genuine human possibility, compatible with our human complexities, and with the demands of social cooperation.

Rousseau presents his ideal of a free community of equals with greatest force in his most important work of political thought, *Of*

the Social Contract. That book carries the subtitle *Principles of Political Right*. His aim, as the subtitle indicates, is to provide principles that distinguish right and wrong in the organization of our social and political life. "Man is born free, and everywhere he is in chains. One believes himself the others' master, and yet is more slave than they. How did this change come about? I do not know. What can make it legitimate? I believe I can solve this question" (SC 1.1.1).

Solving it—finding the principles of political right—requires that we address a "fundamental problem." We need "[t]o find a form of association that will defend and protect the person and goods of each associate with the full common force, and by means of which each, uniting with all, nevertheless obey only himself and remain as free as before" (SC 1.6.4).¹ What kind of society ensures that the individual members of the society are both secure—protected in their person and goods by the collective power of the society—and also fully autonomous—each a self-legislating member, obedient only to him- or herself?

That is the fundamental problem because self-love and freedom are both basic to our human nature. Self-love is a sense of our own worth and concern about our well-being. Because it is essential to our nature, we have a basic interest in ensuring protection of our person and of the goods we need to survive and live well. But not just any kind of protection will do. Not, say, the protection of a benevolent lord, nor of the sovereign in Hobbes's leviathan state. More generally: not protection that depends on submission and thus insults our freedom. Human beings are "born free," with the capacity to resist the pull of our inclinations, make judgments about the best aims and proper principles of our conduct, and regulate our own conduct in light of those judgments: "It is . . . not so much the understanding that constitutes the specific difference between man and other animals, as it is his property of being a free agent. Nature commands every animal, and the Beast obeys. Man experiences the same impression, but he recognizes himself free to acquiesce or to resist" (D2 140–1). Moreover, this capacity—this "power of willing, or rather of choosing" (D2 141)—is the source of humanity's special worth, and the basis of our standing as responsible, moral agents, with rights and duties. So "[t]o renounce one's freedom is to renounce one's quality as man, the rights of humanity, and even its duties" (SC 1.4.6; D2 141, 179).

Rousseau aims, then, to describe a form of social association that provides security without demanding an alienation of our freedom because he is concerned that our “chains”—the social rules and expectations necessary to establish a “common force” that protects person and goods, as well as the rules established by that power—may conflict with the freedom that belongs to our nature and lies at the basis of our worth. We need security without renunciation, a political order that is morally legitimate because it both provides protection to each person, and also respects the dignity of its free members by ensuring their full autonomy: by establishing a form of self-rule, each as free as before.

This problem—combining autonomy and the order on which our security depends—resists easy solution. Consider a collection of people who live in a common territory and regularly interact. Each person’s security and well-being—meeting the concerns that grow from self-love—depend on how other people act. Ensuring the security of person and goods then (arguably) requires *authoritatively imposed* constraints on the conduct of others. Why authoritatively imposed? If each person is to be secure, then *some* constraints on conduct are needed; and unless we are dealing with a world of angels—looking for a scheme that works “for the people of Utopia” but is “worthless for the children of Adam” (M2 270)—effective constraints must be backed by power sufficient to motivate compliance. But such power, again arguably, requires backing from an authority that is regarded as rightfully imposing the constraints, and fixing obligations to obey: “[t]he stronger is never strong enough to be forever master, unless he transforms his force into right, and obedience into duty” (SC 1.3.1). But if authoritatively established constraints are necessary to “defend and protect the person and goods of each associate,” then how can we meet the concerns that grow from self-love while also ensuring “moral freedom”—the full political autonomy that consists in “obedience to the law one has prescribed to oneself” (SC 1.8.3; LM 232)?

According to a familiar line of thought—Hobbes’s *Leviathan* provides its classical formulation: we cannot. Security and the pursuit of happiness depend, as Hobbes said, on peace. But given the “known natural inclinations of mankind” and the facts of human interdependence, peace requires *submission*. Each person must exchange self-government rights in return for safety

and the hope of happiness: an agreement that proceeds “as if every man should say to every man *I authorise and give up my right of governing myself* [my emphasis] *to this man, or to this assembly of men, on this condition, that thou give up thy right to him, and authorise all his actions in like manner.*”¹² Hobbes deployed this argument in support of a leviathan state, whose members are subject to the sovereign’s unconditional authority. It is widely agreed that his political absolutism exaggerated the necessary terms of trade: the extent to which we must sacrifice our self-government for the sake of security. Suppose, then, that an authority less expansive than an absolute Hobbesian sovereign suffices for achieving peaceful order. Still, we may wonder how each citizen could achieve *full* political autonomy—meaning that each person gives the law to him- or herself, regards him- or herself as its author—within an organized society. A sphere of personal freedom within the bounds of law: no large problem in theory. A share of public freedom, as joint author of laws: also no large problem in theory. But remaining “as free as before” by giving the law to yourself, by being the legislator of the authoritatively imposed constraints that apply to yourself and other members: that is another matter.

But it is Rousseau’s idea. Hobbes’s large purpose in *Leviathan* is, he says, “to set before men’s eyes the mutual relation between protection and obedience” (Lev. 491). Contrast this with Rousseau’s idea that “the essence of the political body” is “the concurrence of obedience and freedom” (SC 3.13.5), a harmony that ensures protection, without demanding a morally unacceptable subordination of will: “in the relations between man and man the worst that can happen to one is to find himself at the other’s discretion” (D2 176). The idea is fundamentally different, and it may strike us a nice thought. But what could it possibly mean? How *could* each “[unite] with all” under common rules for security, while obeying only him/herself and so remaining “as free as before”? How is it possible to combine the social union under common, enforceable rules that provides protection with the political autonomy Rousseau describes? How can we achieve the “moral freedom” that consists in giving the law to oneself, while living together under authoritatively imposed constraints with others who share with us the dignified, freedom-affirming status of self-legislators (SC 1.8.3)?

That is the question, and Rousseau's answer—a realistic ideal of a free community of equals—has two components, corresponding to two kinds of doubt about the possibility of an answer: doubts about *content* and doubts about *realism*. I will consider them in turn in this chapter, and then in detail in the rest of the book.

Before proceeding, however, I want to mention a third possibility problem, which I do not discuss in any detail in the rest of the book. This problem arises from doubts about "accessibility": is there any route leading from current circumstances to the society of the general will? The society of the general will might be humanly possible but inaccessible, if we have become too corrupt. Corresponding to the three problems of possibility, we can distinguish three ways that political thought might be utopian: it might rest on values that simply cannot be jointly realized under any conditions; it might endorse values whose realization is incompatible with human nature; and it might embrace an ideal that cannot be realized by a social trajectory that begins from current conditions (barring some catastrophe that "wipes the slate clean"). Rousseau certainly focused less on the problem of accessibility than on the other problems, though his proposed constitution for Poland suggests serious—which is not to say successful—engagement with it.³ In any case, I will not address it here.

Free and in Chains? The Society of the General Will

Accepting authority *appears* to be a matter of letting oneself be ruled by the decisions of others (perhaps the majority), by treating those decisions as binding, as decisions that are rightly made and with which one ought to comply. How can *self*-government—the moral freedom or autonomy that consists in giving the law to yourself—be reconciled with these bonds of political authority? To show that the idea of autonomy (self-legislation) in chains (a common lawmaking authority) is even coherent, we need some way to dispel the appearance that acknowledging political authority requires letting oneself be ruled by others.

Rousseau addresses this problem with his conception of a society guided by a *general will*. The idea of the general will is one of the essential ideas in Rousseau's political philosophy, and I will be describing the ideal of a society regulated by a general will in more detail later. Suffice to say here that ideas of equality and the common good are fundamental to it, thus fundamental to achieving autonomy in a political society. In the society of the general will, citizens share an understanding of the common good and that understanding is founded on the members' commitment to treat one another as equals by refraining from imposing burdens on other citizens that those members would be unwilling to bear themselves. Thus the content of the understanding of the common good reflects an equal concern for the good of each citizen; citizens take that shared understanding to be the ultimate basis of their political deliberations, and express it by jointly settling on the laws of their community; finally, they acknowledge political obligations as fixed by laws founded on the common good, and the limits of collective legal regulation as fixed by the need to justify such regulation by reference to the common good (reading together SC 2.1.1, 2.4.5, 2.6, 2.11.2): "From whatever side one traces one's way back to the principle, one always reaches the same conclusion: namely, that the social pact establishes among the Citizens an equality such that all commit themselves under the same conditions and must all enjoy the same rights" (SC 2.4.8). Moreover, when they are "subjected only to conventions such as these, they obey no one, but only their own will" (SC 2.4.8).

How does the general will, thus interpreted, provide a solution to the fundamental problem? To sketch briefly an answer that I will discuss in detail later: Because the content of the conception of the common good that lies at the basis of the laws reflects an equal concern with the well-being of each citizen, the society provides security for person and goods. And because citizens share the conception, and the laws emerge from that shared conception, each citizen remains free in fulfilling his legal obligations.⁴

The essential point about content is that Rousseau's solution requires that individuals commit to regarding themselves as belonging to a political community whose members are committed to regarding one other as equals: acknowledging one another as political equals, with equal status in establishing the

laws; recognizing one another as equally subject to the laws; and agreeing to regulate their association by reference to reasons of the common good, which give equal weight to the good of each citizen. Moreover, Rousseau proposes to institutionalize the general will's supremacy through a direct democracy, whose equal citizens regularly assemble to reaffirm their social bonds and decide on the fundamental laws best suited to advancing their common good, and in which limits on social-economic inequality help to sustain the institutions.

Rousseau's solution to the fundamental problem, then, aims to reconcile full autonomy with the authority required for personal security, and full autonomy with equality and community. Under conditions of social interdependence, we achieve full autonomy or self-government only by living in a community of equals. Requirements of equality do not stand as limits on free association, but instead are both ingredients of and preconditions for such an association. And community is not the enemy of liberty and equality, but a setting defined by a commitment to both.

In short, Rousseau's solution to the fundamental problem is his ideal of a *free community of equals*: *free*, because it ensures the full political autonomy of each member; a *community*, because it is organized around a shared understanding of and supreme allegiance to the common good; and a *community of equals*—a democratic society—because the content of that understanding reflects the good of each member.⁵

Realism? Natural Goodness and Democracy

But can we live this way? Assume for now that the ideal of a free community of equals solves the fundamental problem. In any case, it has its attractions. But can people live this way? Can human beings really live in a free community of equals, or is that ideal a utopia, well beyond our reach?⁶

Here we have the second problem of possibility—the problem of realism—and it has two elements. I will call the first element the problem of *motivational possibility*. A free community of equals requires a shared understanding of and allegiance to the

common good; it is founded on a commitment to treat others as equals, to refrain from imposing burdens on them that one would not be prepared to shoulder oneself. In the face of widespread vice—pettiness, pride, jealousy, envy, and selfish indifference to human suffering—reasonable people may wonder whether that ideal is humanly possible, whether its motivational demands are compatible with our human nature.

Hobbes would doubtless have raised this objection. He argued that we needed to alienate our rights of self-government to a leviathan state. And his case for that unhappy conclusion rested on a philosophical anthropology—an anthropological pessimism—that makes Rousseau's solution motivationally unrealistic, whatever its attractions as a political ideal. Surveying the "known natural inclinations of mankind" (Lev. 489), Hobbes found desires for individual preservation and happiness; he noted the strength of human fears about violent death; he observed (in at least some people) passions of pride, envy, and greed rooted in a sense of natural differences of worth and a concern that social standing mirror those presumptively natural differences; and he found that people are often blinded by those passions—prompted by them to act for near-term advantages and against their own longer-term interests in preservation and happiness.⁷ Departing from these observations, he concluded that human beings need to live under the rule of a sovereign with unconditional authority. Only a sovereign with such unbounded authority would have power sufficient to overawe subjects—to tame their passions (pride in particular) with fear, and thus ensure the peace required for preservation and felicity (Lev. chap. 18). "For by this authority, given him by every particular man in the commonwealth, he hath the use of so much power and strength conferred on him, that *by terror thereof* [emphasis added] he is enabled to form the wills of them all, to peace at home" (Lev. 120).

Hobbes's case for political submission is driven, then, by a general pessimism about human capacities for self-regulation, even an individual's own prudential self-regulation in pursuit of a longer-term good.⁸ But Hobbes was particularly skeptical about the idea that people might be motivated by reasons of the common good. When people act rationally, they are moved by long-term benefits to themselves, and not by the

thought that their conduct treats others as equals, or is part of a system of conduct that ensures such treatment. And insofar as we are prone to passions of pride, we will reject equality as inconsistent with our naturally superior worth and an insult to our dignity.⁹

To be sure, Hobbes believes that we *are* naturally equals in our fundamental bodily and mental powers. Moreover, his ninth law of nature condemns pride and commands “*that every man acknowledge other for his equal by nature*” (Lev. 97). But the principal reason for following that command—as with the other laws of nature—is that compliance increases chances for peace, which in turn increases chances for one’s own preservation and felicity.

Rousseau’s answer to the problem of motivational possibility would have been easier if he had found Hobbes an insufficiently acute observer of humanity. But Rousseau largely accepted Hobbes’s dismal description: “Men are wicked; a sad and constant experience makes proof unnecessary” (D2 197). We observe widespread vice—selfishness, pride, jealousy, envy—and underlying that vice a “frenzy to achieve distinction” (D2 184), an “ardent desire to raise one’s relative fortune less out of genuine need than in order to place oneself above others” (D2 171). This frenzy for distinction and desire for relative gain have their roots, Rousseau argues, in an inflated, false sense of self-worth: the same sense of pride that drove Hobbes to endorse a leviathan sovereign to rule over the proud, the vainglorious who are not prepared to regard others as their equals by nature (Lev. 220–1). In principle, a sense of duty, its content tied to the common good, could override these tendencies to vice, thus taking care of the problem of motivation. But Rousseau did not put much weight on this possibility because he was skeptical about the motivational strength of the sense of duty. When the passions oppose our sense of duty, the sense of duty cannot be expected to win (see below, pp. 123, 144–5).

Achieving full political autonomy, then, requires a community of equals. But if a commitment to treating others as equals has no basis in human psychology, then autonomy is not in the human cards. We might then reject Rousseau’s solution as objectionably utopian, or condemn human nature as sadly barren soil for the demands of political morality: find that a free

community of equals is too demanding for humanity, or that humanity is too low for justice. Whatever the right response, if we could directly infer intrinsic properties of human nature from observed motivations—if, for example, we were entitled to conclude that the observed “ardent desire” for relative advantage is a direct expression of an original human predisposition, part of our human nature—then the society of the general will would be incompatible with our nature. We would be forced to pessimistic conclusions about the possibility of a free community of equals.

Rousseau’s response is that we have no reason for confidence in any such direct inference from observations about human motivation and conduct to the intrinsic properties of our nature. We, therefore, have no reason to endorse the thought or its pessimistic implications. That, in brief, is the point of Rousseau’s idea of the “natural goodness” of humanity. He aims to defeat an argument that begins with the evidence provided by “sad and constant experience” and concludes by rejecting an ideal of a free community among equals. In response to the problem of motivational possibility, Rousseau advances an account of human nature organized around the contention that human beings are *naturally good*, “but that society depraves him and makes him miserable” (RJ 213). And this account blocks the inference from dismal experience to a dismal human nature, and, more to the point, underwrites the possibility of our being well-motivated citizens, with suitably public concerns, in the society of the general will.

Showing that a free community of equals does not make impossible demands on human motivation is, however, insufficient to address concerns about realism. A free community of equals must also be *socially and politically possible*: can an ongoing society meet the conditions described by that ideal? I will call this the problem of *institutional possibility*. To show that a free community of equals is institutionally possible, we need, for example, to understand how, in a workable political society, people might come to acquire a general will, with its characteristic regard for others as equals and associated concern for the common good; how members might come to assign the general will priority in public decisions; how the general will might regulate the terms of cooperation; and whether the general will

can retain that regulative role, despite a range of pressures from other passions and interests that work against it.

Three Aims

That is Rousseau's program. I will explore it here, examining the fundamental problem and the solutions to the two problems of possibility. In Chapters 2 and 3, I discuss the fundamental problem and its solution: the first possibility problem, the problem of content. In Chapter 4, I present the conception of human nature and natural goodness, and then take up the first aspect of the problem of realism, the problem of motivational possibility. In Chapter 5, I discuss the second problem about realism, its institutional side. I explore the more concrete proposals, including democratic lawmaking, public participation, and law, about how to institutionalize the general will in ways that will elicit the humanly possible motivations that support it.

I have three aims. First, I want to highlight the general plausibility of the program of reconciling the values of autonomy and equality both with one another and with a conception of human community. Urging the possibility of such reconciliation—the ideal of a free community of equals—was Rousseau's distinctive contribution to political philosophy.

Second, I want to present a way to think about Rousseau's distinctively "participatory" conception of democracy, with its emphasis on direct citizen involvement in lawmaking. I will suggest that that conception of democratic order derives from the conjunction of Rousseau's normative ideal of a free association regulated by a "general will" and his psychological views about the formation of self-understandings and motivations. Once we distinguish the normative ideal from the specific institutional implications that Rousseau draws from it we will be in a position to consider how that ideal might be realized or approximated under modern conditions, in which Rousseauian direct democracy is implausible. In Habermas's terms, we will be in a position to consider how "the old promise" of a community of free and equal members, guiding their collective conduct through their common reason, can be redeemed if it is "reconceived under the conditions of complex societies."¹⁰ I will not describe the

terms of such redemption here, but open some space for its consideration.¹¹

Third, I want to explore the relationships between two tendencies within Rousseau's political thought. Rousseau's political ideal of a free community of equals has a strongly liberal cast: it is founded on values of individual self-love and freedom, justified through a compact among individuals conceived of as free and equal, aimed at advancing the basic interests of individuals, concerned to establish relations of equality under law, and it requires that equal citizens give priority in politics to their common good. The arguments are secular; the only reason for the exercise of political authority is public utility; there is no trace of an organic conception of society; nor is authority designed to serve the cause of human perfection.

Another strand in Rousseau's thought might be called communitarian, or republican: for the elements that concern me, "communitarian" strikes me as more accurate. What defines this strand is an emphasis on a shared social and national solidarity, an attachment to a distinctive way of life, and on political agreement as the natural outgrowth of that common attachment. Rousseau often (though not always) seems to treat political disagreement as a sign of impending disaster: he worries about what happens when "votes are no longer unanimous" (SC 4.1.4), and sometimes endorses very demanding requirements of civic unity and solidarity. Among those demanding conditions are a mandatory civil religion, strong attachments to place, compatriots, and distinctive usages, and a pervasively vigilant sense of civic responsibility. Thus, in a striking passage in *Emile* (I will return to it later), Rousseau says: "Good social institutions are those that best know how to denature man, to take his absolute existence from him in order to give him a relative one and transport the *I* into the common unity, with the result that each individual believes himself no longer one but part of the unity and no longer feels except within the whole. A citizen of Rome was neither Caius nor Lucius; he was a Roman. . . . The Lacedaemonian Pedaretus runs for the council of three hundred. He is defeated. He goes home delighted that there were three hundred men worthier than he to be found in Sparta" (E 40).

I will suggest a way to understand the connections between these different tendencies in Rousseau's view. Rousseau, I will argue, is prepared to entertain the possibility that human nature

requires us to accept demanding conditions of civic solidarity as preconditions for a free community of equals: absent pervasive ethno-national devotions, he suggests, we will be psychologically unable to sustain the free community of equals that is authorized by the social compact. The social compact itself and the conception of a free community of equals do not establish strong communitarian demands of solidarity: those demands are part of Rousseau's political sociology of a free community of equals. Although I do not find that sociology compelling, and do not think that Rousseau so fully endorsed the most exaggerated statements about ethno-national solidarity, I think it is essential to appreciate the substance and the roots of this more communitarian side of Rousseau's political thought. One result of reading Rousseau this way—as philosophically liberal, and sociologically communitarian—is that it suggests, as a more general thought, that the issue between more communitarian and more liberal traditions of political thought may lie less in the core of their political philosophies than in their distinctive social psychologies and political sociologies. And that means that, while their disagreements run deep, they may also share more common ground than we sometimes suppose. In any case, convictions about the possibility of a free community of equals need some story about civic solidarities. If Rousseau's is too narrowly confining, an alternative is needed. There is no evading the issue, not in a political philosophy that deserves the name "political."

2

The Society of the General Will

Rousseau aims to reconcile human autonomy with the fact of social interdependence and the possibilities for human life that it creates. In describing this reconciliation, he needs to address two problems of possibility: the problem of content and the problem of motivation. I will start by exploring the issue of content: what would a reconciliation of autonomy and social association consist in? How could it be that we live with others in ways that provide basic protections for our good, while at the same time remaining as free as before, achieving the moral freedom of self-legislation?

The question, and the answer, too, are usefully divided into two parts—the first concerns the prospects of reconciliation at the level of abstract principle, whereas the second part concerns the institutional implications of reconciliation, what I called the problem of institutional possibility, the second part of the concern about realism.

In this chapter, I explore the first part—essentially, Rousseau's theory of the general will. After setting out more fully Rousseau's "fundamental problem," I sketch the main features of his proposed solution: what I will call "the society of the general will." In Chapter 3, I explain some features of the ideal in more detail, elaborate and clarify certain aspects of the notion of the general will and of a social order regulated by such a will, and discuss some reasons for thinking that the general will

solves the problem of content. In Chapter 5 I explore the second part—essentially, Rousseau’s theory of democratic institutions.

The Fundamental Problem

The *Social Contract* gives us an account—as its subtitle says—of “principles of political right.” Rousseau offers a solution to a problem, and understanding the solution requires understanding the problem.

As background, Rousseau assumes that individuals are socially interdependent, and have acquired some understanding of notions of justice. These assumptions are not stated explicitly, but are implicit in the discussion of the background to the social contract (SC 1.6.1), at least when that paragraph is read alongside the discussion of the emergence of conflict in his *Discourse on Inequality* (D2 170–4). Thus the argument of the *Social Contract* is not addressed to isolated individuals (on Rousseau’s understanding, such individuals would lack language and developed cognitive powers), or to individuals for whom such isolation is a genuine practical possibility, or individuals who are in—or whose conception of themselves is identical with that of individuals who are in—an amoral state of nature prior to human interdependence, lacking moral categories. Nor is he concerned to explain how individuals come to be interdependent in the first place, or how they acquire moral ideas, or how they acquire the motivation to act morally.

None of this is said, not in so many words. But Rousseau’s writing is as laconic as it is elegant. He does not waste words, so you need to pay close attention to the ones he uses.

The aim in the *Social Contract* is not to explain anything at all. The aim is to *justify* terms of association (“principles of political right”), to show that a certain form of political association is legitimate by showing that individuals would themselves agree to that form, where those individuals are understood to be interdependent, aware of their interdependence, endowed with the capacity to distinguish just from unjust arrangements, and endowed with a capacity for freedom: “Man is born free, and everywhere he is in chains. One believes himself the others’ master, and yet is more a slave than they. How did this change

come about? I do not know. What can make it legitimate? I believe I can solve this question" (SC 1.1.1). More particularly, then, the aim is to show that legitimate authority is compatible with human freedom, and to indicate the conditions that political authority must meet if it is to be legitimate.

Abstracting from certain matters of detail, Rousseau assumes at least the following three conditions as background to the justification (see SC 1.6).

First, given basic human nature, each person is motivated by self-love, by, that is, a concern for his/her own good. This general concern is expressed, more particularly, in a concern for *self-preservation* and *personal security*, and for the *goods* required for individual well-being. He assumes that people want their own preservation and well-being, that they value these (self-love is, as I explain later, a matter of valuation as well as affection), and that a concern for these goods is something that each person owes to him- or herself (*les soins qu'il se doit*) (SC 1.6.3). Moreover, he appears to assume a basic interest in the development and exercise of our faculties (SC 1.8.1), a human good which depends on social cooperation: in the civil state, he says, a person's "faculties are exercised and developed, his ideas enlarged, his sentiments ennobled, his entire soul is elevated" (SC 1.8.1). Rousseau is sometimes said to have exalted the simple life, but this is entirely misguided. His concern is always about the price that we pay for the good of developing and exercising our powers, and about whether it is possible to achieve that good without paying a terrible price.

Second, individuals are *interdependent*. Thus, the satisfaction of the needs and interests associated with self-love—minimally, interests in personal security and protection of goods, more expansively, interests in the development of our faculties—depends on how others (as well as oneself) act(s). Though not a basic fact of our nature, this interdependence is part of the background of the need for political institutions and so part of the backdrop for an argument about their justification: the link in our fates is one of life's givens, not an option. That social interdependence has, moreover, a particular character. If each acts solely with the aim of advancing his or her own interests, all do less well than they could through the coordination of their actions: "men having reached the point where the obstacles that interfere with their preservation in the state of nature prevail

by their resistance over the forces which each individual can muster to maintain himself in that state" (SC 1.6.1).

This point about the nature of interdependence—that uncoordinated interaction, with each acting on a separate plan, leaves each of us less well-off—can be put in terms of Rousseau's notion of the "will of all." Rousseau distinguishes the will of all from the general will (SC 2.3.2). I will come to the general will later on. When he describes the will of all, he says that while the general will "looks only to the common interest," the will of all "looks to private interest, and is nothing but a sum of particular wills" (SC 2.3.2). What makes a will particular is its content, the aims that are regulative for it in practical deliberation. The crucial point is that the regulative aims of the agent do not give any weight to the good of all members of the association. The agent could itself be a collective or "moral" person within a larger political society: say, an interest group, or a political party, or the executive itself. The dominant particular will of an agent may often aim at the good of the agent. But the essential point is the narrowness of concern, not its selfishness.

In saying, then, that the will of all is "nothing but a sum of particular wills" I take Rousseau to mean that the will of all, unlike the general will, is really not a single, unitary will at all. Instead of referring to the "will of all," he might instead have said "the wills of each." Thus a state of affairs results from the will of all (that is, the wills of each), rather than the general will, just in case it results from conditions in which each agent in a set of interdependent agents—whose well-being depends on how others act, and who all know that to be true—acts with a view to his/her own advantage (or some other advantage less encompassing than the good of every agent), without explicitly coordinating with the actions of others. For example, each person pursues his/her own security, without concern for the security of others, and without coordinating with those others to provide common security. An outcome results from or reflects the will of all, then, just in case it results from each person acting for his/her own advantage, without coordinating actions with others. The idea of the will of all comprises two conditions, then, and both are important: that the content of the aims of the agents is narrower than the good of all who are interdependent, and that there is no coordination to achieve aims.

The justification of political arrangements, then, assumes circumstances in which when each person acts for his/her own advantage—when the outcome is determined by the will of all—the result is collectively disadvantageous. Here, the relevant measure of advantage is provided by the basic interests of each person—in particular, the interests in security of person and goods. So we are assuming conditions in which independent, uncoordinated action is less advantageous (for example, less secure) for each person than alternatives available through coordination: uncoordinated interaction is suboptimal.

I assume too—as part of this second condition—that individuals *recognize* their interdependence. In particular, they understand the dangers they face from the uncoordinated pursuit of particular interests, and know, too, that mutually beneficial coordination is possible. None of these assumptions are especially controversial: they certainly would not have been in dispute between Hobbes and Rousseau.

Third, individuals have views about the claims that they can legitimately make on one another, and those views tend to conflict. Conceptions of justice and entitlement tend to be highly contested, because they arise under conditions of interdependence from several distinct sources. For example, we tend to think we are entitled to the goods that meet our basic needs (D2 171); similarly we tend to think we are entitled to the products of our labor (D2 169). At the same time, those with power and wealth seek to maintain that power and wealth by representing it to others as legitimate and thus as properly acknowledged by others (D2 173). For “the stronger,” Rousseau says, “is never strong enough to be forever master, unless he transforms his force into right, and obedience into duty” (SC 1.3.1).

Thus coordination is required for mutual advantage. In the absence of such coordination, we face conflicts of interests. Moreover, we face a conflict-deepening tendency to moralize the opposition of interests. We see others as aiming not only to get what we would prefer to have, but also to get what we have a right to, what is legitimately ours. Conflicts of interest, in short, easily transmute into contests of right, honor, and worth.

The *Social Contract* aims to identify norms of social cooperation that are reasonable, given these three conditions, and given as well the fundamental human capacity for *freedom*. By the

“capacity for freedom” I mean that each of us has a capacity to resist our inclinations, reflect on how best to act, and act from our own judgments about what is best overall.

“Nature commands every animal, and the Beast obeys. Man experiences the same impression, but he recognizes himself free to acquiesce or to resist; and it is mainly in the consciousness of this freedom that the spirituality of his soul exhibits itself” (D2 141). Or again: “I always have the power to will, I do not always have the force to execute. When I abandon myself to temptations, I act according to the impulsion of external objects. When I reproach myself for this weakness, I listen only to my will” (E 280). When Rousseau (here speaking through the Savoyard Vicar of *Emile*) goes on to explain what does determine the will when we resist the call of inclinations, he identifies the cause with an agent’s own judgment: his/her own reflective judgments of what is suitable to do. “One chooses the good as he has judged the true; if he judges wrong, he chooses badly. What, then, is the cause which determines his will? It is his judgment. And what is the cause which determines his judgment? It is his intelligent faculty, it is his power of judging: the determining cause is in himself. Beyond this I understand nothing more” (E 280).

This capacity to choose—understood as involving the capacity to suspend the impelling force of inclinations, reflect, and act guided by evaluative judgments—is distinctive to human beings among all animals (D2 141). It is also the source of our special worth (it is our “noblest faculty”). And it is the basis of our nature as moral agents: that is, of our capacity to act as responsible agents, and proper subjects of rights and obligations, of praise and blame. For that reason, “to renounce one’s freedom is to renounce one’s quality as a man, the rights of humanity, and even its duties” (SC 1.4.6). Because freedom, an “essential [Gift] of Nature,” is the source of our standing as moral agents, alienating our freedom is impermissible (D2 179).

What would it mean to alienate one’s freedom? It would be to agree to accept someone else’s judgments about what to do as supremely authoritative, and through such agreement to take oneself to be under an obligation to follow that person’s judgments: I need not agree with their judgment about the proper course of conduct, but I am obliged in any case to follow it. And,

just to underscore, that obligation of course remains in force, even if the other person's judgment conflicts with one's own. To alienate one's freedom is, as Rousseau puts it in the first draft of the *Social Contract*, to be "subjected without qualification to the will of another" (GM 1.5.10). The person who alienates freedom places herself under an obligation to regulate her conduct by another's will—a supreme obligation, which takes precedence over other normative requirements.

Rousseau offers at least three arguments against the permissibility of such alienation.

The first argument is, broadly speaking, instrumental, and parallels an argument Locke offers in his *Second Treatise*. Freedom, Locke says, is a fence to my preservation, and "reason" thus "bids me look on him as an enemy to my preservation who would take away the freedom which is a fence to it."¹ And again, "I have no reason to suppose, that he, who would take away my liberty, would not when he had me in his power, take away everything else."² For these reasons, I cannot give up my freedom. Similarly, Rousseau says that my freedom is one of the "primary instruments" for my own preservation (SC 1.6.3). Thus, each person is, as a matter of natural disposition and affection, bound to preserve him- or herself: "His first law is to attend to his own preservation, his first cares are those he owes himself" (SC 1.2.2). But each person is, from the age of reason, the best judge of the means needed for such preservation. To alienate freedom, then, would be to place one's preservation under the care of someone less capable of judging how best to secure it, and thus to neglect "the cares he owes himself" (SC 1.6.3).

This argument—we find variants in Hobbes and Locke—has some force, but a force limited by Rousseau's larger purpose. If freedom is preservation's instrument, then it can be alienated up to the point that would result in a threat to preservation. So perhaps the implication would be an inalienable liberty of self-defense of the kind that Hobbes embraces in his doctrine of the true liberties of subjects (see below, p. 38). But Rousseau seems to want more than that: to require that I remain as free as before, having my own will, thus my own judgment, as a more comprehensive guide to conduct. And the instrumental case seems poorly suited to that ambitious conclusion.

The second argument, which I will call the "rationality argument," is that freedom is the "noblest faculty" and "most

precious of all . . . gifts" (D2 178), of supreme value. To alienate freedom, then, by relinquishing ultimate authority over one's own conduct would be to trade a greater for a lesser value (exchanging your birthright for a mess of porridge, self-government for protection). On this assumption about value, such a trade would be irrational, and therefore void: "illegitimate and null, for the simple reason that whoever does so is not in his right mind" (SC 1.4.4). And "madness does not make right" (SC 1.4.4). This argument of course depends on a substantial claim about the value of liberty, and of the importance of retaining ultimate authority over one's conduct in one's own will and judgment. Rousseau endorses such a claim: he thinks that we have a sense of our special worth, and that it is tied to the thought of ourselves as free, as having the power to act under the guidance of our own sense of what is best. But the rationality argument assumes this claim, it does not establish it.

A third line of argument—I will call it the "moral nature" argument—helps to support the thesis of the value of freedom that figures in the rationality argument. Thus Rousseau says that "it is important to me that my freedom not be abused, and I cannot risk becoming the instrument of a crime without incurring the guilt of the evil I shall be forced to commit" (D2 179). The thought is that if I alienate my freedom then I oblige myself to follow the will of a superior, an authority. And if that superior commands me to do something evil—say, kill an innocent person—I will be required to do it. Because the other person's will is supremely authoritative ("subjected without qualification"), the obligations take precedence over other practical requirements. But—here is the crucial assumption—I cannot absolve myself of responsibility for the conduct, or avoid the sense of guilt for having done it. My sense of guilt shows that I attribute the conduct to myself—hold myself responsible for it—rather than to the agent who ordered it. And that self-attribution expresses my thought that I retain control of my conduct, am bound by standards of conduct from which I cannot escape by submitting myself to another person, and therefore that my effort to alienate my freedom is unsuccessful. It is unsuccessful because, as the sense of guilt indicates, I am subject to and aware that I am subject to norms that have priority over the will of the agent under whose supreme authority I sought to place myself, and

of an undiminished capacity to comply with those norms: my awareness that I ought to have and could have acted otherwise than I was told.

In *Emile*, Rousseau underscores (in the voice of the Savoyard Vicar) this connection between our sense of ourselves as free and moral feelings of guilt and remorse: "When I abandon myself to temptations, I act according to the impulsion of external objects. When I reproach myself for this weakness, I listen only to my will. I am enslaved because of my vices and free because of my remorse. The sentiment of my freedom is effaced in me only when I become depraved and finally prevent the voice of the soul from being raised against the law of the body" (E 280). Thus I take myself to be free when I experience feelings of remorse because such feelings reveal standards of conduct to which I am attached and my sense that I could have followed them.

The rationality argument and the moral nature argument, then, work together as follows: if alienating freedom is irrational, then agreements to alienate are void. Such alienation is irrational if freedom is the supreme value (and if it is the fence to my preservation). But I am committed to taking freedom to have this supreme importance because freedom is required for acting on standards to which I take myself—as my moral feelings indicate—to be supremely bound. "To renounce one's freedom is to renounce one's quality as man, the rights of humanity, and even its duties. There can be no possible compensation for someone who renounces everything. Such a renunciation is incompatible with the nature of man" (SC 1.4.6). And I suppose that Rousseau is thinking that this renunciation is incompatible with taking ourselves to be under an *obligation* to keep a contract of submission: for the thought that we are subject to the obligation to keep the agreement, thus to submit to the will of another, comes with the thought that we are free, thus responsible agents.

The third argument connects the value of freedom to our nature as moral beings, a nature revealed in our susceptibility to feelings of guilt and remorse. So long as we are susceptible to such feelings, and so long as they are connected to standards of conduct other than those that are dictated for us, we are unable to alienate our liberty—to accept that we are supremely bound to comply with the will of another person. We must take our own will as a rule because the conception of ourselves as free is

so deeply embedded in, essential to, moral thought and feeling, to our membership in the moral order.

This combination of self-love, interdependence, and the capacity for and importance of free action leads us to “the fundamental problem”: to “find a form of association that will *defend and protect the person and goods of each associate* with the full common force, and by means of which each, uniting with all, nevertheless *obey only himself and remain as free as before*” (SC 1.6.4, emphases added). The problem—as the two italicized clauses indicate—is to characterize a form of social interdependence that both permits each to secure his or her well-being—the protection of person and goods that one owes to oneself, the development and exercise of capacities, the broadening of ideas and feelings (SC 1.8.1)—and at the same time requires no sacrifice of the freedom that defines our nature. Put in the terms that I used earlier, Rousseau’s question is: Given the fact of self-love, conditions of social interdependence in which the common advantage requires coordination, and the natural tendency to endorse competing views of justice that deepen conflicts, can we agree to mutually beneficial terms of cooperation—terms that protect and defend the persons and goods of each—that enable us to be governed by our own judgments of what is, on the whole, the best thing to do?

This fundamental problem—the first problem of possibility—is the “problem to which the social contract provides the solution” (SC 1.6.4). To see how it is possible to reconcile autonomy and social connection, we need to answer the question: What form of association would be rationally agreed to by equal persons, who are moved by self-love and, above all, by an interest in securing their freedom?

A Solution: The Society of the General Will

Rousseau’s contractual problem has, he claims, a determinate solution (SC 1.6.5): “The clauses of this contract are *so completely determined by the nature of the act* [emphasis added] that the slightest modification would render them null and void; so that although they may never have been formally stated, they are everywhere the same, everywhere tacitly admitted and

recognized" (SC 1.6.5). To be sure, there are many good reasons for variation in the more particular shape of legitimate political arrangements, depending on their context. For example, there is not a general answer to the question "which is absolutely the best Government" (SC 3.9.1). But it is different with the fundamental problem, expressed in the question: what *would* (not could) individuals, themselves moved by self-love and bound to preserve their freedom, agree to? And Rousseau, like Hobbes, Locke, and Rawls, thinks that *that* question has a determinate answer. For Rousseau, as for these others, the idea of an agreement is not simply a way to think about political justification in particular settings by asking what all can consent to in those settings, but a way to fix the basic principle of political right for all such contexts. What is that principle? If we have an answer to this question available, then we may be able to see how it does provide the solution to the fundamental problem.

The outcome of the social contract, thus the solution to the fundamental problem, is a political society in which each member "puts his person and his full power in common under the supreme direction of the general will" (SC 1.6.9). In such a society there is a shared framework for collective judgments and decisions, and final authority rests in the shared understanding of the common good that partially characterizes that framework. By a "shared framework for collective judgments and decisions," I mean a shared set of principles and political values which all members regard themselves as bound to appeal to in political reasoning, whether they are reasoning and deciding on their own about a public issue, or offering argument to others about that issue. We could also say that in such a society, final authority rests with the people—that the people are sovereign. But we need then to keep in mind that the people is not simply a collection of individuals in a territory, or a collection under a common authority, as in Hobbes's theory of the people as an agent constituted through mutual subordination to the will of a single agent. The people is an agent—a moral person—that exists as such only through a shared understanding of the common good and shared recognition of the authority of that understanding: the people is constituted as an agent through its having a will, and what defines the will of the people is its orientation to the common good of those subject to it.

What, then, is it for an association of persons to be regulated by a general will that provides supreme direction? Such an association meets four conditions:

Particular Interest Condition (GW_I)

Members of the society of the general will have separate, particular interests (SC 1.7.7)—in particular, interests in their bodily security and material well-being—and those interests provide them with reasons for action. In Rousseau's terms, each person has a particular or private will (SC 2.3.2). The "public person" is composed of "private persons . . . whose life and freedom are naturally independent of it" (SC 2.4.2).

The particular interest condition is of course not at all distinctive to Rousseau's conception of the general will, but I begin with it because it is an element of his view that is easy to overlook or deny. In particular, the force of this first condition is that Rousseau does not embrace the strong conception of civic unity that we find in Plato's *Republic*, the Platonic conception—"the greater the unity of the state the better"—that Aristotle criticizes in his *Politics*.³ "Is it not obvious," Aristotle asks, "that a state may at length attain such a degree of unity as to be not longer a state?—since the nature of a state is to be a plurality."⁴ Plato's strong conception of unity is suggested in his wondering whether there is "any greater evil for a city than what tears it apart and turns it into many cities instead of one"⁵ and emerges with particular force in his account of the class of guardian rulers. As Plato describes that group, its members have shared knowledge of what is good in human life; they identify their own good with the good of the political society, and believe that they are doing well as individuals if and only if the political society is doing well; they identify their own good with the good of other guardians, and each thinks that he or she is doing well if and only if others are faring well; and, because there are not separate families and private property in the guardian class, the members of that group think of the same things as "mine" and "not mine" and—most strikingly—experience "feelings of pleasure and pain in common."⁶

Rousseau's view may seem more platonic than I have suggested here. In an illuminating essay distinguishing Rousseauian from liberal constitutionalist political views, Allan Bloom

says: "The struggle between inclination and duty, obstinate and irreconcilable, is the psychological price paid for the liberal social contract [which Rousseau rejects]. Only the man whose private will *wills only the common good* [emphasis added] would experience no tension between his individuality and society, freedom and duty."⁷ Bloom takes Rousseau to endorse this no-tension ideal, in which individual interest and identity are fully absorbed into a public identity. Similarly, Ernst Cassirer suggests that membership in the society of the general will requires "the *complete* [my emphasis] renunciation of all particular desires," that "Man does not give himself to the state and society without giving himself completely to both. We may speak of a real 'unity' of the state only if the individuals are merged in this unity and disappear in it."⁸ Durkheim, too, says "each individual will vanishes into a common, general will, which is the basis of society."⁹

These views about *complete civic unity*—about willing *only* the common good, and about a *complete renunciation* of particular desires—are not entirely lacking in support from Rousseau's texts, which sometimes suggest—sometimes rhapsodically—an ideal of a perfectly civic self. Thus, in his description of the lawgiver who founds a people and proposes a constitution for them, Rousseau says: "Anyone who dares to institute a people must feel capable of, so to speak, changing human nature; of transforming each individual who by himself is a perfect and solitary whole into part of a larger whole from which that individual would as it were receive his life and his being. . . .; of substituting a partial and moral existence for the independent and physical existence we have all received from nature. . . . The more these natural forces are dead and destroyed, the greater and more lasting are the acquired ones, and the more solid and lasting also is the institution: So that when each Citizen is nothing and can do nothing except with all the others . . . the legislation may be said to be at the highest pitch of perfection it can reach" (SC 2.7.3). In *Emile*, describing the public education that is most powerfully depicted in the *Republic*, he says: "Good social institutions are those that best know how to denature man, to take his absolute existence from him in order to give him a relative one and transport the *I* into the common unity, with the result that each individual believes himself no longer one but part of the unity and no longer feels except within the

whole. A citizen of Rome was neither Caius nor Lucius; he was a Roman. . . . The Lacedaemonian Pedaretus runs for the council of three hundred. He is defeated. He goes home delighted that there were three hundred men worthier than he to be found in Sparta" (E 40). Similarly, in his *Government of Poland*, Rousseau proposes that to create a stable political state in Poland, "the execrable proverb must be reversed, and every Pole must be made to say in his inmost heart: *Ubi patria, ibi bene*" (P 186). If our non-civic capacities are all "dead and destroyed," if there is not a Lucius, but only a Roman, and if our personal good so closely tracks the good of the country, then it would be wrong to think of the members of the society of the general will as having separate interests (while they might still have some uncivic desires, their interests would be entirely aligned with the political society).

Rhapsody notwithstanding, the complete civic unity view suggested by Bloom, Cassirer, and Durkheim gets Rousseau wrong. They exaggerate his view; or perhaps it would be more fair to say that they focus on Rousseau's own exaggerations, leaving aside the less rhapsodic moments in which he focuses on the possibility of a free community of equals under conditions in which the Particular Interest Condition is clearly in place: "Ancient peoples are no longer a model for modern ones; they are too alien to them in every respect. You above all, Genevans, keep your place, and do not go for the lofty objects that are presented to you in order to hide the abyss that is being dug in front of you. You are neither Romans, nor Spartans; you are not even Athenians. Leave aside these great names that do not suit you. You are Merchants, Artisans, Bourgeois, always occupied with their private interests, with their work, with their trafficking, with their gain; people for whom even liberty is only a means for acquiring without obstacle and for possessing in safety. This situation demands maxims peculiar to you. Not being idle as the ancients Peoples were, you cannot ceaselessly occupy yourselves with the Government as they did: but by that very fact that you can less constantly keep watch over it, it should be instituted in such a way that might be easier for you to see its intrigues and provide for abuses" (LM 292–3; LD 67).

Consider, for example, the "execrable proverb" in the *Government of Poland*. Rousseau's comment is clearly ambiguous in its substance—for example, it does not say that the good

of my *patria* exhausts my good—and seems clearly polemical in its intent. Moreover, elsewhere in *Poland*, where Rousseau emphasizes the importance of generating passionate national attachments—attachments of citizens to one another and to their fatherland (*patrie*)—as prior to the task of devising laws and constitutions, Rousseau indicates that the nation must be the “principal business” (*plus grande affaire*) of citizens (P 185), not their exclusive business, and that patriotism must be their “dominant passion” (P 188), not their sole passion.

Moreover, the great-legislator passage about the pitch of civic perfection conflicts with what Rousseau says elsewhere about the persistence of separate interests and a dimension of private personhood in the society of the general will. Here, as elsewhere, Rousseau’s disposition to exaggerated expression should prompt interpretive caution. So, for example, he tells us that the “private persons” who constitute the public person have a “life and freedom” that are “naturally independent of it” (SC 2.4.2). No suggestion here of the utter lack of independence indicated in the description of the legislator’s work. Moreover, the legislator passage itself does not say that the destruction of natural independence is *required* for a society founded on the general will, but only that it represents the *perfection* of the legislator’s work. As a general matter, Rousseau does not present the members of the society of the general will as having transcended conflicts between inclination and duty. Thus he says in *Emile* that “what is forbidden to us by conscience is not temptations but rather letting ourselves be conquered by temptations. It is not within our control to have or not to have passions. But it is within our control to reign over them. All the sentiments we dominate are legitimate; all those which dominate us are criminal” (E 445, see also 446, 473; RJ 158). The “virtuous man,” then, “is he who knows how to conquer his affections; for then he follows his reason and his conscience; he does his duty” (E 444–5).

More fundamentally, Bloom, Cassirer, and Durkheim conflate two very different ideas: the idea that the general will is sovereign or “supreme,” and the idea that it is exhaustive—“complete,” as Cassirer puts it. It is not true that the members of an association regulated by a general will lack separate interests, or that joining requires a renunciation of all particular interests. To put one’s person and powers under the *supreme* guidance of a general

will—which is what the social compact involves—is not to put them under the *sole* or exclusive direction of that will; the general will is sovereign and authoritative, not the exclusive source of reasons for conduct. When I describe someone as devoted to duty, what I mean is that he takes care to learn what it is and to fulfill it, not that he thinks that life is entirely a matter of acting from duty.

Consider again, now in light of this distinction between *supreme direction* and *exclusive direction*, the passage I quoted from *Emile*:

Natural man is entirely for himself. He is numerical unity, the absolute whole which is relative only to itself or its kind. Civil man is only a fractional unity dependent on the denominator; his value is determined by his relation to the whole, which is the social body. Good social institutions are those that best know how to denature man, to take his absolute existence from him in order to give him a relative one and transport the *I* into the common unity, with the result that each individual believes himself no longer one but part of the unity and no longer feels except within the whole. A citizen of Rome was neither Caius nor Lucius; he was a Roman. (E 40)

This passage seems to deny that, in the well-ordered political society, there are any interests apart from those associated with the role of citizen: indeed, that there is any identity apart from that of citizen. Tensions within Hobbes's view—expressed in his account of the true liberties of subjects, liberties which are retained as permissions to act against the sovereign's will—reflect the absence of any comparable story about civic unity (Lev. chap. 21). All the unity we find in Hobbes is unity in the submission to a sovereign will. Rousseau seems here to go to the opposite extreme. He seems to entertain the possibility of a civic unity so complete as to obliterate any separate interests and identity.

Consider, however, Rousseau's summary of this same discussion: "He who in the civil order wants to preserve the *primacy* [emphasis added] of the sentiments of nature does not know what he wants. Always in contradiction with himself, always floating between his inclinations and his duties, he will never be either man or citizen. He will be good neither for himself nor for others" (E 40). Here, once more, Rousseau indicates that the issue is "primacy." He is concerned with what takes precedence in practical

reason, in our deliberations about appropriate public conduct, not what exhaustively characterizes our will and identity.

One of Rousseau's chief concerns is to consider how, if at all, an association might be regulated by a general will, *given* the "opposition of particular interests [that] made the establishment of societies necessary" (SC 2.1.1). It is, Rousseau says, "the agreement of these same interests [i.e., the private interests] which made it possible. What these different interests have in common is what forms the social bond, and if there were not some point on which all interests agree, no society could exist" (SC 2.1.1). These remarks would make no sense were there not such separate interests. And they also would make no sense if the separateness and potential conflict of individual interests were a passing condition prompting social association, but to be transcended in a well-ordered polity, rather than a permanent background fact, helping to define the conditions in which the chains of authority are necessary and possibly legitimate. The separateness, potential conflict of, and possibly mutually beneficial coordination of interests should be counted among the "circumstances of political right," on analogy with what Rawls calls "the circumstances of justice."¹⁰

To be sure, the social bond is not simply the sum of private interests: those interests are in some ways opposed, and it is, again, their opposition that makes authoritative social coordination necessary. Each of us wants, for example, to be secure, but would perhaps prefer others to bear the costs of providing the good of peace and domestic security. So there is always the danger that "Each person, in detaching his interest from the common interest, sees clearly enough that he cannot separate them entirely, but his share of the public evil seems to him as nothing compared to the exclusive good which he seeks to make his own. Except for this particular good, he wills the public good in his own interest just as forcefully as anyone else" (SC 4.1.6). Our interest in peace is common; but we divide about—and so need to coordinate on—paying the costs of maintaining peace. But absent the interest of each person in his or her security, there would also not be the common interest in peace—which the different and opposed private interests have in common—that underlies social union.

Plato's guardians, then, have no private interests, or, apparently, private feelings. Rousseau's citizens do have such interests,

and civic unity in the society of the general will must take a correspondingly different form from Plato's ideal. I will call it "unity through ordering," as distinct from the "unity through integration" characteristic of Plato's ideal city. But before sketching that distinction and describing the structure of civic unity, we need to consider a second essential element of the society of the general will.

Common Good Condition (GW₂)

Citizens *share* and it is *common knowledge* that they share a conception of their "common good," although they may have—and indeed can in general be expected to have—different beliefs about what will advance that good. The fact of such differences of belief limits the expectations of unanimity that may be suggested by the idea of a shared view of the common good; I will come back later to the implications of these differences, and to the possibility of different conceptions of the common good itself. But first I want to say more about the idea of the common good.

Rousseau unfortunately says almost nothing very definite about the notion of the common good. Because the "common utility is . . . the foundation of civil society" (GM 1.5.18), and the social bond is founded on what our different interests have in common, this lack of definiteness is a weakness.

Contrast Rousseau's view in this respect with Rawls's. Rawls observes that a settled convention of political justification in democratic societies is to appeal to some conception of the common good. His theory of justice as fairness presents a case for a particular conception of the common good—the difference principle, as one element of a conception of democratic equality—as the most reasonable way to understand the idea of the common good in a democratic society, given the "ethos" of such a society. "[W]e may note," he says, "that it is a political convention of a democratic society to appeal to the common interest. No political party publicly admits to pressing for legislation to the disadvantage of any recognized social group. . . . The difference principle can . . . be interpreted as a reasonable extension of the political convention of a democracy [viz., to appeal to the common interest], once we face up to the necessity of adopting a reasonably complete conception of justice."¹¹ The difference principle is a way to give more determinate content to the notion

of the general will as a willing of the common good, in a society of equals.

Rousseau has nothing so definite to say about how best to interpret the idea of the common good. Still, he does suggest some constraints—I will mention four—on what will count as a genuine conception of the common good.

Distributive/Aggregative. First, the common good is not understood, as in classical utilitarianism, as a maximization of aggregate pleasure. This conception is deficient in two ways. First, the *good* that concerns Rousseau in his account of the common good is characterized in terms of shared human interests in security of person and goods and in liberty, not pleasure. Thus, society is to be directed to the “common interest,” where this is understood as what private interests that are widely shared have in common, the “point on which all [private] interests agree” (SC 2.I.I). In reply to the question “What is the aim of the political association?”, Rousseau responds: “the preservation and prosperity of its members” (SC 3.9.4). He adds in a footnote that we ought to use a similar measure in assessing the periods of history that have done most to advance the “prosperity of humankind.” Instead of assigning large weight to the “letters and arts,” we should look to the well-being of entire nations and “above all of the most numerous estates” (SC 3.9.4 n.). Rousseau thus has a different account of basic interests than classical utilitarianism.

Second, and of more immediate importance, advancing the common good seems to be understood in distributive, not aggregative, terms. It is not a matter of maximizing the sum of individual satisfactions, but of advancing each member’s fundamental, shared interests (SC 2.I.I, 2.II).¹² Underscoring his deep hostility to aggregative views of the common good, Rousseau says that the “maxim” that it is permissible for the government to “sacrifice an innocent man for the sake of the multitude” is “one of the most execrable maxims that tyranny ever invented, the most false that might be advanced, the most dangerous that might be accepted, and the most directly contrary to the fundamental laws of society” (PE 17).

This hostility expresses an individualizing concern that lies at the core of Rousseau’s view. In explaining the connection between the general will and common good in the *Social Contract*, Rousseau tells us that “while the opposition of particular interests made the establishment of societies necessary, it is the

agreement of these same interests [i.e., the particular interests] which made it possible. What these different interests have in common is what forms the social bond, and if there were not some point on which all interests agree, no society could exist. Now it is solely in terms of this common interest that society ought to be governed" (SC 2.1.1; GM 1.5.18). Moreover, "what generalizes the will is not so much the number of voices, as it is *the common interest which unites them*" (SC 2.4.7, emphasis added). Thus the common good is to be understood in terms of interests shared by the members of society, and as requiring attention to the interests of each member, in a way that gives equal consideration to each.¹³

Equality/Common Good. Indeed, the idea of according equal consideration to the individual members of the political society, each of whom is assumed to take an interest in advancing his/her own interests—associated with self-love—is arguably more fundamental to Rousseau's conception of the society of the general will than is any substantive idea of the common good (the same can be said of Rawls, who defends the democratic conception of equality expressed in the difference principle by reference to the more abstract ideal of fair terms of cooperation among free and equal persons). Arguably, the social compact is only derivatively an agreement to have the common good (our common interests) as our rule. In essence, we have an agreement to live together in a political society governed by a shared commitment to limit burdens on others (imposed through laws) to those burdens one would be prepared to live under oneself. Thus the claim that the social compact establishes a "condition [that] is equal for all, and since the condition is equal for all, no one has any interest in making it burdensome to the rest" (SC 1.6.6). Or again, "the social pact establishes among the Citizens an equality such that all commit themselves under the same conditions and must all enjoy the same rights. Thus by the very nature of the pact every act of sovereignty, that is to say every genuine act of the general will, either obligates or favors all Citizens equally" (SC 2.4.8). More simply, "The first and greatest public interest is always justice. All wish the conditions to be equal for all, and justice is nothing but this equality" (LM 301).

Moreover, this commitment—to impose only those burdens on others that one is prepared to accept for oneself—serves as

the principal guide in subsequent collective decision-making, rather than some definite conception of the common good: “in this institution, everyone necessarily submits to the conditions which he imposes on others; an admirable agreement between interest and justice which confers on common deliberations a character of equity that is seen to vanish in the discussion of any particular affair” (SC 2.4.7). Interest and justice agree because the members, who are prepared to impose on others only those conditions that they are prepared to accept for themselves, are also assumed to be concerned with their own good: “Why is the general will always upright, and why do all consistently will each one’s happiness, if not because there is no one who does not appropriate the word *each* to himself, and think of himself as he votes for all?” (SC 2.4.5)

Having said this, I nevertheless take the idea of the common good (defined in terms of common interests) to be fundamental to an understanding of the general will because Rousseau supposes it to be obvious that individuals who are prepared to impose on others only those conditions that they would be prepared to live under themselves will thereby be committed to advancing the common good, because each is concerned to advance his or her own interests: “the equality of right and the notion of justice it produces follows from each one’s preference for himself and hence from the nature of man” (SC 2.4.5). Or again, “Every condition imposed on each by all cannot be onerous to anyone” (LM 261). But I also emphasize that the common good needs to be interpreted against the background of the fundamental commitment to treat associates as equals.

Let’s return, then, to the common good. Consider again the example I mentioned earlier: I have an interest in my protection; you have an interest in yours. While we each have an interest in social peace, we each want to avoid the costs of ensuring it: our interests are opposed because while we are each willing to bear the costs as a price of social peace, our private will is not to bear those costs. The “sum of particular wills” (SC 2.3.2), then, is a world in which no one bears the costs, and so there is no protection. And the common interest—what remains when we “subtract off” the particular interest of each in not bearing the costs—is in the conditions of peace required for general security. This provides a way to interpret Rousseau’s famously obscure remark that the will of all “looks to private interest, and is

nothing but a sum of particular wills; but if, from these same wills, one takes away the pluses and minuses which cancel each other out, what is left as the sum of the differences is the general will" (SC 2.3.2).

Content of the Interests: Self-Development and Independence.

As to the content of the interests, of the individual goods that form the common good, Rousseau seems to limit the range of common interests to interests in security of person and protection of resources—that is, the very interests that he indicates as providing the foundation of the social contract. Thus his assertion in his account of the “signs” of whether a people is well governed or not, that the “aim of the political association . . . is the preservation and prosperity of its members” (SC 3.9.4).

Consider two other interests that seem to be lacking from the list of shared interests. In describing the advantages a person gains in the civil state, Rousseau says: “his faculties are exercised and developed, his ideas enlarged, his sentiments ennobled, his entire soul is elevated” (SC 1.8.1). Let’s say then that there is an interest in *self-development*. Although Rousseau acknowledges the good of the cultivation of human capacities—at least when those powers are kept in balance with our desires, producing a kind of equilibrium and happiness (E 80)—he does not seem to count such cultivation or development as part of a common good to be promoted through collectively authorized laws and regulations. Perhaps he thinks that all we can do to achieve such cultivation through collective efforts is to establish a legal order in which the people are collectively sovereign.

More fundamentally, Rousseau seems not to acknowledge a fundamental, shared interest in “individual independence.” I have in mind an interest in guiding the conduct of one’s own life by a set of values and principles with which one identifies, which may well be different from the values and principles of the political society and those of the other citizens of that society. I say he *seems* not to acknowledge an interest in individual independence, because he does say—in a discussion of religious toleration in his *Letters from the Mountain*—that the basic principle of Reformation Christianity is the central role of private judgment and interpretation in guiding religious belief and conduct (LM 154–5). This discussion underscores the importance of independence of judgment, at least in religious matters: “the individual mind is established as the sole interpreter of

Scripture; thus the authority of the Church is rejected; thus each is put under his own jurisdiction for doctrine. Such are the two fundamental points of the Reform: to acknowledge the Bible as rule of one's belief, and not to admit any other interpreter of the meaning of the Bible than oneself" (LM 154).

The unifying idea of the Reform, he continues, was that all Protestants "acknowledged each of them as competent judge for himself. They tolerated and they ought to tolerate all interpretations except one, namely that which removes liberty of interpretation" (LM 154). Thus "[e]ach remains the sole judge of them [doctrines] for himself, and does not acknowledge any authority in them other than his own. Good instruction ought less to fix the choice we ought to make than to put us in a condition to choose well. Such is the genuine spirit of the Reformation; such is its true foundation. Individual reason pronounces in it . . . and it is so much of the essence of reason to be free, that even if it wished to subject itself to authority, it would not be able to do so" (LM 155). Moreover, in his earlier *Letter to d'Alembert* he endorses, as a basic principle—without specific reference to the doctrines associated with Reform Christianity—that "human reason has no well-determined common measure and that it is unjust for any man to give his own as the rule to that of others" (LD 11 n.). And he indicates that this principle would provide a strong case for religious toleration.

But this interest in conditions of individual independence plays an uncertain role in the theory of the general will. And this makes the place of what Benjamin Constant called the "liberties of the moderns"—liberty of conscience and of the person—uncertain, because of their close association with independence of judgment. Rousseau does not say, for example, that the preservation of individual independence is itself among the basic, shared interests that help to define the content of the common good. Were he to have held that view, then he could have argued that a regulation that advances interests in security of person or resources but restricts individual independence might—because of that very restriction—not advance the common good. Nor, alternatively, does he claim—at least not explicitly—that we need to weigh considerations of the common good against the value of individual independence in deciding whether a regulation ought to be imposed. So we have an essential role for freedom as collective

self-legislation, but a less clear place for independence as shaping the content of self-legislation.

Instead the protection of individual independence—insofar as it receives such protection—results from the fact that restrictions on conduct must be founded on law, and the laws are to be justified by reference to the common good. As a result, restrictions that lack such a common-good justification—proposed regulations that cannot plausibly be represented as contributing to the common good—are arbitrary and impermissible, impermissibly arbitrary because unsupported by reasons of the appropriate kind. So when there is no common good justification for a regulation of, say, religious belief and conduct, individuals are to be left at liberty. This is the point of Rousseau's comment, in his case for the "limits of sovereign power," that "the Sovereign, for its part, cannot burden the subjects with any shackles that are useless to the community; it cannot even will to do so" (SC 2.4.4). He does not say that the sovereign *ought* not to burden subjects with useless regulations, but that the sovereign *cannot*. The sovereign cannot—and cannot even will to do so—because the sovereign general will aims at the common good. If a regulation is useless to the community, then, it cannot be an expression of the general will: no matter how much support it may have from a majority, a regulation that lacks plausible connection to the common good cannot be interpreted as an expression of the general will. And this fact provides the justification of toleration in the *Social Contract*: "Subjects . . . only owe the Sovereign an account of their opinions insofar as those opinions matter to the community" (SC 4.8.31). In this way, some measure of individual independence and thus of personal liberty emerges as a requirement that is internally related to the supremacy of the general will, not a constraint on the general will.

Still, the fact that a regulation would restrict independence of judgment does not, on this line of argument, itself count as a reason against endorsing it; and the fact that a regulation fosters individual independence does not count as a reason in favor of it. So, for example, insofar as opinions do matter to the community, insofar as the good of the community depends on them, they can permissibly be regulated. Thus the sovereign people may legislate a civil religion that contains principles useful to a society, and banish from the state anyone who does not endorse the dogmas

of a civil religion: "The existence of the powerful, intelligent, beneficent, prescient, and provident Divinity, the life to come, the happiness of the just, the punishment of the wicked, the sanctity of the social Contract and the Laws" and the rejection of tolerance for the intolerant (SC 4.8.33). The basis for the exclusion is not the non-believer's impiety, as measured by the standard of some broad religious outlook: the civil religion contains "all dogmas truly useful to either a universal or a particular society, and leaves out all the others, which may be important to faith but not at all to worldly well-being, the unique object of Legislation. For how can the mystery of the Trinity, for example, contribute to the good constitution of the State; in what way will its members be better Citizens when they have rejected the merit of good works; and what does the dogma of original sin have to do with the bond of civil society?" (LM 148). Instead the basis of exclusion is the secular concern that the person who rejects the dogmas of civil religion cannot be relied upon to comply with the laws: "it is impossible to be either a good Citizen or a loyal subject" because he is "incapable of sincerely loving the laws, justice, and, if need be sacrificing his life to his duty" (SC 4.8.32) Moreover, unless citizens love the laws, "the laws will invariably be evaded" (P 179).

Having entered these remarks about Rousseau's apparent inattention to the interest in individual independence as an ingredient in or constraint on the promotion of the common good, I want to add some qualifications. I do not think that Rousseau does in the end have a crisp account of the importance of individual independence and its role in the society of the general will. Nevertheless, I think we can find suggestions of it, for example in the remarks cited earlier on the basis of religious toleration in the central principle of Reform Christianity, and in the remarks about the essentially free nature of reason. Moreover, a larger role for it is compatible with much that he says. More particularly, I have said that Rousseau does not explicitly endorse the idea that regulations advancing the common good are to be endorsed by the people only if the benefits to the common good are of sufficient weight or importance to outweigh the burdens on individual independence: that is, individual independence does not appear as a weighty value. At the same time, however, he never endorses the idea that, in justifying a regulation, it *suffices* to show that the enactment and enforcement of

the regulation would advance the common good; he says only that such a showing is necessary.

For example, in his important discussion of the limits of sovereign authority, Rousseau says that the sovereign “cannot burden the subjects with any shackles that are useless to the community” (SC 2.4.4), not that all useful regulations are permissible, irrespective of the interests that they burden; and he says, too, that sovereign authority “does not . . . exceed the bounds of public utility” (SC 4.8.31), but he does not say that any regulation respecting those outer bounds is permissible. Indeed, the very use of the terms “burden” and “shackles” suggests that he might have had some hesitation about sufficiency.

Moreover, in the one place where he does appear to present a sufficient condition he says: “It is agreed that each man alienates by the social pact only that portion of his power, his goods, his freedom, which it is important for the community to be able to use, but it should also be agreed that the *Sovereign is alone judge of that importance*” (SC 2.4.3, emphasis added). The force of the passage is uncertain. Rousseau is pointing to a claim on which he takes there to be general agreement—that there is some qualification on the alienation by each in the social compact—and then both agreeing with the claim and qualifying its implications: yes, each person alienates only *what is important*, but the sovereign is judge of importance. It is easy to attend only to the latter point of qualification—that the sovereign is sole judge—and forget what Rousseau is agreeing to. Let’s keep in mind that the sovereign judge of importance is the sovereign people, and ask: why does the case for collective regulation, made by citizens to one another, need to show that the regulation is *important* for the community—that the regulation is *important* in advancing the common good—rather than simply that the proposed regulation is of *some* public benefit, however minimal?

To fix the question more precisely, let’s say that we make a case for the importance of a regulation by showing that there is no less burdensome way than the proposed regulation to advance the common good; or perhaps, more strongly, that there is no less burdensome way than the proposed regulation to advance the common good *and* that the contribution to the common good is considerable. So why would it be necessary to show that there is no less intrusive way to bring about the common good than

through the regulation under contemplation? A natural response is that individual independence is an important political value, not to be burdened unless such burdening is important: that regulations are “shackles.” And this thought does fit with the case Rousseau presents in his discussion of religious toleration for banishing those who deny the dogmas of civil religion. For he says not only that the case is based on the need for sociability, rather than a requirement of piety, but also that those who reject the dogmas of civil religion *cannot* be good citizens or loyal subjects who love the laws. He says that their loyalty is “impossible,” and presumably he thinks (as indicated in the passage cited earlier from *Poland*) that those who fail to love the laws cannot be expected to obey the laws: we do not have to wait on their disobedience, because we can be sure that it will follow. Right or wrong (and it is surely wrong) the argument respects the value of individual independence in two ways: first, and most clearly, by resting the case on a political interest in sociability rather than a religiously founded requirement of piety, and second, by accepting the burden of demonstrating the great importance of laws that restrict such independence: that they are not permitted unless they are important to advancing common interests—perhaps even necessary to such advance—and that the interests they advance are considerable.

In sum, the interest in self-development appears not to play any role in Rousseau’s account of the common good. The interest in individual independence, associated with the modern liberties, has an uncertain role, but several features of Rousseau’s view of the common good, its content, and its role in justifying laws, seem consistent with acknowledging a role for the interest in individual independence.

Baseline for the Common Good. I have said that advancing the common good is a necessary condition for an acceptable regulation, and that such advance consists in promoting common interests: the points on which the interests that form the social bond are in agreement. But this leaves three important questions open.

First, there are, of course, many ways to advance the basic interests of each, depending in part on where the “baseline” is set for deciding whether the conditions of each have been improved. That is, we have an improvement for each, and thus the promotion of the common good, only if (and arguably if and

only if) each person's interests are better served than they would be under some specified alternative condition: the problem of the baseline is the problem of saying what that alternative condition is.

Second, something needs to be said about what the test under contemplation is to be applied to: individual laws or whole systems of law, including expenditures as well as the taxes used to support them, for example.

And third, even if we fix the baseline, there are infinitely many ways to achieve improvements for all from the baseline. A strong pareto condition says that one outcome is better than another if it improves the conditions of each person (the interests of each are better satisfied). But this condition does not by itself select a distribution of those improvements (a point on the "Pareto frontier"). And it is not clear where Rousseau wants to set it.

I will start with the second and third of these issues, and then come to the first.

As to the target of assessment, then: Rousseau often speaks as though each regulation must work to advance common interests. But it is not clear why this needs to be so as a matter of principle rather than, perhaps, as a matter of political strategy—not clear, that is, why redistributive policies targeted on groups that are less well-off ought to be excluded as possible objects of the general will (even if they are not mandatory). If the system of law as a whole advances the common interests, why should it matter that some regulations have targeted benefits, but are paid for out of general revenues?

Consider a remark made with an evidently different concern in mind, but which bears on this issue: "Thus the law can very well state that there will be privileges, but it cannot confer them on any one by name; the law can create several Classes of Citizens, it can even specify the qualifications that entitle to membership in these classes, but it cannot nominate this person or that for admission to them" (SC 2.6.6) To be sure, it might be said in this case that the rationale for establishing the system of privileges under contemplation (as an example, consider the discussion of the Roman system of voting by *comitia* by centuries, SC 4.4.28–34) is that that system advances the common good, but that is not said, nor is it clear why an individual law that did not advance the interests of each, set within a system that does, violates the basic conception. To

underscore the point, let me return to the alternative account I noted earlier of what the general will commits us to: namely, to refraining from imposing on others conditions that we would not be prepared to accept for ourselves. A law that benefits groups that are less well-off at the cost of those who are better-off does not seem to fall foul of this condition, inasmuch as those who support the law could sincerely say that they were not imposing conditions on groups who are in any case better-off than they as less well-off would be unwilling to live by.

As to the issue of constraints on the distribution of advantage, beyond the pareto improvement requirement, it is difficult to see that Rousseau imposes such constraints. I see no anticipation of, say, Rawls's difference principle. To be sure, he says—apparently imposing a very demanding additional constraint—that “every act of sovereignty, that is to say every genuine act of the general will, either obligates or favors all Citizens equally” (SC 2.4.8). But it seems clear from what he says elsewhere—as in the passage cited just above about legally established systems of privilege—that he meant that every genuine act of the general will must treat citizens *as equals*, as of equal importance and equally worthy of respect, when it imposes obligations or confers benefits, not that it must confer equal benefits on them.

Coming then to the baseline issue itself: With respect to what baseline should the system of law generate improvements for each member? The force of Rousseau's conception of the supreme authority of the general will is that the general will itself is the foundation of all rights, of all claims that citizens can make against one another. Put otherwise, all such rights—including a right of independence, if there be such—are founded on the social contract itself; no rights within the political community are understood to have foundations prior to it; all are founded instead on the considerations of the common good that provide the supreme standard in the political society: “the total alienation of each associate with all of his rights . . . is made without reservation . . . and no associate has anything further to claim: For if individuals were left some rights, then, since there would be no common superior who might adjudicate between them and the public, each, being judge in his own case on some issue, would soon claim to be so on all, the state of nature would subsist and the association necessarily become tyrannical or empty” (SC 1.6.6–7). The force of this point—directed, I believe, against

Hobbes's idea that subjects reserve natural rights to preserve themselves—is brought out sharply by Rousseau's remark that the possessors of property in the society of the general will are "considered to be the trustees of the public good" (SC 1.9.6), and that "the right every individual has over his own land is always subordinate to the right the community has over everyone" (SC 1.9.7).

Consider, by way of contrast, the Lockean social contract. Locke's agreement to form a political society subject to the ultimate authority of the people is made against the background of a system of prior property rights, assumed to be established according to natural principles and pre-political conventions, in particular a convention to establish money, which enables individuals to increase their holdings without violating Locke's prohibition on letting resources go to waste. When Lockean individuals judge whether their association benefits them, they do so by considering whether the terms they are contemplating are better for them than the conditions they would have in a state of nature, with their different levels of property: some owning only their bodies, others owning large estates. So the different individuals will make different judgments about the proper terms of association, depending on where they stand in the prior scheme of property rights: some may favor a political arrangement controlled by owners of property, while others may reject such a system.¹⁴ Similarly, in Gauthier's modern version of the Lockean contract, the agreement of the contractors is modeled as a bargain, and the bargaining proceeds from distinct antecedent positions, which provide a basis for measuring relative improvements and the magnitude of the parties' relative concessions.¹⁵

Though Rousseau's account of how a community claims authority over a territory supposes that the members acquired rights to land antecedent to the community, those antecedent claims play no role in political argument within the community. The baseline for assessing improvements cannot be a system of unequal property rights prior to the social contract: "since the alienation is made without reservation, the union is as perfect as it can be, and no associate has anything further to claim" (SC 1.6.7). Correspondingly, the agreement cannot be thought of as a bargain from an antecedently defined position, because there are no prior rights that play a role within the organized political

society. So it seems natural to think of the baseline as a notional state of equality, with members having equal claims on existing resources.

Moreover, while Rousseau does not take distributive requirements to be built directly into the conception of the common good, he does think that there are other sources of such requirements. Thus, he urges that we can only sustain a scheme of political equality if we fix constraints on economic inequality. He indicates that equality is, along with liberty, an aspect of the common good (SC 2.11), and then adds: "with regard to equality, this word must not be understood to mean that degrees of power and wealth should be absolutely the same, but that, as for power, it stop short of all violence and never be exercised except by virtue of rank and the laws, and that as for wealth, no citizen be so very rich that he can buy another, and none so poor that he is compelled to sell himself" (SC 2.11.2). By way of clarification, he explains that it is important to "bring the extremes as close together as possible" (SC 2.11.2 n.). Or, as he puts it elsewhere, "the greatest evil has already been done when there are poor people to defend and rich people to restrain" (PE 19). "Is it in these two extremes, the one made to buy, the other to sell itself, that one should look for love of justice and the laws? It is by means of them that the State always degenerates: the rich man holds the Law in his purse, and the poor prefers bread to liberty" (LM 300). The limits on inequality that Rousseau mentions here are fixed not by the content of the general will itself, which requires general improvements from a notional baseline of equality, but by an account of the conditions required for the stability of the society of the general will.

In sum, then: Rousseau endorses a substantive understanding of the general will, according to which the general will, by its nature, is directed to the common good construed in a non-aggregative way. For this reason, he does not think that citizens need to be *protected from* the general will: "the general will is always upright and always tends to the public utility" (SC 2.3, 1.7.5). The supremely regulatory role of the general will advances the basic interests of each, taking a notional condition of equality as the baseline, while securing to members their independence in areas that do not bear on the common good: "the sovereign power does not and cannot exceed the limits of the general conventions,

and . . . everyone may fully dispose of such of his goods and freedom as are left him by these conventions" (SC 2.4.9). At the same time, there is some suggestion—though no more than that—that the protection of individual independence should itself be counted among the shared basic interests of the members of the society of the general will. The plausibility we are likely to find in the idea that the general will is always right will depend on whether that interest is counted as an ingredient in the common good, or as a fundamental good that must be balanced against the common good in assessing the justification of collective regulation.

And even if it is, we still must allow—as Rousseau does—that individuals may well need to be protected from the judgments of majorities—certainly of simple majorities, in which Rousseau does not put much faith (SC 4.2.11). For those judgments are at best imperfect expressions of the requirements of the general will: "it does not follow from it that the people's deliberations are always equally upright" (SC 2.3.1). But the general will itself—aimed at the common good—is their protection. So a central institutional problem is to ensure that actual deliberations and political outcomes reflect the general will—to ensure the supremacy of the general will, not to constrain its expression.

Priority Condition (GW₃)

The fact that an institution or law advances the common good provides citizens with a reason for supporting that institution or law, and they recognize that such reasons ought to take precedence over other reasons, in particular over those provided by considerations of individual advantage (reflecting the private will of individuals or groups within the political community), at least when they have reasonable assurance that others will act on such reasons. As citizens, they are not to rely on their private will in assessing regulations or deciding whether to comply, nor are they to balance considerations of personal advantage and of the common good. Instead, they are to deliberate about what the laws should be by reference to considerations of the common good (common interests), and to deliberate about conduct by giving first consideration to reasons of the common good, as articulated in laws that provide standards of coordination for the common

good. Because the notion of the common good reflects the idea of treating other citizens as equals by giving their interests the same weight in public judgments as one's own, the priority of reasons of the common good expresses the idea—fundamental to the genealogy of vice—of establishing conditions of association in which the equality of the members of the association is more fundamental than their inequality.

Thus, where Plato's civic unity requires a guardian class whose members have no identity apart from their role as philosopher-rulers of a well-ordered city, Rousseau's conception assigns to citizens a hierarchy of practical reasons, with supremacy assigned to reasons of the common good. In the society of the general will, we have, as I called it earlier, *civic unity through ordering* (the structure of practical reasons, with an ordering of reasons of different kinds) rather than *civic unity through integration* (that is, a lack of differentiation of reasons of different kinds). Correspondingly, civic troubles, for Rousseau, come not simply from the fact the citizens have separate interests or a sense of personal independence, but from the subordination of their commitment to the common good to their private interests or personal independence.

The idea of civic unity through ordering—of a *we* that is constituted through a common ordering of practical reasons—might, however, be interpreted in two different ways, more and less demanding on members in a free community of equals. Moreover, these different interpretations have different implications for individual conduct and, perhaps, for the plausibility of Rousseau's conception of civic unity.

According to the first, and less demanding, interpretation the idea is that members of the society of the general will occupy the role of citizen, along with other social roles; when they are called upon to play that role—for example, to make judgments about the right legislation—they are to base their reasoning and judgment on their view of the common good; and they are then to give priority to their judgments as citizen over their judgments about their own good. So if they are asked what the right tax rates are, their answer is to reflect their judgments about how best to advance the common good: for example, wealthier citizens should not oppose a progressive rate structure because of the additional burdens it would impose on themselves. Moreover, they are to regard the obligation to pay taxes, imposed through

decisions of the sovereign, as taking priority over their personal judgments about how best to use their private resources.

Correspondingly, on the less demanding interpretation, citizens fail to give proper priority to considerations of the common good when, in occupying their role as citizen, they make their decisions about what the laws should be and whether to comply with them on the basis of their private will. Thus, when the state is "close to ruin," a member "no more states opinions as a Citizen than if the State had never existed, and iniquitous decrees with no other goal than particular interest are falsely passed under the name of Laws" (SC 4.1.5). But—and here is the crucial point—"[e]ven in selling his vote for money he does not extinguish the general will within himself, he evades it. The mistake he commits is to change the state of the question, and to answer something other than what he is asked: So that instead of saying with his vote, *it is advantageous to the State*, he says, *it is advantageous to this man or to this party that this or that opinion passes*" (SC 4.1.6).

The second way to interpret unity through ordering is more demanding on members of a free community of equals. Here, the idea is that considerations of the common good are to have priority in practical reasoning over considerations of one's own good, *without adding the qualification that this pertains to the special role of citizen, and judgments made within that role*. On this more demanding view, the concern with the common good is to be more pervasive in the community of equals, in just the way that personal morality is meant to be pervasive in the lives of individuals, not confined to decisions made in some special role. This more demanding interpretation appears to be the force of the remarks cited earlier, from the discussion of *Poland*, about love of country becoming the dominant passion: there, the passion is dominant throughout one's life—in one's affections, attachments, and practical reasoning quite generally—and not simply in one's role as citizen, even if that role is itself understood as taking priority over others. On this more demanding interpretation, we have unity through ordering when, and only when, all members of the political community take the promotion of the common good of the community as their guiding aim, and use that aim as the basis for their decisions about all aspects of their life conduct, including for example their choice of occupation. Thus, "[t]he better constituted the

State, the more public business takes precedence over private business in the minds of Citizens. There even is less private business, because, since the sum of the common happiness contributes a greater share to each individual's happiness, he needs to seek less of it in his personal pursuits" (SC 3.15.3). But later in that same paragraph, Rousseau may suggest the weaker interpretation of unity through ordering: "As soon as someone says about affairs of State *What do I care?* the State has to be considered lost."

Here, once more, Rousseau seems to move between stronger and weaker expectations of community or collective attachment in a free community of equals. We saw such movement earlier in connection with the issue of whether private, particular interests are present in the society of the general will at all, and also in connection with the issue of the interest in individual independence. I don't think that there is a determinate answer to the question: what was Rousseau's considered judgment about the required strength of community attachment? But I do want to suggest a way to give some structure to the shifting conceptions: to understand why Rousseau seems drawn to weaker and stronger views about the nature and content of community attachment.

The starting point for the suggestion is to take notice of the location within *The Social Contract* of the passage making the strong demands of civic unity. Rousseau is there discussing the degeneration and death of the body politic, and what might be done to forestall it (beginning in SC 3.12, on "How the Sovereign Authority is Maintained"). His strong assertions about the need for attention to public business and the importance of common happiness as an ingredient in individual happiness are offered, I suggest (I will fill the point out later, in more detail), as devices to prevent the dissolution of the society of the general will: as conditions of the stability of a legitimate order, not as constitutive elements of the conception of legitimacy itself. The claims are elements of a political sociology, not of the philosophical conception of political legitimacy. In the case at issue, then, I suggest that the basic idea of unity through ordering is provided by the first, weaker interpretation: the priority of reasons of the common good is presented as part of a conception of the role of citizen and of the judgments that are to guide individual conduct in that role.

Rousseau may be attracted, then, to the stronger, more demanding conception of unity through ordering because—here is the interpretive idea—he thinks it may be *motivationally necessary* to ensuring the weaker one: that, for example, only a society whose members find their personal good in advancing the public good and are therefore *pervasively* concerned with the public good will, as a matter of motivational psychology, be able in a sustained way to give the priority to reasons of the common good that is ingredient in the general will. In particular, only those who are pervasively concerned with the common good will be able to act on reasons of the common good when those reasons make substantial demands on citizens—in the limit, when they demand that citizens give their lives to save the republic from invasion or ruin. In short, the stronger conception of unity through ordering is not an essential ingredient of the society of the general will, but comes from reflection on the conditions required for achieving such a society, in the face of certain kinds of human temptation.

Reasonable Confidence Condition (GW₄)

Citizens have reasonable confidence that the institutions conform to their shared conception of the common good, and those social institutions do in fact generally conform to it (suggested at SC 2.1.3). By “reasonable confidence,” I mean both that citizens believe that the institutions conform to their conception of the common good, and that their belief is not simply a matter of faith but one which they could support on the basis of evidence and reasonable inference. Think of GW₄ as a *non-ideology condition*. The society of the general will is not merely an association with a shared commitment to the common good, but an association that operates in ways that are responsive to that shared commitment, and whose members have reasonable confidence about that responsiveness.

Together, then, conditions GW₁–GW₄ explicate the notion of a social association regulated by a general will, thus of a free community of equals: the kind of association that would, Rousseau argues, issue from the initial agreement. The central idea is that in a society regulated by the general will, supreme authority rests with an understanding of the common good shared among citizens: citizens acknowledge that considerations

of the common good provide the principal reasons for action, and are willing to act as those reasons require, on condition that others do so as well. They are prepared to act for the common advantage, under circumstances in which the general failure to so act would yield suboptimal outcomes.

The ideal is a political community—a *we*, a people—unified by a shared understanding of the common good, defined in terms of the common interests of the members, and a shared willingness on the part of citizens to place considerations of the common good above other considerations, and in particular considerations of personal advantage, with well-founded confidence that the society conforms to their understanding. It is a free community of equals: a *community* because of the shared allegiance to the common good; a community of *equals* because the content of the idea of the common good reflects a commitment to treat other members as equals by giving their interests equal weight along with one's own; and a *free* community of equals because the members, assumed to endorse the common good as the basis for legitimate law, have their own will as a rule. Thus, "The better constituted the State," Rousseau says, "the more public business takes precedence over private business in the minds of Citizens" (SC 3.15.3). By contrast, "As soon as public service ceases to be the Citizens' principal business, and they prefer to serve with their purse rather than with their person, the State is already close to ruin. Is there a call to battle? they pay troops and stay home; is there a summons to Council? they name Deputies and stay home. Finally, by dint of laziness and money they have soldiers to enslave the fatherland and representatives to sell it" (SC 3.15.1).

3

Reflections on the General Will's Sovereignty

Solving the fundamental problem requires a political community regulated by a shared understanding of the common good. Only then, in the society of the general will, can people both be assured the protection of their person and goods, and express the freedom that belongs to their nature. I have not yet discussed the reasoning that leads Rousseau to this conclusion. I will come to that at the end of this chapter. First, however, I want to explore the idea of the society of the general will more deeply.

Groups, Sovereignty, Consensus, Majorities, and Rights

I have described, in very general terms, the ideal of the society of the general will. To clarify the essential idea, I will take up five questions about the ideal:

1. What is it for a group to have a will?
2. Is sovereignty in Rousseau the authority of a person, like the Hobbesian sovereign, or of a principle?

3. Does the ideal of a free community of equals assume too much agreement?
4. How and why can the majority serve as the people's tribune?
5. Does the central role of the notion of the common good push rights out of Rousseau's conception?

Group Wills

My explication of the notion of a general will is founded on a general account of what the will is, focused on reasons for action. Thus, having a will is having the capacity to act for reasons. This underlying idea enables us to understand such otherwise puzzling features of Rousseau's view as that individuals have multiple wills, and that groups have wills. And the understanding it provides fits with the methodological individualism that Rousseau appears to endorse, and enables us to understand how a person could be free—having his own will as a rule—when he complies with laws reflecting the general will.

Thus a person has different wills when he/she uses different reasons for ranking social outcomes and, by implication, the policies that bring them about. Rousseau says, for example, that the magistrate in a well-ordered state has "three essentially different wills": the magistrate might will, so to speak, *as an individual*, or *as a magistrate*, or *as a citizen*. What we need is an account of "as an X." The idea is that the magistrate's conduct might be guided by three distinct orderings of policies and social states, each based on a distinct consideration: an ordering reflecting particular interests, an ordering reflecting interests shared with members of the executive, and an ordering reflecting an attachment to the common interest (SC 3.2.5). The case of the magistrate with three essentially different wills is really an instance of the more general case. Generally speaking, citizens will have a particular will founded on personal interests, a corporate will (or several corporate wills) founded on the shared interests of the members of a group within the wider community, and general wills founded on the shared interests of the members of the political society. Given the different interests that provide the basis for distinct reasons and thus found these different wills, an individual will have a private will that is "contrary to or differing from the general will he has as a Citizen" (SC 1.7.7). The fact that individual members of a group with a general will

can themselves each be said to have a general will—because they are assumed, as citizens, to take considerations of the common good as reasons—will be important in understanding what Rousseau has in mind when he says that those who act on the general will are acting on their own will, and thus are free and self-governing.

As to the notion of a group will, when a collection of individuals meet the Common Good and Priority Conditions (GW₂ and GW₃)—where they each assign priority over the reasons provided by their own particular interests to reasons provided by the shared interests of that collection—there seems to be a perfectly straightforward and intelligible sense in which the set of individuals itself has a single will. In particular, the members of the set each rank social states on the basis of the same reasons; they are motivated by those reasons; and there is common knowledge that both of these previous points are true. Under these conditions, it seems unproblematic to treat the collection as a single agent. For a single system of aims coordinates and guides and is known to coordinate and guide the actions of each member.

Hobbes and Rousseau on Sovereignty and Law

A second way to clarify the notion of the general will is to characterize more specifically the nature of the generality of the will and the consequences that flow from the requirement that it be suitably general. To provide such characterization, I want to contrast Rousseau's notion of a sovereign general will with Hobbes's conception of sovereignty and authority.

To begin with, then, Hobbes and Rousseau share at least four ideas: both think that we can represent a political society or body politic as a single person—as an agent with a will—thus as a *we* formed from a collection of *Is*, as a people. Moreover, what makes the collection of persons a *we* is that the single will of the political society is authoritative: it takes precedence over the private wills of its members. Furthermore, the content of that authoritative will of the body politic—the will of the people, *our* will—is given by the content of the sovereign's will. And finally, legitimate law is the expression of the sovereign's will.

But the differences are far more fundamental than these four abstract, common points.

The essential idea in the Hobbesian state is that each person subordinates his/her will to the particular will of a determinate individual, the sovereign. Whether that sovereign is an individual or an assembly, each subject agrees to accept the sovereign's commands—imperatival expressions of the desires of the person who holds sovereign power—as providing reasons for action that preempt his/her own private reasons, whatever the content of those commands (more or less). We have a political society just in case we have such common subordination to a determinate agent. The *we* is constituted by common submission to a determinate will: "A Multitude of men, are made *One* Person, when they are by one man, or one Person, Represented; so that it be done with the consent of every one of that multitude in particular. For it is the *Unity* of the Representer, not the *Unity* of the Represented, that maketh the Person *One*. And it is the Representer that beareth the Person, and but one Person. And *Unity*, cannot otherwise be understood in a multitude" (Lev. 114).¹

To be sure, Hobbes's discussion of the true liberties of subjects (Lev. 150–2) imposes a condition on that otherwise-unconditional acceptance: because the desire for self-preservation is, Hobbes says, the fundamental human desire, no one is assumed to alienate his right to self-preservation. Self-preservation being the most important good, the liberty to take measures needed to ensure it cannot be relinquished. Thus the "true liberty of a subject" comprises "the things, which though commanded by a sovereign, he [the subject] may nevertheless, without injustice, refuse to do" (Lev. 150). In particular, if the sovereign commands me to kill myself, I am not obliged to follow that command, nor am I obliged to hurt myself, "to abstain from the use of food, ayre, medicine, or any other thing" (Lev. 151) required for life, nor to serve in the military, at least if I can find a substitute. I have a basic liberty to preserve myself, a permission even in the face of the sovereign's will to the contrary. But this exception underscores the rule: for if the sovereign commands someone else to kill herself, I am not for that reason released from my obligations of obedience to sovereign and law. The command exempts the person to whom it is directed, but does not challenge the legitimacy of the sovereign's authority, or the obligations of subjects generally to obey. Put otherwise, the liberty in question is merely an obligation-limiting permission, not a claim-right to

which obligations of others correspond. So others—including the sovereign—have no obligation to refrain from infringing on the liberty.

In effect, then, Hobbes's conception of the state is organized around the idea of a supreme *particular* will—the supremacy of the will of an identifiable agent. What this means is that there is no requirement of *generality in the content of the agent's will*. That is, Hobbes imposes no requirement, at the level of norms or institutions, that the sovereign's will aim at the common good. The Hobbesian state—the people, the we, the commonwealth—is constituted by the common and reciprocal acceptance of the authority of a single agent, whatever the content of that agent's will. Put otherwise, we have a way to identify the sovereign and the legitimate acts of the sovereign to which we are subject and which we are bound to obey, quite apart from the content of what the sovereign wills. Hobbes imposes *no content constraints* on an act's being an expression of sovereignty—no necessary conditions, or even relevant considerations, as to the content that the acts must meet to be interpretable as acts of the sovereign—and therefore no content constraints on laws or legal obligations. Sovereign acts are all about *source*, and not at all about *content*. When we know that a regulation has its *source* in a command—in a public expression of will—by the agent who is entrusted through common submission to make the laws, then we know that it is a law and that it ought to be obeyed. In this very general sense, Hobbes is a legal positivist: we have only source conditions on legal validity; content requirements are not necessary conditions for legal validity, or even part of a test for legal validity. Thus “no law can be unjust. The law is made by the sovereign power, and all that is done by such power, is warranted, and owned by every one of the people; and that which every man will have so, no man can say is unjust” (Lev. 239).

To see the full force of the point, consider Hobbes's distinction between good laws and just laws. “To the care of the sovereign belongeth the making of good laws. But what is a good law? By a good law, I mean not a just law: for no law can be unjust. The law is made by the sovereign power, and all that is done by such power is warranted, and owned by every one of the people; and that which every man will have so, no man can say is unjust” (Lev. 239). Good laws, then, are laws that are *necessary* and

perspicuous. A law is necessary when it is required to advance the “safety of the people,” where their safety is understood as comprising both preservation and “all other contentments of life, which every man by lawful industry, without danger, or hurt to the commonwealth, shall acquire to himself” (Lev. 231). Laws that are needed for the safety or good of the people are not arbitrary impositions on liberty, but instead fulfill the true purpose of law, which is not to keep people from acting, “but to direct them and keep them in such a motion, as not to hurt themselves by their own impetuous desires, rashness, or indiscretion” (Lev. 239). Good laws, in short, are commands issued by the sovereign that restrict liberty for the general welfare.

Moreover, laws are perspicuous when they are briefly stated, and promulgated along with the reasons for them, which “shewes us the meaning of the legislator” (Lev. 240). Indeed, once the purpose is expressly stated, we only need brief laws, because we know the guiding purpose and can recur to it in determining the proper interpretation.

Suppose, then, that we want to know whether a law is good or bad. As the nature of needful laws indicates, we have to know the law’s content, and not simply its source. In particular, we have to know whether it advances the general welfare. Knowing that plainly requires an understanding of what the law requires, and not simply about who made it, not simply the source.

But the justice of laws, on Hobbes’s account, is an entirely different matter. Any command of the sovereign establishes a just law. We do not need to know the content to know whether it is just, or whether it is a law: we only need to know the source. When subjects accept the authority of the sovereign, they do so unconditionally, without limiting that acceptance to the sovereign’s good laws: “and that which every man will have so, no man can say is unjust.”

To be sure, the sovereign has an obligation under the natural law to make only good laws. But that obligation is owed to God, and only God is to hold the sovereign accountable for fulfilling it. Moreover, making good regulations is an important part of the official responsibilities of the sovereign. But, to reiterate the central point, a regulation’s goodness is no part of the test of its *validity as law*. More generally, the validity of a regulation—its standing as valid law—is, as the goodness/justness distinction

underscores, entirely independent of its content. Of course, if we could only identify a command as the command of the *sovereign* by relying on a content test, then we would still have a content test for legal validity, even if all commands of the sovereign, without any further requirements on content, were valid laws. But on Hobbes's theory of sovereignty and law, the content of a regulation—and a fortiori its goodness—has no bearing on whether it is genuinely an act of sovereignty.

Rousseau, in sharp contrast, rejects the identification of the sovereign with any natural person or collection of such persons, including the collection of all members of a political society. What each member accepts as authoritative and obligation-imposing is not the will of any individual or collection of individuals, identifiable independently from the content of the will of that collection. Instead sovereign authority lies in effect in the shared understanding of the common advantage in which the interests of each are taken into account. This is the force of Rousseau's remark that the sovereign is not a determinate individual or a determinate collection of individuals, but a "collective being" (SC 2.1.2), a *we* that is constituted by a collection of persons who share an understanding of the common good and accept the authority of that common good in matters of collective decision, not a collection united in submission to a single will.

The general will is *general*, then, not simply in that it comes from all, acting on their judgment of the common good, but because it is by its nature—that is, by the terms of its construction and authorization—directed to the advantage of all: its content is intrinsically general; the enacting will is general, as is its matter. So just as we might say that a piece of behavior is an action of an agent only if we can interpret the behavior as enacting a coherent scheme of values and beliefs, similarly we can say that a regulation is an act of the sovereign only if we can interpret it as enacting a coherent conception of the common good of those who are subject to it: only if guided by a conception of our good, as the good of persons who are moved by self-love. We have no way to identify an act of sovereignty—and therefore no way to identify a regulation as a law—apart from such an interpretation: content conditions help to distinguish an act of the general will from an expression of the will of all. Because legal validity is in this way content-dependent, Rousseau is not a legal positivist.

This requirement of generality implies, as I noted in the earlier discussion of the right of independence, that the general will is by its nature limited in what it can will: limited, in particular, to regulating conduct in ways that are understood to advance the common good. "The right which the social pact gives the Sovereign over the subjects does not . . . exceed the bounds of public utility," and therefore—Rousseau adds in a footnote quoting d'Argenson—"everyone is perfectly free with respect to what does not harm others" (SC 4.8.31). The sovereign cannot arbitrarily—that is, without reason—infringe on personal liberty: that much is hard to dispute. What matters is the standard of arbitrariness, the account of acceptable reasons. Rousseau tells us that "worldly well-being [is] the unique object of legislation" (LM 148), not, for example, piety or salvation: "the science of salvation and that of Government are very different" (LM 149). More particularly, Rousseau here gives the standard as the common good. The sovereign cannot make regulations that lack a justification in the notion of the common good. It is therefore bound by its own construction to provide some protection of personal liberty, even if there is not a fundamental interest in personal independence. And if there were understood to be such an interest, then the protection would correspondingly be stronger: political argument would need to show an important connection between regulations and the common good before such regulations would be acceptable.

In either case, whether individuals are assumed to have a fundamental interest in personal independence or not, we get a version of what Juergen Habermas has called "the internal connection between popular sovereignty and human rights."² That is, we have the result that at least some minimal assurance of personal autonomy is built into a system of popular sovereignty. It is not built in in the way that rights to discuss and vote are: those must be in place if the process of collective decision-making is to treat citizens as the equal authors of laws. But at least some rights of private autonomy are presupposed because the outcomes or results of the exercise of popular sovereignty can extend no further than the common good, understood in terms of the basic interests of each, and such reasons for restricting conduct as impiety are excluded because they pertain only to conviction, not to conduct that bears on the common good.

To put the point once more in terms of the contrast with Hobbes: Rousseau's sovereign is confined by the terms of its construction, which defines the sovereign's nature, to what Hobbes calls good laws; that's because we have an exercise of sovereignty only when collective power is guided by the general will (SC 2.4.1), and because we, therefore, cannot identify a sovereign act except by reference to a determination of such guidance. An act of the sovereign must be something that we (can) collectively endorse; but as equals, each concerned with his or her own good, we can only endorse the exercise of collective power when it is guided by the common good. So the limits on the sovereign's authority are not given "from the outside"—say, by a higher authority, or by moral laws or principles with which the people regard itself as bound to comply—but by the intrinsic nature of sovereignty as constructed through the social compact. And the confinement of sovereign acts to good laws implies at least a limited protection of personal liberty.

To be sure, this leaves a large question: how to ensure, through political institutions or culture, that collective decisions do indeed remain within the bounds of legitimate authority, as those bounds are established through the social compact. But it is question that Rousseau must address because his social compact does not simply authorize the exercise of power by some determinate individual or set of individuals, but authorizes such exercise only for the purposes for which the social compact is made: namely, the common good (SC 2.1.1).

Unanimity?

Rousseau's view assumes important elements of unanimity or consensus in the society of the general will, and this may raise serious troubles for his view: after all, general consensus seems too implausible, even in relatively small groups. But the degree of implausibility and corresponding extent of those troubles depend on just how much agreement he is supposing, or needs to suppose, to make the account of the society of the general will work. Two sorts of disagreement should be noted, as compatible with the idea that the political society is regulated by a supreme general will aimed at the common good, or by the fundamental commitment to treat others as equals by only imposing burdens that one is prepared to live under oneself. To be sure, there are

other sources of political disagreement, but the ones I mention here are distinguished by being fully compatible with the idea that citizens keep faith with the social compact.

First, agreement on the ends of association—the common good—is consistent with considerable “factual” disagreements about how best to advance those ends. Such disagreements can run quite deep, extending, for example, to judgments about the appropriate form of government—there being no answer to that question that applies across all circumstances—or about which collective decision rules provide the most reasonable balance of deliberativeness and efficacy (SC 4.2.11), or about the level of revenue needed by the government to carry out its business, or about which goods are necessities and therefore to be exempted by taxation (LD 114).

Second, people who broadly agree on a conception of the common good may nevertheless disagree about how exactly to interpret the conditions that define the common good or are presupposed by it. Here the role of public deliberation is in part about how to interpret the terms of the agreement. For example, I indicated earlier that the notion of the common good is understood as barring the existence of material conditions in which people are forced to sell themselves. But it is a familiar fact that there are reasonable disagreements about just what sorts of material circumstances force people to sell themselves—disagreements that trace in part to disputes about how to understand “force” and “sell.” If, say, a person is forced to sell himself when the person has no reasonable alternative to selling himself, then disagreements about *reasonableness* will translate into disagreements about *forced*. Thus there might be debate about whether working for a wage is selling oneself, about the level of material well-being below which people will be forced to sell themselves, and about how to ensure that people do not fall below that level.

Consider a second example. Citizens are to be left at liberty, not subject to law, unless such subjection is “important” to the community. “It is agreed that each man alienates by the social pact only that portion of his power, his goods, his freedom, which it is important for the community to be able to use” (SC 2.4.3). Earlier, we discussed this passage in connection with Rousseau’s account (or lack of account) of individual independence. Here I wish to draw attention to the remark that follows: that “the Sovereign is alone judge of that importance.” Let’s read

“alone judge” as “authoritative judge.” Thus, I do no wrong in making a judgment, but only in acting on it, when it contravenes the public judgment. So the sovereign people are alone judge—that is, authoritative judge—of whether, for example, the benefits flowing from an increase in public expenditures are sufficiently important—meaning, sufficiently important to advancing common interests—to justify the greater restrictions on liberty that follow from the increased tax rates required to fund the expenditures; or the benefits from mandatory public service are sufficiently important to advancing common interests to justify the restrictions on liberty required by such service; or the benefits from anti-loitering legislation are sufficiently important to advancing common interests in personal safety and security to justify the restrictions on personal liberty that come with the enforcement of such legislation. On each of these three issues, we can expect disagreement among citizens, even if all citizens accept that regulations ought only to be imposed if they are sincerely understood as important to advancing common interests, agree about what the interests are that figure in the account of common interests, and agree about all the effects of the law on those interests and on the regulated activities. Even if they do agree, they may still disagree about whether the gain in, for example, public safety, is of sufficient importance to justify the burdens imposed by the regulation, perhaps because they disagree about the relative importance of interests in personal safety and personal liberty.

In short, the unanimity required by the general will seems compatible with substantial political disagreement, and that observation should limit concerns about unreasonable demands of consensus. That said, I hasten to add that Rousseau himself says very little about disagreements among citizens with conflicting interpretations of the demands of the general will: he says little about the sources of such disagreements, their extent, or the means for resolving them. Indeed, he appears to hold the view that disagreement itself is a sign of degeneration: “So long as several men united consider themselves a single body, they have but a single will, which is concerned with their common preservation, and the general welfare. Then all of the springs of the State are vigorous and simple, its maxims are clear and perspicuous, it has no confused, contradictory interests, *the common good is everywhere fully evident and requires only good*

sense to be perceived" (SC 4.1.1, emphasis added). Under such conditions, everyone can sense when new laws are needed and what they should be, and all anyone needs from others is the assurance of such common recognition.

"But", he goes on, "when the social knot begins to loosen and the State to weaken; when particular interests begin to make themselves felt, and small societies to influence the larger society, the common interest diminishes and meets with opposition, votes are no longer unanimous, the general will is no longer the will of all, contradictions and disagreements arise, and the best opinion no longer carries the day unchallenged" (SC 4.1.4). Similarly, he says that "The more concord reigns in assemblies, that is to say the closer opinions come to unanimity, the more the general will also predominates; whereas long debates, dissensions, disturbances, signal the ascendancy of particular interests and the decline of the State" (SC 4.2.1).

To be sure, Rousseau does not say here that votes fail to be unanimous *only when* particular interests intrude, but rather that such intrusion is sufficient for division. Nor does he say that the absence of agreement on what the common good requires is incompatible with the regulative role of the general will. Moreover, these critical remarks about disagreement might be understood as making the plausible point that the existence of considerable division and conflict in views about the substance of the common good and how best to advance it fosters suspicion that people are masking particular interests under the guise of the common interest (all the more when their interpretations of the common interest and their judgments of importance correlate with their particular interests), and not the implausible and objectionable point that such masking is the only source of disagreement.

The fundamental idea of the society of the general will does not, then, exclude political disagreement of either of the two kinds just noted—either about the best means for advancing the common good, or how best to interpret its constitutive elements. Still, Rousseau is generally suspicious about political disagreement, suspicious because disagreement at least *suggests* (even if it does not imply) an intrusion of private interests into judgments about how to advance the common advantage, an intrusion that signals and fosters political degeneration. And he presumably expects political disagreements to

provoke comparable suspicions among citizens. Rather than seeing disagreements in political judgment as expressing legitimate differences of opinion, reflecting differences of circumstance, information asymmetries, or the complexities of political reasoning, citizens themselves (not just the suspicious theorist) will see such disagreements as reflecting private or group interests. To be sure, they will not openly acknowledge the aim of using collective power to advance individual or group interest. Instead, they will disguise their interests as expressions of sincere concern for the common good. Referring to Frenchmen, Germans, Spaniards, and even Englishmen, he says that “all will declare themselves disinterested and be cheats; all will speak of the public good and think only of themselves; all will praise moderation and wish to be Croesuses” (P 184). The presentation of private wills as general wills—as sincere convictions about the common good—is a kind of all noise, no signal, cheap talk: no evidence against the hypothesis that disagreements reveal illegitimate private concerns, not conscientious disagreements. After all, if you are trying to persuade others, you of course claim to be incorporating their good within your concern. So the fact that you say you are is no evidence that you are.

Suppose we had little evidence outside politics that people have separate and opposed interests, or compelling evidence of powerful, overriding, and encompassing attachments to a national-political community. Then we would have less reason to suspect that political disagreements are ill-founded. Perhaps this gives yet another reason for the stronger form of civic unity described earlier, in which individuals are pervasively concerned with the common good, to which they give priority throughout their lives (and not only in the ways required by the Priority Condition). Such pervasive concern might reduce political disagreement, and—what is immediately at issue—would help to put such political disagreement as remains beyond suspicion. Each of us would have good reason to think that others have a genuine concern for the common good, apart from the special settings of political debate with their evidentially weightless announcements of concerns for the common good: evidentially weightless because, once more, we have good reason to announce such concerns in making our public arguments even if we do not have them. But suppose that others exhibit that concern throughout their lives, in their everyday interactions, thus revealing the

authenticity of the political expression of that concern (anyway, they give us more evidence of its authenticity). This pervasiveness would help lay to rest our suspicions about the appeals to the common good. Once more, concerns about political stability would result in more demanding requirements of community: but this time for epistemic reasons, rather than for motivational-psychological reasons.

To be sure, we might not wish to live with the pervasively high-minded, or to live pervasively high-mindedly. And it seems like a very high price to pay in order to lay to rest the doubts of our fellow citizens about our political sincerity, or our doubts about theirs. But that is another matter (to be sure, an important matter, but still another matter). Here I mean only to draw attention to the interesting idea that a stronger and more pervasive form of community might be attractive not because it *eliminates* political disagreements, as if disagreements as such were intrinsically illegitimate, but because it helps make it possible to *have* political disagreements without the destructive suspicions that commonly accompany them. A strong form of civic unity would be recommended not as an essential feature of the society of the general will, say as part of the Common Good Condition (GW₂), or as supporting the Priority Condition (GW₃), but as a requirement needed to support the Reasonable Confidence Condition (GW₄).

Majority Rule

In a society regulated by a general will, Rousseau says, each person achieves the full political autonomy of giving the law to him- or herself. This requirement is obviously very demanding, more demanding than the familiar (and important) democratic idea that those who are subject to the laws ought to have fair chances to participate in the processes from which those laws issue, and that the results of such fair processes of argument and collective decision should be authoritative. Rousseau has this idea, too: "The People subject to the laws ought to be their author; only those who are associating may regulate the conditions of the society" (SC 2.6.10). But he also has the stronger demand of full political autonomy: that subjects in the society of the general will "obey no one, but only their own will" (SC 2.4.8). Because of this allegedly self-governing character of an association regulated by a general will, such an association

achieves the reconciliation of protection and freedom required to resolve the fundamental problem.

But Rousseau also emphasizes that coordination for the common good requires general laws that apply to citizens. Moreover, he is fully aware that people will sometimes disagree about which laws are best, at least for the two non-objectionable reasons canvassed in the discussion of the limits of unanimity. Acknowledging such disagreement, he holds that majority rule is the best way of settling what the laws should be—though the majority in question will often not be a simple majority (SC 4.2). Indeed, he defends supra-majority requirements on important political questions.

Majority rule certainly has important attractions: for example, it is arguably a fair system of collective choice, because of the ways it treats people as equals in assigning them equal influence over collective decisions. But whatever its attractions on grounds of fairness, there appears to be a considerable difference between giving the law to yourself and getting it from the majority. Suppose I lose in a collective decision. How is it, then, that I obey only my own will when I obey the law supported by the majority? Indeed, even if I vote in favor of legislation that eventually wins, that legislation may itself be the product of compromise and bargaining, and not reflect my own judgment of what the best solution would be. So even if I am in the majority, compliance to the law can hardly be construed as obedience to my own will, except in a highly attenuated sense.

Rousseau is acutely aware of this difficulty, of the tension between compliance with laws as majority decisions, and compliance with laws as obedience to one's own will. Thus he acknowledges that "Except for this primitive contract [the social contract], the vote of the majority always obligates all the rest." But then, he adds: "the question is raised how a man can be both free and forced to conform to wills that are not his own. How are the opponents both free and subject to laws to which they have not consented?" (SC 4.2.7). Rousseau's response is that "the question is badly put," put badly because it assumes that the minority has not consented to the laws. But—and here it is important to quote the passage in full—in fact:

The Citizen consents to all the laws, even to those passed in spite of him, and even to those that punish him when he dares to violate any one

of them. The constant will of all the members of the State is the general will; it is through it that they are citizens and free. When a law is proposed in the People's assembly, what they are being asked is not exactly whether they approve the proposal or reject it, but whether it does or does not conform to the general will, which is theirs; everyone states his opinion about this by casting his ballot, and the tally of the votes yields the declaration of the general will. Therefore when the opinion contrary to my own prevails, it proves nothing more than that I made a mistake and that what I took to be the general will was not. If my particular opinion had prevailed, I would have done something other than what I had willed, and it is then that I would not have been free. (SC 4.2.8)

It is not clear how best to interpret these remarks, but they are essential to Rousseau's idea that we can combine political authority with self-legislation, on the reasonable expectation that there is not unanimity in the community. When a person votes "his opinion," then, on whether a proposal does or does not "conform to the general will," what precisely is the opinion about? Apparently, it is about some matter of fact: that is the force of saying that the vote expresses an opinion. But what matter of fact? And, correspondingly, if the person loses in that the majority votes otherwise, what matter of fact should he conclude that his opinion was wrong about?

A first interpretation—call it the procedural interpretation—might be that the opinion is, in essence, a prediction about the result of the vote: let's say a vote about a general rule, where the vote is assumed to take place under reasonably favorable conditions of good information and absence of factional division (SC 2.3.3). I call this the procedural interpretation because the idea is that the general will consists in—is fixed by—the result of a procedure of voting under conditions of good information and non-factionalization: as Rousseau puts it, under the conditions just described, "the general will would always result from the large number of small differences, and the deliberation would always be good" (SC 2.3.2). Or elsewhere, "the tally of the votes yields a declaration of the general will" (SC 4.2.8). I have presented a more substantive interpretation of the general will, focused on an idea of the common good, but it is worth exploring the implications of the procedural view for the passage in question.

Before going any further with a discussion of the procedural interpretation, however, I want to digress briefly about an

issue to which I will return later, but which requires attention now because it emerges in the passage about the importance of good information and the absence of factional division as background to a vote that yields a declaration of the general will. In that passage, Rousseau adds that there is to be no communication among citizens: with no communication and adequate information, "the general will would always result . . . and the deliberation would always be good" (SC 2.3.3). I have omitted mention of the no-communication condition from my sketch in the previous paragraph and think it is commonly misconceived. That there is something odd about the condition should already be clear from the tension between an assumption of adequate information and an assumption of no communication: how, as a practical matter, could citizens be adequately informed in the absence of communication? More to the point, in the sentence immediately following the one in which he indicates that the general will would emerge from deliberation by citizens without discussion, he begins: "But when factions arise . . .", and proceeds to express concerns about group divisions, and particularly about the emergence of a majority faction ("one of these associations is so large that it prevails over all the rest"). So it seems clear from the context—and from the contrast signaled by "But when factions arise . . ."—that his concern is not with communication as such but with factionalization. This point is immediately highlighted when Rousseau goes on to insist that, in making collective decisions, every citizen should "state only his own opinion" (SC 2.3.4). Here, too, Rousseau is best understood as expressing hostility not to communication but to factionalization, which leads individuals away from expressing their own judgments, and pressures them instead to express the opinions of the faction to which they belong, which arguably implies a reduction of information.

To conclude the digression: Rousseau never criticizes political discussion. Indeed, near the beginning of Book 4 he says the following: "I could offer quite a few reflections here on the simple right to vote in every act of sovereignty; a right of which nothing can deprive Citizens; and on the right of voicing opinions, proposing, dividing, discussing [motions], which the Government always takes great care to allow only to its own members; but this important matter would require a separate treatise, and I cannot say everything in this one" (SC 4.1.7).

There is no suggestion here that the government is right to confine discussion in these ways; to the contrary, the right of citizens to voice opinions and discuss is located on the same plane as the right of citizens to vote. And as we will see later, when, in his *Letters Written from the Mountain*, he directly confronts the Genevan government's practice of regulating discussions in citizen assemblies, he condemns that practice as a basic violation of popular sovereignty.

Let's return now to the procedural interpretation. Suppose we say that a vote under conditions of good information and non-factionalization defines the content of the general will on the issue under consideration. Then the voting citizen would, through the vote, express his opinion about whether a proposal conforms to the general will by voting his prediction about what the result of the collective vote will be. Thus, I would vote in favor of increased salaries for the police if I believed that the result of the vote on increased salaries would be positive, and I would otherwise vote against. Adopting this procedural interpretation, we have a straightforward account of the idea that when the vote goes against me, I am thereby shown to be mistaken. I thought the result would be favorable, so I voted in favor. The result was negative. So I was wrong. Everything fits. But the view of voting is perverse: it is perverse to think of individual votes on a question as predictions of the outcome of the collective vote. Individual votes cannot be understood as, in this sense, opinions about the content of the people's collective will.

Consider then the alternative to the procedural interpretation that I presented earlier. I suggested that the general will wills the common good, according to some shared understanding of that good, guided by an account of the basic interests of members and their equal standing; and that though the understanding is shared, there can be disagreement both about the content of the common good and how best to advance it. Against this background, we can make reasonably straightforward sense of Rousseau's point about the right question, and in particular his essential distinction between expressing approval or rejection of a proposal and expressing an opinion about whether a proposal conforms to the general will, "which is theirs."

Thus, think of the problem of majority rule as follows. Assume a group of people, each of whom wants above all to advance the

common good of the group, defined in terms of the good of the members of the group. Assume as well that people decide how to vote by first asking themselves: Does this law advance the common good? They are not addressing the question: Do I like this law? Or is this good for me and those in my group? Much less are they asking the question: Is this the law that will win the support of most of us? Thus, in voting, citizens do not simply express their preferences about what the laws should be: that would be to express approval or disapproval. Instead they express their opinions, their beliefs about whether the proposal does indeed advance the common good, knowing that those opinions will differ across citizens. Thus understood, a collective choice does not aggregate preferences but pools opinions about a question to which there is a right answer.

Against this background, we can think of Rousseau as claiming, then, that the judgment of the majority about how to advance the common good is more likely to be right than the judgment of the minority. So if I vote my opinion, and the majority supports an alternative, then that support itself gives me some reason to think I was wrong. Moreover, since I wish that the law we pass is best suited to advancing the common good, I don't want my own proposal to be accepted, not after I have the additional evidence provided by the outcome of the vote. Previous to the vote, I thought my proposal was the best way to advance the common good. But the vote itself gives me new evidence, not simply evidence about what other people want, but about the best way to advance the common good. Given my overriding concern to advance that good—expressed in the Priority Condition—and the new evidence, I no longer should want my proposal to win. Given what I now know, I can say, in retrospect, that if my proposal had won I would have ended up acting on laws that are not best suited to promoting the common good, and so "I would not have been free" (SC 4.2.8).

So far this is simply interpretation, but it does provide a way to refine Rousseau's question about majority rule. Thus, we want to know why we should have the confidence in collective intelligence required by the argument? Some support is provided by what are called the "Condorcet jury theorems."³ The idea of the jury theorems is to model collective decisions generally on decisions by juries. The fundamental idea is that with juries, there is a right answer (innocent or guilty), and members of the

jury are to vote their opinions about the answer. Generalizing, we are to attribute a “judgmental competence” to the members of a group, where this competence is understood as the probability of an individual member’s getting the correct answer to a question (e.g., “Is this person guilty or innocent?”). The jury theorems then assess decision rules by considering the probability that a group of individuals, each with a certain judgmental competence, will arrive at a correct answer by using those rules. So, for example, it turns out that if there are two alternative policies, and the average individual competence on some question is greater than 0.5, then the probability that a simple majority will get a correct answer to the question is greater than 0.5; moreover, that probability increases very rapidly as the size of the group increases, as the size of the majority increases, and as individual judgmental competence increases.

The jury theorems provide some reason for confidence in collective intelligence about the common good, on condition, of course, that individuals are good judges of what advances the common good. If average competence is below 0.5, then, as the size of the group increases, group competence approaches 0 just as rapidly as it approaches 1 if average competence is above 0.5. And using this Condorcetian perspective to account for Rousseau’s defense of majority rule underscores that Rousseau’s conception of democracy is not majoritarian. A reasonable order advances the common good—it operates for *the people*, not simply for the majority. The role of the majority is to identify the rules that are for the good of the people as a whole, including those rules that ensure the liberties of the people.

But the Condorcetian argument is very limited, in several ways. First, the theorem requires that individual judgments be independent. But if people are talking to each other the judgments do not meet that condition: we have in effect fewer draws from a sample. That is why it is important both that people be well informed and that they do not form into factions.

Second, it is not clear what the basis for assigning judgmental competence to individuals is, or how we might revise such assessments in light of experience of past choices. Someone who consistently is in the minority might be a very competent judge.

A third difficulty, however, is arguably more fundamental than either of the two just noted. Once we acknowledge that political disagreements—I have in mind legitimate political

disagreements—are not confined to disagreements about the best ways to advance the common good, but also reach to the content of the conception of the common good, then the Condorcet argument seems not to address the principal difficulty in the case for majority rule. More precisely, the epistemic argument does not address the principal obstacle to the view that in following the judgment of the majority, I am following my own will, or taking it as a rule.

Suppose, for example, we have a group of citizens who disagree about the relative importance of individual independence, and therefore about how compelling a showing of collective benefit is needed before a regulation can be permitted; or they disagree about the relative importance of preservation and prosperity in the social welfare function that each uses to assess the performance of the political association (SC 3.9.4). Suppose then that I disagree with the majority view about the content of the common good because I think that members of the majority undervalue individual independence, or put too much weight on security relative to prosperity. I might nevertheless think that they are very competent judges about how to advance the common good as *they* interpret it, or that they would be good at judging how to advance the common good as I interpret it, if they were only to agree with me: they are not epistemically incompetent. So the result of the collective decision is a law that I think will likely advance the common good as they see it. But I reject that conception of the common good. While I may think of the process that issues in the law as *fair*, in that it gives equal weight to the judgments of all, and that those judgments express some view of the common good that gives equal weight to the interests of all, it is not clear how I will be following my own will in obeying the law, or how I have given the law to myself. After all, I reject the rationale for the law as given by the majority's conception of the common good.

I see three ways that Rousseau might respond to this concern, while preserving a commitment to majority rule.

First, he might reject political disagreements on the content of the conception of the common good as compatible with full political autonomy: he might argue, that is, that we can only achieve the compatibility of political authority with freedom, and therefore can only achieve political legitimacy, if people can think of themselves as following their own view of the common

good when they obey the law. The Condorcet argument permits them to think that they are in this way autonomous—political self-legislators—even when they lose, but only with a strong requirement of underlying unanimity. As an interpretation of Rousseau, this may be the right thing to say. The implication would be that legitimacy, which depends on self-legislation, presupposes strong community, with agreement on an understanding of the common good. But if it is right, it restricts substantially the scope of the view.

A second possibility would be to argue that citizens can regard obedience as an expression of political autonomy even if they reject the rationale for the law so long as they attach fundamental importance to a fair process of collective decision-making, whatever the inputs to that process. So I follow my own will in obeying the law insofar as my own will is that regulations issue from a process of collective decision-making that treats everyone as equals *by* assigning equal weight to the inputs of each in the process of collective decision-making. But this cannot be Rousseau's view, nor is it, I think, an especially attractive view. In Rousseau's terms, if the result of the collective decision gives no weight to my good, then I cannot endorse it, even if the process was fair in giving equal attention to each voice.

A third possibility would be that I follow my own will when I follow the judgments of the majority so long as I think both that the majority is making its decisions on the basis of a sincere effort to advance the common good as they reasonably (though not correctly) see it, and that the members of the majority are competent judges of how to advance the common good (again, as they see it). In this case, I would not simply be endorsing the result because it issues from a process that treats members as equals in assigning them equal voice, though I do think that that is an important consideration in itself; but also because I could see the result as advancing what I take to be the fundamental interests of myself and others in the association, according to an account of the common good—a way of combining those interests—that I think is not unreasonable, at least in the following sense: I can understand the considerations that might have led me to endorse that account of the common good, to see it as respecting the idea of not imposing conditions on others that one would not be prepared to live by oneself. So I am free in

obeying the law because I think the majority is a good judge of how to advance its conception of the common good; because the process respects the judgments of all members, and thus treats them as equals in at least that way; and, crucially, because the result treats people as equals in giving equal weight to the good of each, according to a way to understand that idea that I think is not unreasonable.

This idea has much to be said for it, and does, I think, represent a natural extension of Rousseau's views about the possibility of political autonomy, once we acknowledge the scope for reasonable political disagreement. But it does not strike me as a view that he himself considered, and thus not one that can profitably be discussed using the resources of Rousseau's view.

Rights and the General Will

Finally, the account of the general will provides a basis for understanding what Rousseau means when he says that the original contract involves a "total alienation" of rights to the community (SC 1.6.6). I will discuss some issues about more specific rights in Chapter 5. My aim here is in part to set the basis for that discussion, but also to address the ominous ring of Rousseau's phrase about total alienation. Perhaps the ring can be muffled if we consider, by reference to the theory of the general will, how the community to which rights are alienated is itself constituted.

To be less oracular. What Rousseau means when he refers to the "total alienation" of rights is not that members of a society regulated by a general will lack all rights (see e.g. SC 2.4). Instead, his point is about how we are to understand the foundations of rights in a well-ordered political society. Specifically, the foundation of rights in a political community is the general will itself—that is, the shared understanding of the common good that provides the supreme bond in the community—and not a set of claims that are prior to or independent of the common good. I say the foundation of rights *in a political community* because Rousseau allows that members of one political community can claim rights against members of another—for example, rights over land, which is part of the territory regulated by the political community—without appealing to a general will, but instead to rights based on need and first possession: "with regard

to its members, the State is master of all their goods by the social contract which serves as the *basis of all rights within the State* [emphasis added]; but with regard to other Powers it is master of all of its members' goods only by the right of the first occupant which it derives from private individuals" (SC 1.9.1). But, as this remark indicates, within a political society, the basis of rights is not a natural entitlement. In a society with a general will, there is shared acknowledgment that certain fundamental interests of each member ought to be respected and advanced: this, as I indicated in the discussion of the Common Good Condition, is the force of the idea that the general will wills the common good. Construing rights as claims that ought to be acknowledged by others, then, the existence of a general will requires the acknowledgment of rights, because it implies a shared recognition of the requirement that those fundamental interests are to be protected, and of the supremacy of this requirement in regulating association. That there are basic rights, then, is implicit in the notion of a social association regulated by a general will and aimed at the common good, understood in the non-aggregative way that Rousseau understands it.

The point, then, of Rousseau's remark about the total alienation of rights is precisely *not* that members of the political community lack rights, but that they claim those rights *as members, and that all the rights are founded on the common good, understood non-aggregatively*: "no associate has anything further to claim: For if individuals were left some rights [here, recall Hobbes on the true liberties of subjects], then, since there would be no common superior who might adjudicate between them and the public, each, being judge in his own case on some issue, would soon claim to be so on all, the state of nature would subsist and the association would necessarily become tyrannical or empty" (SC 1.6.7). To claim the rights, then, is to acknowledge that the common good is the ultimate authority in the association; but when I claim rights on this basis, then I am also committed to acknowledging my obligations as a member, because those obligations are equally founded on the common good. The force of the point about "total alienation," then, is that the social compact bars me from basing claims to rights within a political association on rights I would have had in a state of nature, and therefore bars me from claiming them without acknowledging the corresponding obligations.

A Solution to the Fundamental Problem?

With these points as background, we can now address the claim that the general will is the solution to the problem of the social contract. Why, that is, does Rousseau think that the society of the general will—a society that is guided in its operations by a shared commitment to the common good—provides the solution to the fundamental problem? Recall the fundamental problem: What form of association provides protection and security to each, and secures the freedom of each, given conditions of interdependence? When Rousseau says that the social contract solves this problem I assume that he means that we solve the problem by asking: What form of association would individuals themselves agree to who are moved by self-love and by an interest in securing their freedom? The answer is a free community of equals, an association regulated by its equal members, who share a conception of the common good. But why? Why is “the association must be regulated by the general will of its members” an answer to the question: How is it possible to reconcile self-government with the chains of social connection?

Rousseau’s argument for his striking conclusion is very compressed (SC 1.6.6–8). It has two main elements, corresponding to the considerations of self-love and freedom that figure in the description of the contract. To see how it goes, notice that two main features of the association need to be accounted for. The first is that the association ought to advance the common good; the second is that this common good is the common aim of the citizens in the association: “the first law, the only truly fundamental law that flows directly from the social compact, is that each man prefer the greatest good of all in all things” (GM 2.4.12).

Consider the first. The requirement that the association advance common interests flows naturally from the idea that everyone must agree to it (the agreement must be unanimous) and everyone is moved by self-love and so cannot be indifferent to his/her own interests. So were there not some assurance that it would advance each person’s interests, some people would not agree to it.

But why the condition that members themselves share an understanding of the common good, and give priority to advancing it? It is one thing to agree that the association should produce certain results. It is another to agree that the members are to make that goal their own supreme aim, or agree to certain norms of political argument. One reason for agreeing to the norms would be that common allegiance to the common good is the only sure way to achieve the common good: it must be aimed at to be achieved. This instrumental rationale may state an important truth, and I will discuss it in more detail as I proceed. But I do not think it gets Rousseau's argument right. His principal reason for including this clause in the contract is that common allegiance to the common good would be the only way for people to secure their freedom under conditions of interdependence.

To see why he might have thought so, consider two alternative sorts of political society. In the first, the laws of the association advance the common interests, each person is required to comply with the laws, but no member of the association cares about the common interests; instead, each cares only about his/her own interests, or the good of some subset of the population (whether defined ethnically, sociologically, regionally, or religiously). Assuming that the laws leave some room for personal choice, each person will be free to pursue his/her own interests within the laws. But the laws themselves will be understood as imposed constraints since they do not reflect the person's own judgment of what is best. For each person, the best would be for the laws to advance that person's own particular interests.

In the second case, the political society also advances the common interest, and each person again is required to comply with the laws. What distinguishes it is that each also identifies with the goal of the association. Each judges that what is best is to associate under conditions that advance the common good. Assuming this identification with the associational goal (and that the associational goal takes priority over personal goals), when members comply with the laws they follow their own will. Thus, by sharing the conception of the common good that the laws are assumed to advance, they achieve—in complying with the laws—the autonomy that comes from acting on principles they recognize as their own, from "obedience to the law one has prescribed to oneself" (SC 1.8.3). For citizens to have, that is, the

general will as a rule is for them to have “their will own alone as rule,” and because they have their own will as a rule, “it [the social contract] leaves them as free as before” (LM 232; SC 1.8.3, 2.4.8; E 461; FP 483–4). So, if the conditions of self-love and interdependence require an agreement to an order that advances common interests, then the only way to achieve self-rule given self-love and social interdependence is through an agreement that establishes a shared will to advance those interests.

This vision of self-governance achieved through political community is Rousseau’s distinctive contribution to political theory. The emphasis on political community distinguishes Rousseau from Hobbes and Locke, who see the state as protecting individuals from one another, and enabling individuals to pursue their separate ends without interference. For Rousseau, freedom as self-government requires the framework of political community defined by the general will.

To clarify the line of thought, I want to consider an objection to Rousseau’s view, and how he might respond to it.

Stated generally, the objection points to a tension between the idea of contractual legitimacy—the idea that political legitimacy is based on an agreement made by individuals—and the idea that political argument and justification are to appeal to substantive considerations of the common good—that justification is not a matter of what people choose, but a matter of whether the correct goals are advanced.⁴ According to this objection, the contractual argument does not help us to see how the general will reconciles autonomy and social connection. On the contrary, it is the principal locus of a conflict between two Rousseauian commitments: to the importance of reasons of the common good as a basis of political argument and collective decision, and to the importance of individual choice in justifying political authority.

Before stating my response, I want to distinguish this objection from another objection, which strikes me as more superficial. According to this second objection, there is an obvious tension between the two claims about the basis of legitimacy—a general will focused on the common good and a free agreement—since they assert two different sources of legitimacy. But if there is a problem, it is not so obvious. Rousseau argues that the general will is the source of political legitimacy and legitimate law *because* the social contract would produce an agreement

to the authority of the general will. The two sources are, then, not straightforwardly in competition. If there is a difficulty it must turn on the content of the agreement—on the nature of the general will—and not merely on the unobjectionable fact that the proximate source of legitimacy is the general will, understood as a shared commitment to the common good, whereas the ultimate source of legitimacy—which explains the legitimating force of the proximate source—is a social compact between free and equal persons.

My response consists of three points.

1. *If we think of the parties to Rousseau's social contract as moved solely by self-love, then there is—in a subtle way—a conflict between the idea that legitimacy is based on a collective choice and the conception of legitimacy rooted in considerations of the common good.*
2. It is wrong to suppose that the parties to Rousseau's social contract are moved solely by self-love.
3. Once we properly understand their motivations, we will see that there is no tension between the idea of legitimacy based on agreement and the idea that legitimacy resides in the general will.

1. To begin with, then, I want to outline an interpretation of Rousseau's view on which the parties to the social contract are moved exclusively by self-love, and then show how troubles arise on this interpretation. For reasons that will become more clear as I proceed, I will call this the "self-effacing Hobbesian" interpretation.⁵

Understood as a self-effacing Hobbesian, Rousseau supposes that individuals are initially motivated simply by their own long-term interests in protecting their person and goods (hereafter: self-interests). As they become interdependent, they recognize that mutually beneficial coordination is possible; furthermore, they see that their own long-term self-interest is advanced best by cooperating on mutually beneficial terms, if they can rely on the cooperation of others. But they also notice two large hurdles to achieving coordination for mutual benefit.⁶ First, there are a variety of mutually advantageous arrangements, none of which provides a focal point for their cooperation: so they face a bargaining problem. To address this problem, it suffices that they agree to specific terms of cooperation. But they also see

that their agreement would not be self-enforcing both because of temptations to violate the agreement arising from motivations other than long-term self-interests (e.g. passions) and because they need to be assured of the cooperation of others.

Assessing various solutions to this dilemma, they judge that a Hobbesian sovereign will not suffice to solve the problem of motivating general compliance and so producing a stable order. For even with an absolute sovereign, obedience among subjects will depend ultimately on self-interested calculations, and rational calculators, they judge, are unreliable cooperators. They conclude that only a society with a general will can ensure the stable satisfaction of long-term interests. For only if individuals are directly, not simply instrumentally, motivated by reasons of the common good will the association operate over the long term for that good. So they agree that they ought to have a general will. But they are aware that one cannot simply decide to have such a will: it must be cultivated, educated, formed. As a result, they agree to establish and cooperate in arrangements that will, they expect, transform self-interested individuals over time into citizens whose primary allegiance is to the common good—thus producing a “remarkable change in man by substituting justice for instinct in his conduct, and endowing his actions with the morality they previously lacked” (SC 1.8.1). That allegiance might be more abstractly civic—drawing on a sense of duty and public responsibility, concerned to preserve a political arrangement because it realizes universal values—or it might be more ethno-national, drawing on allegiances and attachments to people, places, and a shared, distinctive way of life: “Give a different bent to the Poles’ passions, and you will give their souls a national physiognomy which will set them apart from all other peoples, which will keep them from merging, from feeling at ease, from inter-marrying with them, you will give them a vigor which will take the place of deceptive appeals to empty precepts, which will make them do by preference and passion the things one never does well enough when one does them only by duty and interest. It is upon souls such as these that an appropriate legislation will take hold. They will obey the laws and not elude them because they will suit them and will have the inward assent of their wills” (P 184). Hobbes lacks both of these forms of allegiance, so we have political unity through common subordination to a supreme will. The self-effacing Hobbesian has

either or both, and so can achieve common benefits through civic and/or ethno-national integration.

I said earlier that *if* we follow this line of argument and think of the parties to Rousseau's social contract as simply self-interested, then a conflict will emerge between contractual legitimacy and the conception of legitimacy rooted in the general will. To see where the problem comes in, let's continue the story. Consider what happens once the "remarkable change" has set in. Persons no longer regard themselves as, at bottom, self-interested. While they continue to have separate interests, their chief allegiance has shifted to the common good, and their chief reason for complying with social rules is that those rules advance the common good. Since they continue to have self-interests, *some* justification for their association is provided by the fact that it *would* be rational, in view of the self-interests alone, to bootstrap their way into the general will. But the fact that self-interested individuals would agree to the association cannot be their chief reason for complying with it, since they are not such individuals.

Thus, if individuals regard self-interested reasons as fundamental, then they can agree that they ought to have a general will but they cannot have one; for having a general will consists in part in acknowledging that self-interested reasons are not fundamental. In contrast, if they have a general will, then the self-interested contract does not provide them with the basic justification for having it. There is, then, no single point of view which embraces both the contractual justification and the general will; more specifically, there is no perspective from which the general will can be regarded as the basis for political legitimacy *because* it would be agreed to. A person who endorses the general will will think it false that the social contract provides any justification.

But here is where a subtlety comes in. The absence of such a point of view is not by itself an objection to a theory. A social contract theory could be self-effacing; that is, it could be that the contract-based reasons for having a general will could not be reasons for the person with the general will. It could be, that is, that the social contract mandates that we believe (or mandates something that has as a consequence that we believe) what is in fact false—in particular, that a social contract has no justificatory power. On the self-effacing Hobbesian interpretation, the social contract does justify the authority of the general will, with its

focus on the common good; but it also justifies us in believing (falsely) that the social contract provides no justification.

But I do not see that Rousseau held such a theory. He gives no evidence of believing that a contractual justification for the authority of the general will cannot be given to those who have such a will—no evidence of believing that political legitimacy cannot be reflective and transparent. For example, in his *Letters Written from the Mountain*, addressed to citizens of Geneva, he presents the essentials of his political theory (see Letter 6); and he summarizes it in some detail in the education of *Emile*, presenting it as the basis of *Emile's* decision about political allegiance. Thus he appears to endorse a non-self-effacing interpretation of the social contract, and in fact to believe that public recognition that the social order would be agreed to by free agents contributes to stability of the order: "I posited as foundation of the body politic the convention of its members. . . . Independently of the truth of this principle, it prevails over all the others by the solidity of the foundation it establishes, for what more certain foundation can obligation among men have than the free engagement of the one who obliges himself?" (LM 231). If Rousseau does require that the contractual justification be non-self-effacing, then the interpretation of the parties to the contract as simply self-interested does produce a conflict between contractual justification and the general will. Those to whom the contract is to be offered as justification would, by virtue of having a general will, not think of it as providing such a justification.

2. But self-effacing Hobbesianism misstates Rousseau's problem. That problem is not: What sort of society would rational, asocial, and purely self-loving individuals agree to? Because that is not his problem, it is wrong to think that he holds that such individuals would agree to regulate their actions by reference to non-self-interested reasons. His question is: What form of association would socially interdependent individuals agree to, if they were interested in protecting their person and goods *and in being as free as before*—that is, in expressing the freedom that belongs to their nature? Parties to the social contract do not aim solely to protect their person and goods, and the interest in such protection is not the sole source of the allegiance to the common good that issues from the original agreement. Its source lies in the freedom that socially interdependent individuals aim

to preserve. In short, something other than interests in person and goods is at work in the contract itself, and that something else is what leads something other than self-interests to emerge from the agreement.

This observation about the importance of freedom as a concern that individuals aim to secure through the social compact would not be germane to the tensions between individual and citizen if freedom were simply yet another good—like security or personal well-being—that one seeks to advance by making an agreement. For one could still ask how a person, fundamentally interested in freedom (as well as security and well-being), can also fundamentally be a citizen concerned with the common good. But that misconceives the nature of freedom and its relationship to principles of political right. The right way to think about freedom is not simply as a good that principles of political right promote, in the way that they promote security and well-being, but rather as a value that is *expressed in* principles of political right and in the regulation of conduct in accordance with such principles.

Consider the relationship between freedom and the moral law in Kant's moral philosophy. Kant thought that human beings have a special dignity in virtue of being endowed with a free will, and that the fundamental principle of morality is expressed in the categorical imperative, which says that we are to act on rules of conduct that we could at the same time will as universal laws (to be followed by all rational agents). Kant also thought that there was a particularly intimate relationship between the will's freedom and the moral law: that "freedom and unconditional practical law reciprocally imply each other," and that the moral law is thus the principle of conduct uniquely suited to a free will.⁷ In affirming this connection, Kant did not suppose freedom is simply a consequence that we produce by following the moral law—as if the moral law said "promote freedom in the world"—but that we express our nature as free by following the moral law; the relationship of freedom to the moral law is expressive rather than instrumental. Kant's thought, very roughly put, is that we reveal our capacity for choosing independently from our desires and inclinations by acting from the moral law. Why freedom's law has this particular content is a complicated matter, which is not germane here.⁸ The only point I wish to emphasize is that we show that we are free by acting from the moral law, using it as our supreme practical principle.

Rousseau's view, then, is that, under the conditions of social interdependence that he is assuming to obtain as background to the social compact and the investigation of principles of political right, human beings animated by self-love can only express their free nature by having a general will and using it as the supreme guide to conduct.

3. I said earlier that once we reject the self-interested view of the parties to the social contract, the problem of the tension between contractual legitimacy and legitimacy based in the common good disappears. But that may not be so obvious. Indeed it might be thought that the alternative I have proposed, which emphasizes that the contracting parties seek to secure their interest in freedom, actually sharpens the problem. To see why, consider again the possible tensions between the standpoint of contracting party and citizen, now taking into account the fact that the parties to the contract have an interest in remaining as free as before.

Thus, as parties to the social contract, individuals regard themselves as free agents. From this standpoint, they see an allegiance to the common good as simply one possible allegiance among others. Should they choose to work in service to the common good, they will see themselves as volunteers in that service. Therefore persons who regard themselves as contractors cannot have the kind of supreme attachment to the common good that is required on Rousseau's conception of the general will. Thus from the free contractor's point of view, the contract does not justify allegiance to the general will.

Citizens, on the other hand, do identify with the common good. They do not regard themselves most fundamentally as free choosers of ends, but as members of a state; their choices are always made as members. They take no interest in knowing what they would choose in abstraction from this political self-conception (if the question has any content at all). To treat a social contract among free individuals—for whom an allegiance to the common good is one possibility among others—as providing reasons for endorsing the general will citizens would have to regard the freedom that figures in the contract as more fundamental than their substantive ends as citizens, more basic than their commitment to the common good, founded on the common interests of members. So regarding themselves would, however, undermine their allegiance to those ends. For once they

recognize their freedom as individual choosers, they would also come to regard the social order as simply one among the many alternative choices they might make about how to live. So here again, there is no single standpoint from which the contractual justification and general will might be embraced.

By way of response I want to consider this problem first from the contractor's point of view, then from the citizen's.

Consider first, then, the perspective of an individual who understands him- or herself as free, has an interest in expressing that freedom, but does not have a general will—or has a general will, but wonders about its rationale. To meet this concern, let's begin by recalling two points: (1) Rousseau's defense of a sovereign general will assumes social interdependence. It is not addressed to individuals who are outside of a society, and does not aim to persuade such individuals that they ought to join. In assuming that people are interdependent, he is assuming that people have desires that can only be satisfied socially, developed capacities that can only be expressed socially, and ethical views. And (2) Rousseau is not arguing that individuals ought to have and act on considerations of the common good under circumstances in which there is no reasonable assurance that others will do the same (GM 1.2.10).

Imagine, then, a person who has reasonable assurance that others will act to advance the common good, who recognizes that institutions do advance common interests, and who is aware (as I am now supposing) that the general will would be agreed to by members of the society if they were to reach an agreement: this person wants to know why he/she ought to comply with the terms of that agreement. The person is faced with two alternatives. The first is to have and act on a general will, knowing that that will would be the outcome of a social contract. The second is to conform for instrumental reasons—only when such conformity advances his/her own interests. Apart from a commitment to the general will, is there any reason to choose the former over the latter?

Rousseau supposes that a person who is self-consciously free desires to act in ways that express that freedom (to be "as free as they were before"). In view of that desire, the instrumental attitude is not satisfactory. For such a person wants more than the availability of alternatives within a system of laws and institutions that he/she views as constraints imposed by others.

He/she wants instead to be able to regard those “constraints” as themselves conforming to their own judgments of what is right. Instead of taking the instrumental attitude which regards the social framework as constraining, the free person wants to affirm the framework of rules itself; they want to have “their will alone as rule” (LM 232). But under circumstances in which there is—as stipulated—a widely shared general will to which institutions do on the whole conform, to affirm the arrangements—to have one’s own will as a rule—is just to have a general will. The relationship between the interest in freedom and the general will is, it should be emphasized, not of the same kind as the relationship between long-term self-interest and the general will suggested by self-effacing Hobbesianism. Having a supreme general will is a *means* for advancing long-term self-interests. Having a general will, by contrast, is not a means to autonomy; under the conditions I am assuming here, it is what autonomy consists in.

Consider next the citizen’s point of view: Does having a general will and a conception of legitimacy based on considerations of the common good preclude regarding oneself as autonomous party to a social contract that provides the source of legitimacy?

The alleged conflict can be understood in either of two ways, corresponding to two interpretations of “based on considerations of the common good.” On the first, to base legitimacy on such considerations is to accept the communitarian-relativist theory that the rightness of an action simply *consists in* its advancing the common good of the community to which the agent for whom the action is right belongs. This does make for a conflict between the general will and the idea of contractual legitimacy. But Rousseau’s endorsement of a contractual conception indicates that he did not think that this is the correct account of rightness.

On the second construal, to endorse the idea that legitimacy is based on considerations of the common good is to endorse the substantive thesis that advancing the common good is right—required as a matter of political morality—but not to propose any view at all about what rightness is. But now the problem is no longer clear. To think that legitimacy is based on a contract among autonomous agents and that it is based on a general will focused on the common good is simply to think that the contract among members, understood as free agents, would issue in an agreement to the authority of the general will

and to accept that authority in part because the contract would have that result. But I suggested earlier that this is a plausible claim about the initial agreement and about the conditions of self-government under conditions of social interdependence.

To be sure, if we emphasize Rousseau's strongest communitarian formulations—if members of a free community of equals might genuinely think of themselves and perhaps need to think of themselves as exclusively Spartans, or Poles, or Americans—then the social compact would once more lack a role in public understanding and political argument.⁹ We would have a self-effacing Rousseauism, analogous to the self-effacing Hobbesianism considered earlier. Here, the justification for living in a free community of equals, with its concern for the common good of members—the fact that a free community of equals solves the fundamental problem by enabling people to protect their basic interests without being subject to a morally unacceptable subordination of will—would be unavailable (practically speaking) to the members of such a community. Unavailable practically speaking, because the members would understand themselves in exclusively national terms and need to understand themselves this way in order to be motivated to comply with the terms they would agree to as free agents facing conditions of interdependence: the agreement would be intellectually intelligible to them, like telling Genevans about Spartan self-sacrifice, but the fact that free persons would make such an agreement would not provide them with considerations that they take to be reasons.

Consider the case of Poles once they have "souls [with] a national physiognomy which will set them apart from all other peoples, which will keep them from merging, from feeling at ease, from inter-marrying with them" (P 184). Then, the "appropriate legislation will take hold." They will "obey the laws and not elude them because they will suit them and will have the inward assent of their wills. Loving their fatherland, they will serve it out of zeal and with all their heart" (P 184). Suppose, as Rousseau argues, that free agents, considered apart from these national passions and attachments, would agree to live in a free community of equals, satisfying the four conditions on the society of the general will. For the Poles, this fact about the social compact would be an intellectual curiosity, not a consideration that they take to provide a reason. Moreover, if they started to treat it as providing

a reason, as something more than an intellectual curiosity, this might work as a solvent, destructive of the self-conception associated with more exclusive passions and solidarities that hold the political order together.

Earlier, in the discussion of self-effacing Hobbesianism, I said that Rousseau does not see a problem in the public, practical availability of the contractual justification (indeed, at LM 231 he suggests that it has practical benefit). The same point applies to self-effacing Rousseauianism: thinking, for example, that we live in a place that is conceived in an idea—that we are all free persons—and dedicated to an abstract proposition—that we are all equals—may not undermine the solidarity and mutual identification that Rousseau understands to be needed for a legitimate order. The more extreme communitarian formulations may, then, overstate (by oversimplifying) Rousseau's convictions about the motivational preconditions for the society of the general will. To appreciate better what those motivational demands might be, we need to address the second possibility problem, the problem of motivational possibility.

4

The Natural Goodness of Humanity¹

Walking along the road from Paris to Vincennes in 1749, contemplating a question from the Academy of Dijon—"Has the restoration of the sciences and the arts helped to purify morals?"—Rousseau was, he writes to Malesherbes in 1762, overtaken by a flood of ideas, "a thousand lights." At the heart of this "sudden inspiration" was the thought that dominated his subsequent writing: "that man is naturally good, and that it is solely by [our] institutions that men become wicked" (M1 1135–6). In a 1762 letter to Archbishop Beaumont, Rousseau says that the idea of natural goodness is the "fundamental principle of all morality," and the basis of "all my Writings" (B 935; RJ 213).

The conception of natural goodness—with its idea that institutions account for our most destructive, dangerous vices—is not uniquely Rousseau's. Egalitarian and radical democratic political theories commonly assert that human motivations supporting unjust institutions are not intrinsic to our nature, but explained by those institutions themselves. Such theories, however, often accompany this assumption with implausible assertions about human plasticity, and denials that human beings have a nature (or claims about its complete social determination or construction).²

Rousseau's account of natural goodness separates it from this undesirable company. Human beings have a complex natural endowment including: self-love; a sense of self-worth and an associated concern to be treated with respect; compassion (*pitié*); a power of choice; and a range of cognitive powers. To be sure, our endowment is, in a sense to be explained, *abstract*. Self-love, for example, can take many forms, depending on how a person thinks about herself (the self that is loved); similarly with a sense of self-worth and compassion. The social-institutional environment accounts for the determinate expressions of the intrinsic features of our endowment. But the endowment itself is fixed and provides an underlying structure that helps to explain the variety of human motivations: thus Kant's description of Rousseau as "the Newton of the mind."³

The idea of natural human goodness, then, gives Rousseau's moral and political views their bearings, and is important in its own right. But it is also obscure. How can we join the view that human beings *have* a nature with the claim that current expressions of it—indeed virtually all its observed expressions—are distorted? Jean Starobinski raises a variant of this question when he asks: "How does one reconcile the assertion that 'man is naturally good' with the assertion that 'everything degenerates at the hands of man?'"⁴ But the puzzle raised by Rousseau's view is more general. Moreover, if all our evidence involves such distortion, how can we acquire *knowledge* of our nature? More fundamentally, is it reasonable in the face of the observable patterns of motivation and conduct to endorse a hopeful conception of human nature, or is such endorsement simply a matter of wishful thinking—of projecting moral and political preferences into our nature?

In this chapter, I present an interpretation of Rousseau's account of natural goodness, with a view to addressing these questions. Rousseau's conception of a free community of equals describes a political society in which citizens achieve full autonomy by cooperating as equals for the common good. This ideal faces a charge of being objectionably unrealistic because it is incapable of engaging human affections and imposes demands that our nature makes us unable to fulfill. A central role of the account of humanity's natural goodness is to provide the basis for Rousseau's response to this charge.

I will proceed in five steps.

I begin by presenting the foundations of natural goodness in human nature, emphasizing Rousseau's distinction between the *abstract potentialities* intrinsic to human nature and the *determinate expression* of those potentialities as a result of social circumstances. Chomsky has argued that a distinction of this kind played a central role throughout the tradition of Cartesian linguistics. Although language acquisition is of course triggered by circumstance, and the particular language a speaker acquires depends on the language spoken in his or her social environment, certain fundamental features of the acquired state reflect intrinsic features of human nature.⁵ A parallel thought guides my discussion of the formation of human motivations: although the acquisition of motives is of course triggered by circumstance, certain fundamental features of the acquired state reflect intrinsic features of human nature.

Next, I present the two main elements of the conception of natural goodness itself: that we are originally innocent, in that all the vices can all be explained without attributing them to "the human heart" (E 92; B 28); and that the same properties of human nature—in particular concerns about self-worth—that are now expressed in vice can also be expressed in ways that support the requirements of free association among equals. We need, Rousseau thinks, an alternative way to express our nature, not the suppression of some elements of our nature.

Third, I explain why Rousseau might have thought that affirming the idea of natural innocence is reasonable, despite all the evidence of human vice. This will involve a sketch of his hypothetical history of inequality and vice, aimed at showing why Rousseau supposed that his own conception of our nature can account for all the evidence of vice without attributing it to our nature.

Fourth, I explain how elements of our nature that are now expressed in distorted ways might be alternatively expressed in ways that support a free community of equals.

Finally, I suggest that Rousseau's reasons for affirming the natural goodness of humanity are moral—that the idea of natural goodness is a conception of practical reason, not theoretical reason, though it does not offend against theoretical reason.

My aims are principally interpretive, and I will not be defending the details of Rousseau's view. Still, I think the main outlines have considerable force, and my interpretation emphasizes

three ideas that strike me as plausible: the central role in our psychology of the sense of self-worth; the view that our natural endowment consists of a set of abstract powers capable of diverse forms of expression; and the idea that moral considerations provide good reasons for endorsing a hopeful conception of human nature.

Three Properties of Human Nature

The central tenet of the idea of natural goodness is that “There is no original perversity in the human heart. There is not a single vice to be found in it of which it cannot be said how and whence it entered” (E 92; B 28; RJ 23).⁶ More particularly, the explanation of “how and whence” is “that society depraves [human beings] and makes [them] miserable” (RJ 213).

To understand the content and force of these claims, we need to interpret them in light of Rousseau’s distinction between the abstract potentialities intrinsic to human nature and the determinate expressions of those potentialities.⁷ *Determinate* motivations are not themselves original elements of our nature but instead reflect the way our nature is expressed, given our circumstances and the self-conceptions that arise from those circumstances. Depending on these circumstances, the expression of the intrinsic properties of human nature will take quite different forms.

This distinction provides, as we will see, the basis of Rousseau’s answer to the motivation problem. Though depressing experience suggests that the society of the general will is not in the human cards, such experience does not provide a direct window on our nature. Our observations are confined to the expression of human powers within institutional forms, and nothing in those observations excludes the possibility that we could, under suitable conditions, express the very features of our nature that are now manifested in vice in quite different ways—in particular, by regarding one another as equals and being moved to comply with regulations because they advance the common good of members thus conceived.

To fill this out, and show the distinction between abstract potentiality and determinate realization works, let’s consider

Rousseau's idea that human beings are naturally endowed with self-love, compassion, and a set of cognitive powers. Let's take each in turn, bearing in mind that we want eventually to understand more complex motivations in terms of these essential elements.

Self-Love

Each human being has a natural concern for his/her own welfare—*amour de soi*, or self-love. Self-love is itself not a simple idea, in part because what we love in loving ourselves depends in part on how we conceive of ourselves. But whatever else it involves, self-love includes concerns to preserve ourselves and to protect the means of that preservation. Moreover, self-love, like love of other people, is a matter of valuation as well as affection.⁸ However I think and feel about myself, among the things I value is myself. We each have a sense of our own worth that animates our more specific concerns and aims.⁹

This self-love is not intrinsically comparative. Thus, from the fact that I love myself it does not follow that I love myself more than others; I may have no thoughts at all about others. But when I develop relations with others, self-love, understood to include a sense of self-worth, naturally extends to include a concern for the relations in which I stand to others and for the regard of those others for me. In particular, I am concerned that those others affirm my worth: concerned that they recognize me as someone with the worth I attach to myself.¹⁰ I say that the extension is natural because I am supposing that my sense of my own worth comprises the thought of myself as valuable. Because self-love is not simply a matter of affection for myself or wishing myself well, but of regarding myself as valuable, I suppose that others ought to recognize my value as well. *Amour-propre*, generically speaking, names this concern for my standing in the eyes of others: the concern that they get my value right, and express their recognition of my value. Let's say that *amour-propre* is the abstract concern that others value me as I value myself—the concern that they display for me the regard that I judge to be appropriate, given my sense of my own value (E 213–15).

Amour-propre comes in two forms, corresponding to two ways that I might value myself. Suggesting this variety of forms, Rousseau says: "Remember that as soon as *amour-propre* has developed, the relative *I* is constantly in play, and the young

man never observes others without returning to himself and comparing himself with them. The issue, then, is to know what rank among his fellows he will put himself after having examined them" (E 243). A person with the first form—call it the *egalitarian* form—regards his own worth as equal to that of others: that is the "rank" he puts himself in. As a result, he believes that others ought to take his judgments and well-being equally into account, to treat his concerns as on a par with theirs. This form of self-regard is reasonable in two ways, both intrinsically and instrumentally.

The egalitarian form is intrinsically reasonable because it conforms to a correct understanding of human beings, of our "true relations" as equals in virtue of our common human nature. Rousseau summarizes the "whole of human wisdom" about the development of the passions this way: "(1) To have a sense of the true relations of man, with respect to the species as well as the individual. (2) To order all the affections of the soul according to these relations" (E 219). To order affections according to true relations is to have affections rooted in true beliefs about the relations of human beings—in particular, about the "identity of our natures" with the natures of other people, despite differences of social station (E 221): "Man is the same in all stations. . . . To the man who thinks, all the civil distinctions disappear. He sees the same passions, the same sentiments in the hod-carrier and the illustrious man. He discerns there only a difference in language, only a more or less affected tone" (E 225; also 194, 222).

An egalitarian sense of worth is instrumentally reasonable in this sense: that unlike the view of oneself as of greater worth, it can genuinely (without affectation) be upheld by others, and so need not be a source of discontent and misery. This possibility of being endorsed by others reflects a further psychological premise: that it is "impossible" that "each person should prefer us to all else and to himself" (RJ 113)¹¹ but not impossible that each should regard himself as our equal. If our sense of our worth takes the egalitarian form, then we are not led to make the psychologically impossible demand on others that they think better of us than they think of themselves. In contrast, if we "[prefer] ourselves to others," and if this preference reflects a sense of self-worth, a high valuation of ourselves, then we also "[demand] others to prefer us to themselves, which is impossible" (E 214).¹²

The second, inegalitarian form of self-regard is to think oneself more worthy of regard than others, and find it insulting if they reject this elevated conception—if they reject the view that one is more worthy of consideration and respect than they are themselves. A point of terminology before proceeding: Rousseau often uses *amour-propre* exclusively for the prideful form of self-regard (D2, 218). But in *Emile* in particular, the generic use—which covers both forms—is more clearly in evidence (E 92, 215, 235), and one passage (“you would only inflame his *amour-propre*”) suggests that we might use the phrase “inflamed *amour-propre*” for the inegalitarian, prideful form (E 247); I will follow the latter usage here.¹³

A third possibility would be to regard oneself as an inferior, as worth less than others. Rousseau does not consider this third possibility, perhaps because (as I mentioned above) he does not think it is a psychological possibility: while we may observe a mutual enactment of subordinate and superordinate performances, we should be cautious about inferring that the subordinates genuinely accept their subordination or endorse the regnant justifications for it. Because Rousseau dismisses this third possibility, I will not explore it further here.

Returning then to the second, a person with an inflamed sense of self-worth “set[s] greater store by himself than by anyone else” (D2 218), assigning himself “the first and best place” (RJ 112). If the earlier observations associated with the intrinsic reasonableness of the egalitarian form are right, then this inegalitarian form of *amour-propre* of course conflicts with a correct understanding of human beings: naturalizing status inequalities, it treats them as outward expressions of inner differences of nature and natural worth.¹⁴ “[I]f as a consequence of my care Emile prefers his way of being, of seeing, and of feeling to that of other men, Emile is right. But if he thus believes himself to be of a *more excellent nature* [emphasis added] and more happily born than other men, Emile is wrong. He is deceived” (E 245).

In addition to resting on error—“the error most to be feared, because it is the most difficult to destroy” (E 245)—the inegalitarian form of *amour-propre* provokes inner discontent and anxiety and interpersonal conflict because it demands the impossible self-abasement of others. It is thus instrumentally unreasonable as well.

Although self-regard is intrinsic to our nature, then, its particular form of expression—whether egalitarian or inflamed—depends on circumstances. Describing Emile's entry into the "moral order," Rousseau says that the point at which he begins to make comparisons with others "is the point where love of self turns into *amour-propre* and where begin to arise all the passions which depend on this one. But to decide whether among these passions the dominant ones in his character will be humane and gentle or cruel and malignant, whether they will be passions of beneficence and of commiseration or of envy and covetousness, we must know what position he will feel he has among men" (E 235; also 243).

Later I will explore the connections between circumstances—"position among men"—and forms of self-regard. Suffice it to say here that these connections play a central role in Rousseau's answer to the problem of motivational possibility that I have identified in his political view. An important thought, which I will return to later, is that such vices as jealousy, selfish indifference, greed, cruelty, envy, and cowardice are all fueled by inflamed *amour-propre*: they are not original to human nature, nor do their roots lie in egoism or in an egoistic hedonism, or in an insatiable desire to accumulate goods. Instead they reflect a conception of one's relative worth which issues in a desire—indeed a demand, a sense of entitlement—to live a better life than others; indeed, egoism itself may have this social basis. And this sets the central question for an explanation of vice: *how does a generic, not intrinsically inegalitarian concern to be treated with respect come to be particularized as a demand and desire to be treated as a better?* If Rousseau can explain this—and how the concern to "assign oneself the first and best place" (RJ 112) might be forestalled—then he will have accounted for the origin of vice without supposing it to be an original predisposition, whether express or latent.

Compassion (*Pitié*)

Each person is moved by compassion, a non-derivative predisposition to respond with aversion to the suffering of others (D2 152–4, E 221 ff.). To say that this predisposition is natural is not to attribute a determinate kind of compassion to human beings quite generally. On the contrary, the natural predisposition

to compassion—like the concern about self-worth—can be expressed in various forms, and the form of expression depends on circumstances.

With self-regard the important dimension of variation was rank ordering; here the dimension of variation is extent or scope of concern. In particular, the compassion we tend naturally to experience when we are face to face with the sufferings of others can be generalized to cover wider groups of people; suitably extended, compassion provides the foundation for a concern for the welfare of human beings quite generally (E 253), and so provides the affective soil for genuine virtue: “from this single attribute flow all the social virtues that he [Mandeville] wants to deny men. Indeed, what are generosity, Clemency, Humanity, if not Pity applied to the weak, the guilty or the species in general” (D2 153).¹⁵

As vanity provides the foundations for the vices—though it is not itself an original sentiment—compassion provides the basis for the virtues. But while compassion itself is original to human nature, its extended expression is not, and depends on how we regard others. Later we will explore the conditions that encourage this extension.

Cognitive Powers/Perfectibility

Along with these basic affective powers, we are endowed with a set of cognitive capacities—including capacities for memory, imagination, abstraction, conceptualization, self-reflection, and reason. These powers develop from their form as unrealized capacities into realized abilities only under the pressure of circumstances. “Superfluous faculties” (E 81), they are put “as it were, in reserve in the depth of his soul, to be developed there when needed” (E 80). Indeed, apart from sensory powers (D2 143–4), all forms of cognition involved in representing and thinking about actual and possible states of the world and of the self are triggered by circumstances. “It was by a very wise Providence that the faculties he had in potentiality were to develop only with the opportunities to exercise them” (D2 150; B 39–41).

There is much to be said about the details of these powers, and how they are activated by social setting. But for the purposes of understanding the idea of natural goodness, further discussion

of these points is best pursued by considering how these powers work to generate more complex forms of motivation.

Motivations

The above three properties of human nature lie at the heart of Rousseau's conception of natural goodness. To defend this doctrine, Rousseau must show how these three properties suffice to explain the sad observations about human motivations and conduct. Even if they do, we may of course still wonder whether they constitute the correct account of our nature; but it would be irrational to believe that they do if they are incompatible with the available evidence.

To see how this threshold consistency requirement might be met, we need first to see how these three elements of our nature work in combination.

Cognition and Motivation

The development of our cognitive powers (enlightenment)—"the successive developments of the human Mind" (D2 159)—enables people to think more complex thoughts, to engage in more complex forms of reasoning and deliberation, and, more to the point here, gives more complex shape to the fundamental motivations. Combining with the basic affective aspects of human beings—self-love and compassion—the development of cognitive powers generates the full range of human desires and passions.

More specifically, motivations fall into two broad types. First, we have a restricted class of natural desires, for food, sex, and sleep. These desires correspond to natural needs—the needs associated with preservation of self and species—and pursuing them is an aspect of self-love.¹⁶ These desires—really instinctual urges—reflect "the simple impulsion of Nature" (D2 42).¹⁷

All other desires depend on beliefs: "for one can only desire or fear things in terms of the ideas one can have of them, or by the simple impulsion of Nature; and Savage man, deprived of every sort of enlightenment, experiences only the Passions of this latter kind" (D2 142, 211–12, 215–16; LD 65–75). By contrast with

hunger or primitive sexual appetites or the desire for sleep, all other desires are concept-dependent and belief-dependent. They depend in particular on opinions and judgments, and require that the subject have a conception of the desired object—that the subject be able to represent the particular object of her desire as an individual with certain general properties.

Consider, for example, the distinction between natural and extended forms of compassion. The development of more complex forms of compassion—the generalization of compassion to the weak (generosity), the guilty (clemency), and the species (humanity)—requires enlightenment because those forms require thoughts about the weak, the guilty, or the species as objects of concern (this condition is necessary, not sufficient for the extension of compassion). Similarly, the development of *amour-propre* requires a capacity for representing oneself in relation to others, and that, too, is possible only with the realization of cognitive powers initially only latent in human nature.

The same is true for the sentiment of love, understood as an “ardent, impetuous . . . terrible passion that braves all dangers, overcomes all obstacles, and in its frenzy seems liable to destroy Mankind” (D2 154–5).¹⁸ Although love has a natural or physical aspect—the sexual desires associated with species reproduction—the sentiment as we know it also has a “moral aspect” which directs love to particular individuals. This aspect depends, for example, on ideas of “merit and beauty,” and so on comparisons. Fixed on an individual—a “preferred object” (D2 155)—it rests in part on understanding and appreciating that individual’s distinguishing features. In particular, it requires a grasp of “abstract ideas of regularity and of proportion” (D2 155). By contrast, the physical aspect knows no “preferences” (D2 155). Absent a realized capacity for certain complex thoughts, then, a person cannot be moved by love, or any of the emotions or motivations associated with love—for example, jealousy, despondency following rejection, heartbreak.

Social Bases of Cognition

Social practices explain beliefs, in particular the socially shared beliefs that Rousseau calls “public opinion” and that figure centrally in shaping the determinate content of motivations.

This explanation works in two ways. First, as a general matter the ability to conceptualize itself can only be realized with the evolution of social interdependence. For conceptualization depends on the ability to judge and on a grasp of definitions (D2 148), and these are tied to the mastery of language: “one has to state propositions, hence one has to speak in order to have general ideas” (D2 148). And language in turn depends on social cooperation (D2 144, 149).

Why Rousseau supposes that language requires social association is not clear, but the dependence of thinking on the ability to use a language seems more plausible. Rightly emphasizing that images are always particular—that having the concept of a tree, for example, cannot be a matter of having an image of a tree—Rousseau holds that “general ideas can enter the Mind only with the help of words, and the understanding grasps them only by means of such ideas” (D2 148). That is, having a concept is a matter of having the ability to use a sign (e.g. a word) that is part of a system of signs that can enter into combination with one another in various ways. Thus, “one has to state propositions, hence one has to speak in order to have general ideas: for as soon as the imagination stops, the mind can proceed only by means of discourse” (D2 148).

Second, the specific concepts and beliefs that people have depend on specific features of their form of association. Rousseau’s views about the mechanisms of their formation are obscure. But, for example, acquiring the idea of *property*—that is, having thoughts about things belonging to people—appears to depend on interacting with others under conditions in which people have some measure of de facto independent existence, some real ability to control things on their own (D2 167–71). Furthermore, mastery of such abstract ideas as *person*—exhibited, for example, in the thought that *I am one person among many*—depends on regular connections with other human beings that enable me to see similarities. And the conception of myself and others as equals requires some substantial echo in our social relations.

Putting aside the detailed mechanisms that figure in the formation of concepts and beliefs, the point is that social practices and political institutions figure centrally in the explanation of shared beliefs, including the ideas that people have about themselves: “A people’s opinions arise from its constitution” (SC 4.7.4; LD

73–4). An important consequence of this view is that opinions—including motivation-forming self-understandings—can be shaped by altering institutions. Because “a people’s opinions” depend on “its constitution,” part of the work of institutional design is to devise constitutions in light of this dependence.

Social Bases of Motivations

If cognition has a social basis, and motivations have a cognitive basis, then motivations—other than the purely natural urges—have social roots. Social arrangements, that is, explain motivations by explaining the ideas that make the motivations possible.

Rousseau, for example, famously rejects Hobbes’s “naturalistic” explanation of the state of war because Hobbes “improperly included in Savage man’s care for his preservation the need to satisfy a multitude of passions that are the product of Society” (D2 151). But what accounts for Rousseau’s conviction that the Hobbesian passions are not natural? Certainly not a sustained case drawn from empirical anthropology, a discovery of pre-social individuals, or the “uncertain testimonies of History” (D2 142). Rousseau’s case is founded on the observation that the motivations relevant to Hobbes’s account of human conflict are *cognitively complex*; the desire for long-term happiness, the insatiable desire for “power after power, that ceaseth only in death” (Lev. 70), and a concern for personal honor are possible only for people who are also capable of having certain ideas. The desire for future happiness requires a representation of myself as the subject of various future desires; selfishness and pride (inflamed *amour-propre*) require understandings of myself as one person among others, and estimations of my relative value. But cognitively complex passions require “enlightenment,” and enlightenment is a result of social interdependence. Outside of social interdependence, abstracting from the self-conceptions it engenders, we experience neither pride, nor hatred, nor a desire for vengeance.

Pride, for example, requires that I compare my circumstances to those of others. And such comparisons depend on conceptions of oneself and one’s relations to others that are beyond the competence of separate, asocial individuals. So “in the genuine state of nature, Amour-propre does not exist” (D2 218), because

[I]t is not possible that a sentiment which originates in comparisons he is not capable of making, could spring up in his soul: for the same reason, this man could have neither hatred nor desire for vengeance, passions that can arise from the opinion of having received some offense; and since it is contempt or the intent to harm, and not the harm itself, that constitutes the offense, men who are unable to appreciate one another or to compare themselves with one another, can do each other much violence when there is some advantage in it for them, without ever offending one another. (D2 218)

Similarly, because the moral aspect of love depends on ideas—of beauty and merit, for example—that lie beyond the natural individual's competence, such love cannot be part of our initial constitution. And absent the particular intensity of feeling that surrounds love, we would see "fewer and less cruel quarrels" (D2 155) among men. So it would be wrong to regard a state of war consequent on conflicts of the heart as a consequence of human nature's original constitution.

Apart from the implications of this view for understanding the sources of conflict, it also carries important consequences for the problem of institutional design. Because motivations depend on opinions that cannot be governed, it follows that motivations, too, are subject to the same indirect methods of control as are opinions. "Among all peoples of the world, not nature but opinion determines the choice of their pleasures. Reform men's opinions and their morals will be purified of themselves. One always loves what is fine or what one finds to be so, but it is in this judgment that one is mistaken; hence it is this judgment that has to be regulated" (SC 4.7.3), which, as we noted earlier, turns on the right form of constitutional arrangements. More particularly, the most difficult problem for a stable order regulated by a general will is to engender opinions that support motivations that contribute in turn to the preservation of just arrangements. I will come back to this point later on.

Natural Goodness

With the conception of human nature as background, we are now in a position to present the idea of natural goodness. The statement is complicated and its complexity reflects three features

of the view of human nature: the cognitive powers that are part of our natural endowment are initially present only as a potentiality; our cognitive potentiality is realized only under certain social conditions; and the realization of the potentiality produces important changes in the determinate content of the basic motivations. The consequence of these features is that the idea of natural goodness includes claims both about what we are like *antecedent to* the development of the potentialities (considered in abstraction from the “successive developments of the human Mind” (D2 159)), and about what is possible for us *given the development* of our potentialities.

Thus consider three alternative renderings of the doctrine that human beings are naturally good:

Innocence/Vice: Antecedent to the development of the cognitive potentialities, human beings are *innocent* of vice, but the development of those potentialities must lead to vice.

Brutishness/Virtue: Antecedent to the realization of our cognitive powers, human beings are brutish. But with that development, they can become virtuous.

Innocence/Virtue: Human beings are antecedently innocent and, although our potentialities may be realized in a vicious way, it is possible for them to be expressed in a virtuous form, even after the development of human powers has initially taken a vice-ridden form.

We know that Rousseau rejects Brutishness/Virtue: the idea that civilization corrects natural tendencies to human brutality and cruelty. And while it is not clear that accepting it would cause trouble for his political philosophy, I will put it to the side, and concentrate on the Innocence/Vice and Innocence/Virtue interpretations.

Of these two, Innocence/Virtue seems clearly right. Without the possibility of virtue, the doctrine of natural goodness would reduce to the idea that our nature in its undeveloped form is not evil. Natural innocence, however, is compatible with the view that vice necessarily accompanies the development of our latent cognitive powers. Rousseau suggests this when he says “the state of reflection is a state against Nature, and the man who meditates a depraved animal” (D2 138). But note how Rousseau begins the sentence from which this passage is taken: “If it [nature] destined

us to be healthy then, *I almost dare assert . . .*" [my emphasis]. Perhaps Rousseau is not here making the daring assertion.

But he seems to make it elsewhere. In the *Discourse on Inequality* he says: "It is reason that engenders amour-propre, and reflection that reinforces it" (D2 153). This remark suggests that the distinction-drawing powers of reason and reflection make us evil by separating our own fate from the welfare of others. And in the *First Discourse* Rousseau claims that virtue has ebbed in proportion to enlightenment, and that "luxury, dissoluteness and slavery have at all times been the punishment visited upon our prideful efforts to leave the happy ignorance in which eternal wisdom had placed us" (D1 13).

Both passages sharply underscore the historical correlation of enlightenment and vice. But that correlation is not in dispute between Innocence/Vice and Innocence/Virtue; so these passages provide no evidence for Innocence/Vice. The correlation of enlightenment and vice provides the point of departure and sets the problem for the doctrine of natural goodness; it does not exhaust the substance of that doctrine. On the contrary, the point is to show that the expression of originally latent human powers need not lead to vice, despite the evidence provided by the correlation.¹⁹

Moreover, Innocence/Vice is badly captured by the slogan "There is no original perversity in the human heart" (E 92) or by claims about how our social arrangements are the source of evil and vice. After all, if *Innocence/Vice* is right—if the realization of human powers must lead to vice—then the tendency to evil does lie in latent human powers themselves; there is a fundamental conflict between culture and nature; that civilization must be accompanied by all its familiar discontents. Furthermore, the last parts of *Emile*, which sketch the conditions for goodness without innocence, would have at best an uncertain role in a work that is "nothing but a treatise on the original goodness of man" (RJ 213), and at worst would refute one of the central elements of the doctrine it sets out to defend.

The theory of natural goodness is, then, captured by Innocence/Virtue, which affirms that *unsuitable* social arrangements, and not society itself, explain the unhappy path taken by the development of our natural potentialities.²⁰ Put more constructively, Innocence/Virtue holds that essential human powers *can be realized* without taking the vice-ridden form that they have

in fact taken. But such realization requires, *inter alia*, that we develop a sense of duty and motivational supports for it, thus enabling us to resist the temptations to vice ingredient in social interdependence (see below, pp. 122–7).

Because it states that the development of virtue is merely possible, this statement of the doctrine may seem too weak to capture the idea of *natural goodness*. After all, if virtue is possible, so, too, is vice. Doesn't the idea of natural goodness imply an asymmetry between vice and virtue: that our natural endowment is aimed at the good, or tends to the good, or would achieve the good were it not for distorted circumstances?²¹

This point may have some force as an objection to Rousseau's terminology, though it must be stressed that he affirms our *natural goodness*, not our *natural virtue*. More fundamentally, I do not think that his principal statements of the conception of natural goodness do affirm more than that people are naturally innocent and that virtue is compatible with our human nature—nor, in view of the Hobbesian background, is such affirmation trivial. Furthermore, Rousseau's attention to the degenerative tendencies of even well-ordered societies (SC 3.10–11, 4.1) indicates his strong reluctance to affirm a tighter connection between our nature and a virtuous life—say, a general tendency to virtue, manifest under normal conditions. If there is an asymmetry, it is only this: that under conditions of social interdependence, we can express our nature as free and thus achieve moral freedom only if we have a general will, and having such a will constitutes civic virtue. And that is to say that virtue, which is possible for us, is also required for realizing our nature.

A Genealogy of Vice

That is natural goodness, and it faces a large challenge from “sad and constant experience.” The very constancy of the sad experience suggests that wickedness belongs to our nature. To correct this misleading suggestion, the *Discourse on Inequality* presents a “genealogy” of vice (B 28–9; D2 132). Officially about the origins of inequality, Rousseau's essay offers support to the idea of natural goodness by explaining the origins of vanity, greed, hate, jealousy, envy, covetousness (and misery) without

postulating an original, non-derivative predisposition to vice.²² Why does he think that our natural powers have been expressed in this vice-ridden way?

The explanation is historical in form, and proceeds through six stages—three stages of equality, three of inequality (the inequality of rich and poor, powerful and weak, master and slave). At each stage we find characteristic forms of association, modes of cognition, and types of motivation; the circumstances (forms of association) shape the expression of the cognitive powers, which in turn shape motivation. Evolution between the different stages of association is driven by motivations generated at one stage that eventually alter the circumstances that produced those motivations, and that, by altering the circumstances, eventually transform the dominant ideas and passions, too.

Rousseau's explanation provides (and is expressly designed to provide) an alternative to Hobbes's theory of human nature and the Augustinian doctrine of original sin.²³ Rejecting Hobbes's view, Rousseau aims to explain how the vices arose, without attributing them to human nature itself; rejecting the idea of a choice of evil associated with original sin, or the associated Calvinist conception of a total depravity redeemed only through divine grace, Rousseau thinks that their emergence can be explained as a product of social arrangements that tend to channel motivations in certain undesirable directions: "The scorn which J.J. had displayed for that entire pretended social order, which in fact hides the most cruel disorders, fell much more on the constitution of the different estates than the subjects filling them and who, by this very constitution, must of necessity be what they are" (RJ 176–7).

The doctrine of original sin—whether it has the choice of evil occurring in each generation, or inherited through the generations—requires some assumption of an original predisposition to evil, even if choice is also needed to express this predisposition. Rousseau thinks he can do without that assumption, and without thinking that God is unjust for making us so predisposed (B 29–30). Where Archbishop Beaumont—the Archbishop of Paris, who had formally condemned *Emile*—believes that the source of evil is the corruption of human nature, Rousseau wants to discover the cause of this corruption (B 31). To explain the patterns of motivation and conduct described by Augustine and Hobbes, "it was not necessary to assume that man is wicked by

his nature," because the conflict between "our social order" and our nature "by itself . . . explained all the vices of men and all the ills of society" (B 52).

Structural Interpretation of Vice

Rousseau is concerned principally to account for vices that involve indifference or hostility to the welfare of others, an interest in their well-being only insofar as advancing it also advances one's own: "No one wants the public good unless it accords with his own" (B 29; SC 4.1.6). These are also the vices of most immediate interest here, because they raise the largest troubles for the society of the general will, because they create problems for the idea of an allegiance to the common good and for the underlying idea of regarding others as equals.

Rousseau traces the foundations of such vice to a concern for one's standing relative to others. Vice does not arise simply from a desire to do well, or from egoism, or from an original insatiability of human appetites, but from the thought that doing well requires (perhaps consists in) doing better, a thought rooted in the sense that affirming my worth requires acknowledging me as a better, in short, from the inflamed form of *amour-propre*: "the gentle and affectionate passions are born of self-love, . . . the hateful and irascible passions are born of *amour-propre*" (E 214; RJ 112). Once pride is in place, unhappiness follows, too. For with the desire to have more than others comes an insatiable desire to accumulate powers—the Hobbesian "perpetual and restless desire of power after power, that ceaseth only in death" (Lev. 58). Once that desire is in place, we are bound to be dissatisfied, because our finite powers can never be sufficient to meet its demands (D2 199). But with Rousseau, this insatiability is not original, nor—as Hobbes suggests—a response to such general features of human circumstance as uncertainty and a desire to protect one's current means (Lev. 58). It derives instead from the desire for relative advantage.

Premising, then, that the vices are modifications of *amour-propre*, Rousseau's genealogy of vice reduces to an explanation of this heated concern for relative advantage. More precisely, his central question, as I indicated earlier, is: *how is the generic concern to be treated with respect crystallized and expressed as a demand and desire to be treated as a better?*

Before exploring the explanation, I want to retrace the essential line of thought that leads to it by quoting a long passage from Kant's discussion of "the radical evil in human nature," from his *Religion Within the Limits of Reason Alone*. Though Kant's remarks differ in some details from Rousseau's view as I have presented it here, the fundamental idea is indistinguishable:

The predisposition to humanity can be brought under the general title of a self-love which is physical and yet *involves comparison* (for which reason is required); that is, only in comparison with others does one judge oneself happy or unhappy. Out of this self-love originates the inclination *to gain worth in the opinion of others*, originally, of course, merely *equal worth*: not allowing anyone superiority over oneself, bound up with the constant anxiety that others might be striving for ascendancy; but from this arises gradually an unjust desire to acquire superiority for oneself over others. Upon this, namely upon *jealousy* and *rivalry*, can be grafted the greatest vices of secret or open hostility to all whom we consider alien to us. These vices, however, do not really issue from nature as their root but are rather inclinations, in the face of the anxious endeavor of others to attain a hateful superiority over us, to procure it for ourselves over them for the sake of security, as preventive measure.²⁴

So the question recurs: What explains the troubling expression of the concern to acquire worth in the opinions of others?

The roots of the demand and desire lie, according to Rousseau, in social inequality. That much is clear. Less clear are the particular aspects of inequality that create troubles and the precise mechanism through which they work their psychological effects. Rousseau sometimes emphasizes that inequality leads to conflicting interests, and that such conflict makes it impossible to resist putting one's own advantage first (D2 198). But this account is unsatisfying in two ways. First, we can have conflicts of interest among equals. Moreover, second, it is not obvious why inequality produces a conflict of interest (especially if equality does not). Why not a willingness to explore mutual gains, albeit from unequal starting positions?

To be sure, *if* inequality leads people to "value the things they enjoy only to the extent that the others are deprived of them" (D2 184) or to "find our advantage in what harms our kind" (D2 198), then it is easy to see how inequality turns cooperative interactions into zero-sum conflicts. But these assertions beg the question. For the question is precisely about why inequality leads

to a concern with relative position—to a link between enjoyment and deprivation, advantage and harm. The explanation now on the table simply assumes the link.

Moreover, the troubles with the simple inequality explanation run much deeper. Rousseau does not condemn all inequality. He suggests, for example, in his *Letter to d'Alembert* that a certain measure of inequality of fortunes “can have its advantages” (LD 115). To be sure, Rousseau emphasizes the importance, in the society of the general will, of limiting inequalities of wealth and power among citizens—ensuring that no one is required to “sell himself” in order to secure basic necessities. Nevertheless, he does not think that the society of the general will is without socio-economic inequalities (SC 1.9.8 n., 2.11; PE 19; LM 300–1; P 187, 223). Understanding the connection between inequality and inflamed self-love requires, then, an account of why certain types of inequality, or inequalities under certain conditions, have these psychological effects.

The explanation, to put it very abstractly, is that certain forms of inequality, under certain conditions, lead us to identify the affirmation by others of our worth with our securing advantage over them, and for this good reason: that under the relevant conditions gaining advantage—and a socially recognized entitlement to such advantage—is the only stable, socially recognized way to secure our natural aspiration to such affirmation. To make the point less abstract, let’s consider two idealized social worlds. In the first, the world of public equality, people have equal standing as citizens in a regime that vests sovereignty in the hands of citizens as a body. By ensuring equal rights of citizenship, the political institutions in effect announce the equal worth of citizens. And given this background equality, the socio-economic inequality need not undercut the role of the political institutions in affirming our equality, upholding our worth as equals, and so establishing the conditions for free association.²⁵ In the second, a social world without public equality, there is also socio-economic inequality, but no institutionally acknowledged equality of worth running parallel to the inequalities. Under such circumstances, the natural way to win recognition from others—even for those who might be prepared to accept the egalitarian conception of worth—is to win advantage, and a socially acknowledged entitlement to such advantage. The social arrangements provide no other vehicle for achieving recognition:

as Rousseau explains in the Preface to *Narcissus*: “In Europe, the government, the laws, the customs, self-interest, everything places individuals under the necessity of deceiving one another, and of doing so incessantly; everything conspires to make vice a duty for them; they must be wicked if they are to be wise, since there is no greater folly than to provide for the happiness of scoundrels at the expense of one’s own. Among savages, self-interest speaks as insistently as it does among us, but it does not say the same things . . . ; nothing leads them to deceive one another” (N 101 n.).

I will call the interpretation of vice I have just sketched “the structural interpretation,” and will elaborate on it by sketching Rousseau’s quasi-historical genealogy.

Troubles in Paradise

The story begins with human beings in conditions of independence. Because the realization of latent cognitive powers traces to social interaction, human beings in the initial state lack the ability to conceptualize, and so are moved only by desires that arise immediately from natural needs (hunger, sex, sleep). Such individuals lack the conceptual resources for desiring to live in a better world, or wishing to be treated as equals, or acting strategically, or having a conception of honor and a passionate attachment to protecting it. Given these motivational limits, themselves rooted in the limited realization of cognitive powers, we have sufficient bodily powers to satisfy our desires—no need to compensate for a shortfall of natural powers by subordinating the powers of others—and live in a state of innocence.

This equilibrium is disrupted by “difficulties” in satisfying the natural desires, and those difficulties lead to the formation of temporary associations to facilitate such satisfaction. Temporary cooperation for specific purposes—catching deer, for example—triggers the realization of latent cognitive powers, in particular the power to represent future states of myself and of the world. It is not clear how temporary associations could form antecedent to this triggering, or why the change of circumstances prompts the expression of precisely the latent powers that are functional for the circumstances. But I do not think that these problems raise fundamental troubles, and do not propose to dwell on them here. In any case, the new “enlightenment,” itself elicited by altered conditions, stimulates a desire

for continued cooperation, if only as a way to satisfy the basic desires.

Eventually, then, we arrive at more settled forms of association—families and associations of families. More settled conditions prompt ideas about one's relations to others and thoughts about relative social positions, and those thoughts generate a shift from a natural concern for well-being to a concern about one's situation relative to others, and in particular to a concern to have *advantage* over others (D2 165–6). But while association engenders the desire for esteem—giving rise to vanity, contempt, shame, and envy—that desire and the pursuit of it are limited by the absence of specialization a division of labor and private property. Absent a form of social interdependence that enables individuals to gain and to maintain great advantages over others, individuals lack a conception of such advantages. Their sense of their own worth, and of the conditions necessary for affirming it, therefore do not depend on such advantages; nor do they believe that a failure to provide relative advantages would be a denial of their worth, and that they therefore merit such advantages, deserve them, or in any other way could claim them as a matter of right. To put the point in terms of the emotions, people are not insulted, injured, or made resentful if they are not relatively advantaged.

Here we have the final stage of equality, the happiest and most durable period because of its balance between our desires and our powers. Desires remain limited because people do not desire more than others and so are not driven to endlessly desire more for themselves: they are not egoists, because egoists are people who “pay more attention to others than to themselves” (RJ 148). Because people under these conditions are not plagued by expansive desires that they cannot satisfy, they are relatively happy, and are not impelled to transform the conditions of their association.

Against this background of limited social cooperation and nascent *amour-propre*, the advent of metallurgy and agriculture prompt specialization and private property. This change of circumstance expands the range of opportunities for pursuing the desire for advantage that is already in place. Moreover, the expanded opportunities increase the temptation to pursue the desire for advantage by increasing the benefits that such pursuit may deliver. In particular, the settlement of individual property

gives rise to inequality, in part because of the inequalities of “talents” (D2 169) that people put to work on their property.

So long as property remained common, talents could have only a limited effect on inequality because each person would share in the fruits of the benefits—for example, improvements in the land—created by the talented. But private ownership eliminates this constraint on the social effects of differences of talent: “This is how natural inequality imperceptibly unfolds together with unequal associations, and the differences between men, developed by their different circumstances, become more perceptible, more permanent in their effects, and begin to exercise a corresponding influence on the fate of individuals” (D2 170).

It is this system of private property and specialization, and the growing inequality resulting from it, that stimulates inflated *amour-propre* and vice. The reason—here we come to the structural interpretation—is that circumstances lead people to identify achieving their own aims with achieving relative advantage, and that puts the interests of individuals at odds; the identification turns conflict zero-sum: “if one sees a handful of powerful and rich men at the pinnacle of greatness and fortune while the masses grovel in obscurity and misery, it is because *the former value the things they enjoy only to the extent that the others are deprived of them*, and they would cease to be happy if, without any change in their own state, the People ceased to be miserable” (D2 184, emphasis added). But why is the natural concern with self-worth expressed as inflamed self-love? Because inequalities emerge under conditions in which recognition can only be ensured by establishing advantage—because, for example, there is no institutional form analogous to equal citizenship that enables individuals to be treated as equals irrespective of their differences. Socially speaking, there are only the differences: perceptible, permanent in their effects, and exercising a profound influence on the fate of individuals. Furthermore, differences among people (say, differences in their talents) play such a fundamental role in determining their fate, that they regard those differences of fate as reflecting intrinsic differences of worth. And this of course reinforces the identification of the natural desire for an affirmation of worth with the demand that others treat one as a better.

At the same time, enlightenment enables the formation of more expansive desires; “imagination . . . extends for us the

measure of the possible, whether for good or bad, and . . . consequently excites and nourishes the desires by the hope of satisfying them" (E 81). The more familiar vices fall out of this conjunction of vanity and enlightenment; given a desire for relative advantage, there are benefits from appearing different from the way one really is. So hypocrisy, deceit, and other forms of dissimulation are now rational (D2 170–1). Relative advantage being an essentially scarce resource, antagonisms intensify, are moralized, and tend to break out into open conflict. "Nascent Society gave way to the most horrible state of war" (D2 172)—a Hobbesian war of all against all, now understood as the effect of circumstances of inequality rather than an immediate expression of our nature.

Nature, then, endows us with desires that correspond to our needs, and powers sufficient to satisfy those desires: it provides a basis for an innocent freedom without unhappiness or vice (E 80). Inequality encourages a concern for relative advantage and a sense of deep insult—of being wronged—if it is not achieved. Working together with the enlightenment engendered by social association, this concern expands our desires without triggering a commensurate expansion of our capacity to satisfy them. So we are finally beset by desires that we can only pursue through social association, in particular through subordinating the powers of others to our own, although we also recognize that participation in such association implicates us in a system of mutual abuse, manipulation, hypocrisy, and deceit.

If this explanation is right, then inflamed self-love and the vices flowing from it create no troubles for the doctrine of natural goodness. The social explanation implies that inflamed content is not intrinsic to the abstract desire for the social affirmation of worth. The very same feature of our nature that is expressed in the form of vice could, as a matter of its intrinsic character, be expressed in other ways. In particular under alternative conditions each person could be satisfied that their worth was affirmed if they were regarded as an equal. And if they were, then the worth of each could in principle be affirmed, while the prideful form of self-worth makes such mutual affirmation impossible.

This [when Emile begins to compare himself to others] is the point where love of self turns into *amour-propre* and where begin to arise all

the passions which depend on this one. But to decide whether among these passions the dominant ones in his character will be humane and gentle or cruel and malignant, whether they will be passions of beneficence and commiseration or of envy and covetousness, we must know what position he will feel he has among men. (E 235)

How, then, might the affections take this more attractive path?

Complementary Motivations

According to the doctrine of natural goodness, we are naturally innocent and capable of virtue. I have discussed the idea of natural innocence and the etiology of vice, and want now to sketch the ways in which the conception of human nature underwrites the possibility of virtue, and so provides the basis for addressing the problem of motivational possibility. Premising that compassion and concerns about self-worth are intrinsic to our nature, and that their determinate expression depends on social-institutional setting, how might these properties be expressed in a virtuous form?

The Sense of Duty Is Not Enough

Answering this question requires some initial clarification of the “virtuous form” of expression. To provide that clarification I will make some initial, schematic observations about Rousseau’s account of the sense of duty:

1. Social association inevitably encourages the development of various human attachments and affections—for example, love for particular individuals. Rejecting the Stoic ideal of extirpating the passions, Rousseau thinks that such attachments are unavoidable (see n. 21).
2. Attachments and affections are not a reliable guide to correct conduct: “As man is not guilty for loving another’s wife if he keep this unhappy passion enslaved to the law of duty. He is guilty for loving his own wife to the point of sacrificing everything to that love” (E 445).
3. In a world of passions and attachments, the only reliable guide to right conduct is a sense of duty (the “law of

duty"). In this world, "Goodness is broken and perishes under the impact of the human passions. . . . Who, then, is the virtuous man? It is he who knows how to conquer his affections; for then he follows his reason and his conscience; he does his duty" (E 444–5).²⁶

4. The capacity for a sense of duty, awakened by social association, is part of our native endowment. Speaking through the Savoyard Vicar, Rousseau explains: "There is in the depth of souls . . . an innate principle of justice and virtue according to which, in spite of our own maxims, we judge our actions and those of others as good or bad" (E 289; also 66, 267, 290). And "Rousseau" in *Rousseau, Judge of Jean-Jacques* puts the point still more strongly: "The voice of conscience can no more be stifled in the human heart than that of reason can be stifled in the understanding; and moral insensitivity is as unnatural as madness" (RJ 242).
5. Though a sense of right is part of our natural endowment, the content of that sense is not similarly fixed by nature. In particular, its content in the society of the general will is given by laws issuing from the shared sense of the common good that defines the general will.
6. We cannot have much confidence in the motivational strength of the sense of duty. In political society, citizens face constant temptations to subordinate their concern for the common good to other interests; the larger the temptation the more likely the subordination (see below, pp. 156–64). Indeed, this is the "great lesson of morality": that we should "avoid situations that put our duties in opposition with our interests, and which promise us good in the bad of others; certain that in such circumstances, however sincere our love of virtue may be, sooner or later one weakens, and becomes unjust and wicked in fact, without having ceased to be just and good in the soul" (C 56).

Taking these points together, then, the problem is to find *motivational complements* to the sense of duty. Although motivational complements are distinct from the sense of duty, they generally support it. The content of these motives is such that they support our efforts to comply with our duties, as those duties are specified by the general will, rather than strengthening our disposition to subordinate the general will to

our particular interests.²⁷ My earlier discussion of the *content problem* indicates the kinds of motivational complements that are needed: because the general will is oriented to the common good, the problem is to find possible sources of motivation for treating others as equals—to show that and how the demands of the general will can engage our affections. The earlier discussion of the theory of human nature suggests two such sources of engagement, and I will add a third. When we take all these complements together, we have a picture of how the general will might have motivational grip, without drawing so exclusively on communitarian solidarities.

Self-Worth, Reciprocity, Generalized Compassion

The desire for relative advantage associated with an inflamed sense of self-worth is plainly the enemy of the general will, and an especially important enemy because the sense of self-worth is psychologically fundamental and resistant to change: a sense of one's greater worth "is the error most to be feared, because it is the most difficult to destroy" (E 245). So the requirements for achieving a sense of self-worth must not by and large conflict with the demands of duty, as those demands are specified by the general will. One aspect of psychological support for the sense of duty, then, is to avoid the inflamed form of self-love and to encourage instead an understanding of others as equals—of "the identity of our natures with theirs" (E 221)—and a correspondingly egalitarian conception of self-worth as requiring treatment as an equal.²⁸

Although the sense of self-worth is not intrinsically egalitarian, its intrinsic features can be developed along these lines. Its content depends on our beliefs about our worth, which depend in turn on circumstances. When social arrangements are such that our true relations as equals are manifest in social experience and differences are neither so perceptible, nor permanent, nor influential—say, when individuals have the public status of equal citizen in the society of the general will—individuals can reasonably be expected to acquire an understanding of one another as equals. And if we do regard each other as equals, then those same arrangements will—for example, by establishing citizenship rights, and taking the interests of each into account—confirm our sense of our own worth. The sense of

duty has motivational support from the sense of self-worth, then, when the latter is informed by an understanding of equality.

To strengthen this point about compatibility, let's add Rousseau's idea that reciprocity is important to the formation of our motivations: "those from whom one expects good or ill by their inner disposition, by their will—those we see acting freely for us or against us—inspire in us sentiments similar to those they manifest toward us" (E 213). In this way—through reciprocation—"the gentle and affectionate passions are born of self-love" (E 214). Assume, then, a society with a general will, which encourages the more egalitarian conception of self-worth. If others comply with the requirements of the general will, they uphold my sense of self-worth, generating a disposition to reply in kind and uphold their sense of their own worth. And that disposition binds us by affection to the common good, the source of obligations in the society of the general will.

If reciprocity leads to attachments to others as *agents* who act for our well-being, compassion leads us to concern for others as *sufferers*, "fellow-passengers to the grave": "we are attached to our fellows less by the sentiment of their pleasures than by the sentiment of their pains, for we see far better in the latter the identity of our natures with theirs and the guarantees of their attachment to us. If our common needs unite us by interest, our common miseries unite us by affection" (E 221). Despite its importance, compassion—as all passions—is not an entirely dependable guide to conduct. To ensure that it conforms to duty, it must be "generalized and extended to the whole of mankind. Then one yields to it only insofar as it accords with justice" (E 253).²⁹ When it does so accord, its presence strengthens the motivations to comply with the requirements of justice. In a striking passage in *The Metaphysics of Morals*, Kant perfectly captures Rousseau's thought about the potential for compassion to serve as a motivational complement to the sense of duty: "compassion [*mitgefühl*] . . . is one of the impulses placed in us by nature for effecting what the representation of duty alone would not accomplish."³⁰

But the extension and generalization of compassion is not automatic, not intrinsic to compassion. It, too, requires a sense of others as equals, and having that sense depends, once more, on circumstances:

Why are kings without pity for their subjects? Because they count on never being mere men. Why are the rich so hard toward the poor? It is because they have no fear of becoming poor. Why does the nobility have so great a contempt for the people? It is because a noble will never be a commoner. . . . Do not, therefore, accustom your pupil to regard the sufferings of the unfortunate and the labors of the poor from the height of his glory; and do not hope to teach him to pity them if he considers them alien to him. Make him understand well that the fate of these unhappy men can be his, that all their ills are there in the ground beneath his feet, that countless unforeseen and inevitable events can plunge him into them from one moment to the next. Teach him to count on neither birth, nor health, nor riches. Show him all the vicissitudes of fortune. (E 224)

Here we have the kernel of a program of individual instruction aimed at giving emotional immediacy to the thought “there but for the grace of God go I”—at encouraging a self-understanding in which equality plays a fundamental role, and generalizing compassion to all humankind.

More broadly, Rousseau describes a path of moral development (E221–55) which directs our “nascent passions” to goodness and humanity in three stages. In the first, we direct our compassion on those who suffer, focusing on forms of suffering to which we are susceptible ourselves. Next, we are to consider people in society, and study how they—particularly the more powerful—mask their suffering and weakness: frailty, we then understand, is part of our human condition. Finally, we confront our own weakness, and see ourselves as objects of compassion, not merely as its subjects. At each stage, the confirmation of the sense of equality generalizes and reinforces compassion, ensuring that it takes the common good as its object. Here, an “ordered development of our primitive affections” (E 235) binds our affections to our duties, as both are directed to the common good.

The political problem is to transform this program of individual instruction into proposals about institutional design. Socio-political circumstances must foster the experience of others as equals, thus discouraging inflamed *amour-propre* and the forms of vice and conflict that follow on it, generating—through reciprocity and extended compassion—motivational complements to the general will, and so establishing a stable political order that answers to the concerns of self-love and in which citizens give the law to themselves. A detailed discussion of this issue

awaits Chapter 5. But to clarify the terms of that discussion, I will sketch its contours here.

Rousseau's answer is to establish the position of equal citizen in an association guided by a conception of the common good. Given public conditions of equality—both in the rights associated with the status of equal citizen and in the content of the general will—I come to see myself as an equal; others, therefore, affirm my worth when they treat me as an equal. Because of that affirmation, the society of the general will discourages the inflamed form of self-love. But when there is a supreme general will each person is in fact treated by others as an equal. So the existence of a general will ordering the terms of social association not only assigns equal weight in settling the laws to each citizen's security and liberty interests, it also provides a way to affirm the equal worth of each even in the face of the inevitable differences in the social and economic circumstances of different citizens.

Here, too, we find Rousseau's strongest case for a "participatory" politics that brings citizens together under manifestly equal conditions. Public deliberation proceeding among equals visibly guided by reasons of the common good provides institutional expression for the egalitarian form of self-regard and helps to extend compassion, thus discouraging inflated forms of *amour-propre*. I will return to these issues later. But I have said enough here to make the essential story about compassion clear: appropriate institutional design encourages "an ordered development of the primitive affections," leading to a generalization of compassion; compassion, suitably generalized, prompts us to act for the common good, thus complementing—though never substituting for—the motivations that arise from the sense of duty itself.

Natural Goodness and Reasonable Faith

I have sketched Rousseau's conception of human nature and his doctrine of natural goodness, but have not yet mentioned any reason for endorsing these views.³¹ Each person may be, as Rousseau says, "naturally moved to believe what he wishes" (RJ 239). But does an attraction to the conception of natural goodness merely

reflect this natural disposition to wishful thinking, to resisting views that seem too bad to be true? To show that it does not, we need to address two concerns: that there may be decisive evidence against the doctrine (for example, evidence provided by our experience of human vice); and that even if we can explain away that evidence, there may just be no positive reason for affirming it. Consider these objections in turn.

Recall Rousseau's injunction that we "begin by setting aside all the facts, for they do not affect the question" (D2 132). Aiming to clarify "the question" to which the facts are immaterial, Rousseau says, "Inquiries that may be pursued regarding this Subject [the foundations of society] ought not to be taken for historical truths, but only for hypothetical and conditional reasonings" (D2 132). And he concludes the *Second Discourse* by restating that he has provided "an account of the origin and the progress of inequality, the establishment and the abuse of political Societies, in so far as these things can be deduced from the Nature of man by the light of reason alone" (D2 188). Although his argument is historical in form, then, it is plainly not intended as genuine history, which is not a series of deductions from our human nature, relying solely on reason. This observation, however, simply shifts the question. Setting aside the facts may be perfectly in order in writing a hypothetical history. But what precisely is the point of hypothetical history?

The answer lies in the problem of motivational possibility. Rousseau's ideal of free association is possible for human beings only if the vices are not part of human nature. But "sad and constant experience" suggests that they are: indeed the constancy of the experience may incline us to think that the vices can only be explained on the assumption that they reflect our nature. That evidence needs to be explained away, shown to be consistent with our natural goodness: with our original innocence, and with the possibility of virtue. Rousseau must give us an account of our nature and show us why the facts of vice do not undermine it. He shows this by explaining how his conception of our nature—which supports the society of the general will—could explain the evolution of vice, how it could provide an explanation that shows instead "for each one [vice], how and whence it entered" the human soul (E 92). I emphasize that he must show that the basic properties *could*, working in combination, generate the phenomena. It does not matter if the details are

wrong: even if they are, we can see that the facts themselves do not drive us to a pessimistic outlook on human possibilities. Of course, that leaves the possibility that the vices *are, after all*, intrinsic to our nature: Rousseau has, I believe, no compelling argument against this possibility. But he does have a plausible case that the principal reason for believing that they are part of our nature—the reason provided by dismal experience—is not as compelling as we might have thought.

The genealogy of vice, then, provides part of the support for the theory of human nature that underwrites the doctrine of natural goodness. Nothing that we know “from sad and constant experience” about the ubiquity of vice forces a rejection of the idea that people are naturally good: it is “not necessary to assume that man is wicked by his nature” (B 52). And so nothing forces us to reject the political ideal as humanly unrealizable because motivationally impossible.

The genealogy, however, only supports the claim that the theory of natural goodness is compatible with our observations about human motivations and conduct. It does not itself provide a positive reason for endorsing the theory, because other, more Hobbesian views seem to be compatible as well. Are there such reasons?

The answer lies in the Profession of Faith that Rousseau attributes in *Emile* to the Savoyard Vicar (E 266–94).³² After describing the inability of traditional metaphysical theorizing to resolve fundamental questions about, for example, the nature of the self, the Vicar offers an alternative to a Cartesian “clear and distinct ideas” test for knowledge. He will “accept as evident,” he says, “all knowledge to which in the sincerity of my heart I cannot refuse my consent” (E 270). Even the sincerest heart may strike us as an unpromising place to look for knowledge. But Rousseau attaches a moral force to the phrase “sincerity of my heart.” He explains, for example, that a belief in the immortality of the soul passes his test because if the soul were mortal, then the wicked would triumph—and “that alone” suffices to establish the case for immortality (E 283). Similarly, the Vicar links his faith that human beings are free to the experience of remorse and self-reproach for succumbing to temptations to vice (E 280–1).

The Vicar’s line of thought, then, is that we have reason to accept beliefs that are called for by our moral convictions,

on the assumption—suggested by the Vicar’s own criticisms of metaphysics—that those beliefs are not inconsistent with what we do or can know. And this provides a rationale for endorsing the theory of natural goodness. The ideal of free social cooperation is compatible with our nature only if we are naturally good. Experience of human motivations suggests that we are not. The genealogy of vice provides a way to explain that sad experience while endorsing the doctrine of natural goodness. Moreover, we must endorse the doctrine of natural goodness—or something like it—if we are to hold out the prospect of realizing the society of the general will: if, in particular, we are to address the problem of motivational possibility. Endorsement of the doctrine of natural goodness—Rousseau’s defense of human nature—is recommended, then, on moral-practical grounds. And that defense is not contradicted by the facts and so is acceptable to theoretical reason. We may, therefore, use it to address the problem of motivational possibility. And that address permits us to hope, with reason, for a society that answers to the demands of self-love and freedom. Of course, “sad and constant” experience must limit our optimism. But hope is not optimism. And experience can only overturn moral hopes whose achievement is, in Kant’s words, “demonstrably impossible.”³³

5

Democracy^r

A free community of equals: that is Rousseau's solution to the fundamental problem presented by our nature as socially interdependent, self-loving, free agents. And if we are, as Rousseau urges, naturally good—both originally innocent and capable of virtue—then living in a free community of equals is not excluded by what we humans are like. We are capable of assigning in thought, feeling, and conduct the importance to the common good of the members of a society of equals required in the society of the general will. If Rousseau is right, then Hobbesians are wrong. Protecting the basic interests that grow from self-love does not demand the subordination that Rousseau condemns as morally unacceptable; self-government is not excluded by facts of human nature and social life.

To complete the case that a free community of equals is humanly possible—even abstracting from issues of accessibility—we need to see more precisely how it can be realized in an ongoing society. How can the priority of the common good, mutually willed by members, be realized in a stable institutional system, and more particularly in the thought and motivations of citizens? Even if the society of the general will is motivationally *possible*—compatible with our cognitive and motivational powers—it might be too demanding informationally; or it might require a degree of civic involvement that would defeat other social purposes; or it might require political units too small to be self-sufficient, stable, or secure; or it might be unsuited to conditions of organizational pluralism, economic markets,

and administrative complexity; or it might not be possible to encourage the commitment to the common good required in a community of equals.

This is the second concern about realism: the problem of institutional possibility. Much of the *Social Contract* is devoted to addressing this issue, as are large sections of Rousseau's post-*Social Contract* discussions of constitutional issues in Poland, Corsica, and Geneva. If what Rousseau says is right, then a free community of equals is difficult to achieve, hard to sustain, but not an unrealistic ideal.

The most striking feature of Rousseau's institutional views and the one that naturally draws our attention is his endorsement of a directly democratic system of lawmaking, which has citizens themselves deciding in a legislative assembly on the substance of the laws to which they will be subject. He is correspondingly hostile to political systems—really, forms of subjection—in which representatives fix the substance of the laws. When laws are made by representatives, he says, "citizens" are slaves. The strongest statement is in the *Social Contract*: "Any law which the People has not ratified in person is null; it is not a law. The English people thinks it is free; it is greatly mistaken, it is free only during the election of Members of Parliament; as soon as they are elected, it is enslaved, it is nothing" (SC 3.15.5). In 1764, he worried about the destruction of popular legislative authority in Geneva, where *Emile* and the *Social Contract* had been burned and a warrant issued for his arrest. Although there was a popular legislative assembly (the general Council), it had fallen under the sway of the magistrates, the executive power which was depriving the people (as executive powers do) of their legitimate legislative powers: "In the general Council your Sovereign power is enchained: you cannot act except when it pleases your Magistrates, nor speak except when they interrogate you. If they even want not to assemble the general Council at all, your authority, your existence is annihilated. . . . In sum, if you are Sovereign Lords in the assembly, upon leaving there you are no longer anything. Four hours a year subordinate Sovereigns, you are subjects the rest of your life and abandoned without reserve to the discretion of someone else" (LM 238)

This commitment to popular, participatory lawmaking is an important element of Rousseau's view. But an *exclusive* focus on this feature naturally leads to misunderstanding of Rousseau's

picture of institutional design, confusion about the basis of his commitment to direct democracy itself, and an excessively close link between his more abstract ideal of a free community of equals and a particular institutional idea that may be suited to achieving such a community.

To appreciate the problem, consider a natural criticism of Rousseau's theory of direct democracy. According to the criticism, Rousseau does not proceed from general and abstract standards of political legitimacy that have wide reach to specific institutional forms that satisfy those principles under particular conditions. Instead, he simply identifies the principle of legitimacy—the general will's sovereignty, understood as a solution to the initial contract—with the institutional requirement that laws be decided on in popular assemblies. Thus he is alleged to have conflated an abstract and hypothetical test of consensual legitimacy—would an arrangement be agreed to by individuals concerned to secure their interests and remain free—with a specific claim about how popular consent is to be institutionally organized, in particular with the view that citizens must actually consent in person to the rules of their association.

Juergen Habermas gives us a particularly sharp formulation of the criticism. He claims that Rousseau takes the standard of consensual legitimacy to lead directly to "an a priori preference for a specific type of organization, for example, so-called direct democracy." According to Habermas, this a priori preference collapses "a level of justification of domination" [that is, an account of what makes authority legitimate] "with procedures for the organization of domination" [that is, with ways of exercising authority]. This conflation is, Habermas argues, common to a variety of normative theories of democracy, and has "confused the discussion of democracy right up to this day." The confusion Habermas is here pointing to is the claim that the idea of consensual political legitimacy leads immediately to a participatory conception of democracy, in which citizens must directly, in person and all together, embrace the laws to which they are subject. Habermas contrasts Rousseau's error with what he sees as the proper approach to democratic theory: "If one calls democracies precisely those political orders that satisfy the procedural type of legitimacy, then questions of democratization can be treated as what they are: as organizational questions. For it then depends on concrete social and political conditions,

on scopes of disposition, on information, and so forth, which types of organization and which mechanisms are in each case better suited to bring about procedurally legitimate decisions and institutions."²

Habermas's characterization of the aims of a normative account of democracy is very rough, but seems more or less correct. But I think he has misunderstood the connections in Rousseau between his contractual (or consensual) theory of political legitimacy and his participatory theory of democracy. Rousseau's account of institutions—including his defense of direct democracy—does not show an a priori preference for a particular form of collective decision-making, nor does he simply identify a conception of legitimacy—either the contractualist thesis that legitimate authority depends on agreement or the ideal of the society of the general will, with its mutual commitment to assign priority to the common good—with a particular (directly democratic) way of organizing consent to the exercise of coercive power. Instead his view exhibits a complex structure of argument with intricacies inherited from both the conception of the society of the general will, with the four conditions identified earlier, and the account of the social formation of motivations that figures in the account of natural goodness. It is easy to miss this structure, however, if one concentrates exclusively on his views about direct democracy.

To forestall such confusion, and to get a better understanding of the rationale for directly democratic lawmaking, I begin with an overall sketch of Rousseau's substantive institutional views. The sketch is intended as a description of those views, and is not meant to provide a rationale for them. Then I provide a general framework of institutional argument—a set of conditions that must be met if institutions are to ensure the general will's supremacy. Next, I indicate how Rousseau's various institutional commitments and proposals can be explained within this structure.

I conclude by replying to a line of argument according to which Rousseau is not really much of a democrat. The criticism may seem surprising in view of the prominent place that Rousseau assigns to popular assemblies with a lawmaking function—their essential role in establishing the legitimacy of legal rules, these "assemblies of the people which are the shield of the body

politic and the curb of Government" (SC 3.14.2). According to the criticism, this apparent role is misleading. Rousseau substantially limits, according to this criticism, popular discussion and choice of laws, and gives a large role to background culture and executive power in structuring popular lawmaking. Indeed, the limits are sufficiently great that we would do better to regard the meetings of the popular assembly as a kind of public festival, designed to reinforce social solidarity and national allegiance, and embrace laws proposed by the executive, rather than as occasions for the exercise of popular power and the organization of consent. I think this line of thought is misguided, but also think that there is much to be learned by responding to it.

Some Institutions of the Society of the General Will

To begin with, then, I will sketch out a range of Rousseau's principal institutional commitments. I have grouped these commitments into three categories, though nothing much turns on the categorizations. The description is not comprehensive, but is meant to say enough to meet my two principal aims: to illustrate the range and complexities of the argument about how to bring together the conception of a free community of equals with the account of our nature embodied in the theory of natural goodness; and to explain the place of democracy in Rousseau's view of the society of the general will.³

Legislative Power

Rule of Law. There must be a system of public and general rules that applies equally to all citizens (SC 2.6), and the exercise of political power is to be confined to the application and enforcement of those public and general rules.

Emphasizing the importance of *generality*, Rousseau says: "when the whole people enacts statutes for the whole people, it considers only itself, and if a relation is then formed, it is between the entire object from one point of view and the entire object from another point of view. . . . Then the matter with regard to which the statute is being enacted is general, as is the enacting will. It

is this act which I call law. When I say that the object of the laws is always general, I mean that the law considers the subjects in a body and their actions in the abstract, never any man as an individual or a particular action. Thus the law can very well state that there will be privileges, but it cannot confer them on any one by name." More broadly, the legislative power cannot be concerned with "any function that relates to an individual" (SC 2.6.6).

The importance of *public* rules is implicit in the idea that the laws come only from the people, who in making them, know them. And the importance of cabinining the exercise of political power, in particular executive power, to the application and enforcement of such general and public rules is connected to Rousseau's central theme of autonomy and authority: "A free people obeys, but it does not serve; it has leaders and not masters; it obeys the Laws, but it obeys *only the Laws* [emphasis added] and it is from the force of the Laws that it does not obey men. . . . A People is free, whatever form its Government has, when in the one who governs it one does not see the man, but the organ of the Laws" (LM 261).

Direct Democracy. Among the laws is a specification of the nature of and conditions for the exercise of legislative power: that is, a "constitutional" law defining the legitimate exercise of lawmaking power. In particular, the laws must be made in a legislative assembly open to all citizens (SC 3.12–15, 18), or by representatives to a legislative assembly who are *elected*, subject to *frequent review*, and given *specific and binding instructions* by citizen assemblies. This second alternative—especially prominent in the discussion of constitutional reform in Poland—is a strategy suited to large states, where "the legislative power cannot show itself as such, and can only act by delegation" (P 200–1). To avoid the "corruption" that characteristically (as in the English Parliament) flows from such delegation of lawmaking, and turns "the organ of freedom [the legislature] into the organ of servitude," the remedies include frequent meetings of the legislature with regular elections aimed at increasing the rate of turnover (the idea, in the spirit of term limits, is that "often changing representatives makes it more costly and more difficult to seduce them," P 201), and insisting that "representatives . . . adhere exactly to their instructions, and . . . render a strict account of their conduct in the Diet to their constituents" (P 201).⁴

Law and Legislative Power. Whatever its precise form, whether a legislative body of the whole or a representative assembly, the legislative body must have *regular, periodic meetings*. The schedule of meetings is to be specified at law, though no “precise rules” can be given about the frequency with which meetings of the assembly must be held. Decisions taken at informal, popular gatherings are not authoritative, regardless of the size of those informal gatherings or the unanimity of sentiment displayed in them, “because the order to assemble must itself emanate from the law” (SC 3.13.3; LM 271).

Suffrage. In determining the laws, the *votes of each citizen must be counted* (SC 2.2.1 n.).

Majority Rule. New legislation requires the approval of *at least a simple majority* (i.e., one more than half the voters). But simple majorities should be decisive only in the case of “deliberations which have to be concluded straightaway” (SC 4.2.10–11). It is both legitimate and reasonable to require support from qualified majorities in order to legislate on “more important and serious” (SC 4.2.11) issues. Among the more important issues are such constitutional questions as the form in which legislative power will be exercised, the frequency of legislative assemblies, the form of government (see ‘Executive Power’ below), and the decision rules of the assembly itself.

Information. Legislative decisions should take place after the people is well informed (*suffisamment informé*) (SC 2.3.3).

Executive Power

Delegation of Governmental Power. Government—understood abstractly, as the legitimate exercise of an executive power guided by law (SC 3.1.6–7)—should be entrusted at least in part to *officials* rather than exercised entirely by the sovereign people themselves. Although government in this sense is necessary in a society governed by laws that express the general will, the best form of government—whether executive power should be organized as monarchy, aristocracy, democracy, with one, a few, or many exercising executive power—varies according to the circumstances of the state (SC 3.1–8).

Minimal Democracy Requirement. In its exercise of executive power, the government *represents* the sovereign people. But while the sovereign’s executive power—its authority to give effect to the laws it adopts—can be represented, it cannot be

alienated: “the act by which a people subjects itself to chiefs is not a contract. . . . It is absolutely nothing but a commission, an office in which they, as mere officers of the Sovereign, exercise in its name the power it has vested in them, and which it can limit, modify, and resume, since alienation of such a right is incompatible with the nature of the social bond and contrary to the aim of the association” (SC 3.1.6). Part of the force of this inalienability of executive power is that, under all forms of government, the sovereign people reserve authority to change the form of government (the terms of the commission or agency relationship to the executive) as well as the officials exercising governmental powers (SC 3.18.6–8). But the choice of officials is itself an act of government, an executive act because it applies or executes the basic law defining the form of government and implements the general rules specifying legitimate accession to power. Moreover, each regular meeting of the popular assembly must begin by reviewing the officials who hold executive power, by asking “*whether it please the People to leave its administration to those who are currently charged with it*” (SC 3.18.6–8).

Legitimate government itself must, therefore, be at least to this minimal extent democratic. And this is so, notwithstanding Rousseau’s observation that popular sovereignty is compatible with a fully aristocratic or monarchical form of government, and that a “genuine democracy”—a system of government in which executive as well as legislative power is vested in the people as a body—“never has existed and never will exist” (SC 3.4.3).

Government Accountability. The people, as sovereign authority, have the right to *oversee* the administration of the state—that is, the execution of its laws—and to overturn the government’s interpretation of those laws. “The Legislative power consists in two inseparable things: to make the laws and to maintain them; that is to say, to have inspection over the executive power. There is no State in the world in which the Sovereign does not have this inspection. Without that, all subordination lacking between these two powers, the latter would not depend on the other at all” (LM 247–8). Or again, “the Legislator, existing always, sees the effect or the abuse of its laws: it sees whether they are followed or transgressed, interpreted in good or in bad faith; it watches over it; it ought to watch over it; that is its right, its duty, even its sworn oath” (LM 265, and 238–9, 249–50,

267; P 200). Aside from this power to monitor the government in its execution of the laws, the people retain control of taxation as well, and thus control the government in part by controlling the revenue available to it (LM 243).

Social Background

Rights of Assembly. Citizens have a right to assemble, and in their assembly have a right to discuss public issues. Even “in the most absolute Governments assemblies of communities that are not presided over by any Magistrate are permitted” (LM 270). This right of assembly and discussion may seem to conflict with Rousseau’s proposal that in arriving at political decisions citizens should deliberate with “no communication among themselves” (SC 2.3.3). For reasons I mentioned earlier, however, it would be a mistake to construe the condition on communication as directed against assembly and discussion. The worry expressed about communication in the lead-up to legislation is with the division of the population into a few organized factions each of which shares common interests that set them at odds, and above all with the problem of majority faction, when “one of these associations is so large that it prevails over all the rest” (SC 2.3.3). When there are such divisions, in particular when there is a majority faction, we cannot expect that legislation will advance common interests.

Rousseau endorses (see below) both a no-faction solution and solution with many factions with relatively equal powers: perfected pluralism. But he never suggests that an appropriate way to handle the problem of faction is to simply deny the right to assemble. To the contrary, Rousseau criticizes the Genevan political settlement of 1738 (the so-called *Règlement*) for imposing just such a bar: “In order to assure public tranquility, they [the Mediators who proposed the settlement] judged it appropriate to separate power from right and suppress even peaceful assemblies and deputations from the bourgeoisie” (LM 268).⁵

Associations. There must be either no formally organized political groups, or, if there are such groups, then they should be both numerous and relatively equal in power (SC 2.3.4).⁶

Property Rights. The form of property rights may vary, from more individual to more public ownership. But whether the holders of property are individuals or public agencies, they are

to be understood as “trustees” of goods owned by the public, and “the right every individual has over his own land is always subordinate to the right the community has over everyone” (SC 1.9.6–7). The fact that the community has a right to all is reflected in the fact that the structure of property rights is subject to public judgments of the common good as expressed in legislation concerning property rights, and not determined by rights that are prior to and independent from the common good.⁷

Inequality. While a certain measure of material inequality “can have its advantages” (LD 115), inequalities of wealth and power among citizens must be limited so that no one is required to “sell himself” in order to secure basic necessities. Thus, everyone should “have something” and no one should have “too much of anything” (SC 1.9.8 n., 2.11.2; PE 19; LM 300).⁸ Moreover, this concern with material inequalities extends even to a concern about inequalities within the wealthier class. In his discussion of the social foundation of a new Polish constitution, Rousseau tells us that the “enormous disparity of fortune which separates the high and the lower nobility is a major obstacle to the reforms required to make love of fatherland the dominant passion. As long as luxury reigns among the great, cupidity will reign in all hearts” (P 188).

Principles and Institutions

The reasoning that supports these conclusions is not always explicit or especially detailed. But Rousseau does defend them, and that defense has a discernible and interesting structure. Moreover, the structure of the reasoning is complex, inevitably so, because of the complexity of the four conditions that define the society of the general will, and the intricacies of Rousseau’s psychological views. So several distinct kinds of consideration figure in the theory of institutions.

Consider, for example, the Common Good Condition (GW₂). It requires that there be a shared conception of the common good, founded on concern for the interests of each. But the presence of such a conception (much less the priority and politically regulative role required by the other conditions on the society of the general will) cannot simply be assumed. Rousseau

needs to show how, given certain conditions of association and assumptions about motivation-formation, citizens will acquire a shared conception of the common good. Because of the Priority Condition (GW₃)—which requires that citizens assign priority to reasons of the common good—he must show how the conditions of association can be expected to motivate citizens to act on that conception, and how those conditions provide some assurance that others will not take advantage of their willingness to advance the common good. Because of the Reasonable Confidence Condition (GW₄)—requiring reasonable confidence in the supremacy of the common good—he needs an account of how to aggregate different opinions about the common good into a collective judgment whose implementation can reasonably be expected to advance the common good.

To highlight these different elements and to clarify the way that they work together in providing a conception of institutional design, I will begin by describing two stylized ways of thinking about norms and socio-political institutions, and then suggest that Rousseau builds on both.

Two Conceptions of Institutional Design⁹

One approach to normative questions of institutional design is founded on the notion of individuals as rational agents. Arguably, this approach traces to Machiavelli, but it finds crisp classical articulation in Bentham's conception of an "artificial identification of interests"—in what he calls the "duty-interest junction principle"—and in modern work in the theory of mechanism design and implementation.¹⁰ The second approach is an ethical conception of institutional design characteristic of more sociological theories of institutions.

Rational Choice/Incentive Alignment View. A rational choice conception of institutional design is organized around two fundamental elements: a norm for evaluating social outcomes—say, a social welfare function of some kind—and the idea that individual action is guided by a coherent ordering of alternative outcomes, where the individual ordering is distinct from the ordering given by the norm. Commonly the norm is some sort of welfarist principle—in Bentham's case, the principle of utility—but nothing in the basic logic of the rational choice conception of institutional design requires a welfarist normative conception. Instead, three aspects of the conception of preferences and norm are essential.

First, for the purposes of issues of institutional design, one takes preferences as given, and abstracts from the role of institutions in shaping what people prefer and the intensities with which those preferences are held: call this the *fixed preferences* assumption.

Second, the motivations—the determinants of individual conduct—are understood as “preferences” in that the crucial dimension of their variation is *intensity*. They are not a set of diverse reasons which are appropriate to different practical contexts: call this a “reductionist psychology.”

Third, while the rational choice conception of institutional design specifies a norm for evaluating outcomes, it assumes that individuals are not themselves directly motivated by that norm; they do not treat the fact that an institution or policy satisfies the norm as itself a reason for complying with that institution or for supporting that policy. To capture the fact that the norms do not themselves motivate, I will call them *external norms*.

Taking these three aspects together, then, the rational choice conception of the task of institutional design focuses on arranging a scheme of incentives (punishments and rewards, taxes and transfers, perquisites of office, etc.) such that rational individuals—who are assumed, as the *external norms* condition indicates, to be indifferent to the norms themselves—will act (for example, reveal private information) in ways that lead to outcomes that satisfy the external norm. If we think of the institutional design problem in game-theoretic terms, then the task is for the designer or planner to announce a scheme of incentives such that, in the game’s equilibrium, the conduct of individual agents leads to an outcome that satisfies the external norm. If the norm were, for example, the classical “happiness-maximizing” principle of utility, then the designer would establish a system of incentives to induce rational individuals to act, in equilibrium, in ways that generate a greater sum of happiness than is possible under alternative designs. Bentham called this alignment of conduct and norm—this junction of duty and interest—*artificial* because the agents themselves are, even if cognizant of the norm, not motivated directly by it.

Ethical Conception. An ethical conception of institutional design offers a fundamentally different account of the alignment of motivations with the principles and values that the society ought to satisfy. The central idea is that individuals may come

to be directly motivated by the norms and values that ought to regulate institutions. There are three salient differences from the rational choice conception, corresponding to the three features I just described.

First, then, the ethical conception rejects the fixed preferences assumption. At least some of the politically relevant motivations are not given but rather are learned through participation in institutions—for example, by occupying institutional positions defined by the norms that the arrangement is designed to satisfy.

Second, individuals are assumed to learn the values by mastering the norms that define their various institutional roles, and to develop a desire to advance those values. Thus the ethical conception also rejects the reductionist psychology, in supposing that we internalize norms that are appropriate for guiding conduct in particular settings.

Finally, the external norms condition is rejected: it is assumed instead that individuals will typically seek to fulfill the duties and to live up to the ideals associated with institutional roles because they will recognize that the institutions embody the principles and values and, having internalized those principles and the values, they will want to act on them.¹¹

Assessment. Neither of these conceptions is adequate on its own either for the purposes of a normative theory of institutions generally, or—what is germane here—for the purposes of understanding the structure of Rousseau's views about institutional design.

It should be clear from my earlier discussion of the society of the general will, and the social bases of cognition and motivation, that Rousseau's theory of institutions needs a richer structure than we find in the stylized rational choice conception. Three differences are of particular importance.

First, motivations cannot be taken as fixed and given for the purposes of institutional design, but are rather shaped by institutions. Tastes, passions, desires, and mores all indicate the state of collective opinion, which in turn has institutional roots. This is importantly true of the general will itself. As Rousseau's genealogy of vice makes clear, the existence of a shared, supreme allegiance to the common good needs to be engendered, and cannot simply be assumed, or presented as a kind of natural and spontaneous outgrowth of basic human powers.

Second, Rousseau's norms are not external. Instead, social cooperation is regulated by a general will only if, as the Common Good and Priority Conditions indicate, the members take the fact that laws and institutions advance the common good as sufficient reason for complying with them. An allegiance to the norms figures in the motivations of the members of a just order. Indeed, in the rational choice conception itself, someone—say the mechanism designer, who establishes the scheme of incentives—must have the direct concern with the common good (with the social welfare function) that Rousseau thinks all citizens have in the society of the general will.

Finally, Rousseau's psychology is not reductionist. Rather (taking the Particular Interest and Priority Conditions together), the motivations of citizens have a structure in which reasons of the common good are supposed to take precedence over reasons of particular interest in public contexts: practical reasoning has a structure that is not captured by preference intensities. Representing this system of reasons by a single utility function would be at best unhelpful since it would obscure the different roles of reasons of particular interest and the common good in the deliberation of agents.

Taking these three points together, it is clear that Rousseauian problems of institutional design cannot be construed exclusively as problems of devising a scheme of incentives that will induce socially optimal consequences from individually rational actions.

But Rousseau knows better than to rely solely on the notion of an internalization of values or norms in describing a society that operates according to the general will of its members. Ethical motivations—in the form of a direct concern to act for reasons of the common good, founded on a commitment to treat other members as equals—play a large role in Rousseau's conception, and the norms that provide the content of those motivations are acquired through institutions. He is skeptical, however, about their sufficiency in resolving problems of institutional stability, because he is concerned about their strength in the face of pervasive temptations to vice.

Indeed, this concern about the strength of moral motivation in the face of temptations lies close to the heart of Rousseau's moral psychology. Thus in his *Confessions*, he refers to "this great maxim of morals, perhaps the only one of use in practice,

to avoid situations that put our duties in opposition with our interests, and which promise us good in the bad of others; certain that in such circumstances, however sincere our love of virtue may be, sooner or later one weakens, and becomes unjust and wicked in fact, without having ceased to be just and good in the soul" (C 56).¹² We met with this observation in our earlier discussion of *motivational complements*: the sense of duty often lacks sufficient strength to resist competing pulls from passions and interests. Because we are moved by both interests and a sense of duty, and our interests—with respect to which norms are external—may lead us to violate our duties—with respect to which norms are internal—we will need strands of both rational choice (a concern with incentive alignment) and ethical (a concern with acquiring values and principles) conceptions of institutional design.

Four Strategies of Argument

What is of particular interest, then, in Rousseau's view is the way that it draws on both strands of institutional theory in working out the conditions for a society regulated by the general will. Because members of a social association regulated by a general will have interests that sometimes conflict with the general will (as the Particular Interest Condition states)—because the norms are external to citizens' particular will of citizens, even if they are internal to the motivational system of members—the account of institutions will share certain features with the rational choice conception. Because institutions play a central role in forming a general will, the account shares certain features with the ethical conception.

More particularly, taking the account of the general will and the theory of motivation as a guide, we can discern at least four different sorts of strategies at work in Rousseau's arguments from the principle of the supremacy of the general will to institutions sketched earlier: (1) arguments of *principle*, which defend rights as intrinsic to a free community of equals; (2) considerations of *will-formation*, which aim to ensure that there is a shared commitment to the common good, in a society of equals; (3) ways of providing *assurance* and preventing *temptation*, which aim to

ensure that the commitment to the common good fixes the rules of social cooperation; and (4) *epistemic* arguments, which aim to provide assurance that, if people vote their views of the common good, citizens can be confident that the common good will result.

Principle

Some of Rousseau's claims are based on what I will call "arguments of principle." Generally speaking, arguments of principle are about the basic rights of members of a free community of equals, and draw institutional consequences from those rights. In my earlier discussion of rights and the general will (pp. 82–3), I said that a society with a general will is defined in part by a shared acknowledgment that certain fundamental interests of each member are to be respected. Construing rights as claims that ought to be acknowledged by others, then, the existence of a general will implies the existence of rights, for it implies a shared recognition of the requirement that those interests be protected. Fundamental rights are, so to speak, implicit in the ideal of the society of the general will, with its commitment to a non-aggregative conception of the common good, not claims against the general will: general willing is constitutively protective of these rights.

More specifically, then, citizens have at least the following basic rights:

1. Rights to the basic goods of security and personal liberty that are components of the common good: "The Citizens must have all reasonable securities that while doing their duty they will be able to sleep in their bed" (LM 280).¹³ We have, associated with self-love, a fundamental interest in these goods as requisites of our individual well-being, and we make the social compact to ensure security of person and goods. The existence of a general will, in turn, implies (see GW2) a public commitment to advancing the interests of each.¹⁴
2. Arguably, a right to a level of well-being sufficient to keep individuals from being forced to sell themselves: "with regard to equality, this word must not be understood to mean that degrees of power and wealth should be absolutely the same. . . . [A]s for wealth, no citizen [is to] be so very

rich that he can buy another, and none so poor that he is compelled to sell himself: Which assumes, on the part of the great, moderation in goods and influence and, on the part of the lowly, moderation in avarice and covetousness" (SC 2.11.2) Rousseau explains in an attached footnote that the society of the general will needs to "bring the extremes as close together as possible; tolerate neither very rich people nor beggars."

3. A right to personal independence, including liberties of conscience and worship, within the bounds of civil religion, which emerges from the limits intrinsic to the general will: that is, the need for a justification of regulations by reference to the common good, and a corresponding limit on the extent of such regulation: "freedom," Rousseau says, is one of the principal objects of legislation, "because any individual dependence is that much force taken away from the State" (SC 2.11.1). A person is dependent when he or she is controlled by another individual, and not bound solely by laws which apply to all, and are themselves controlled by judgments of the common good. And for reasons explored earlier, a person who is dependent solely on the law will have scope for individual independence.
4. A right to participate in politics. This right, underscored by Rousseau in his argument for the inalienability of popular sovereignty (SC 2.1), flows from the requirement (see GW4) of public confidence that institutions advance the common good. Thus it violates the terms of the social contract, and ultimately the fundamental interest in autonomy that the society of the general will secures, for citizens to be put in or to put themselves in a position in which they are unable to assess whether the rules and institutions by which they are governed conform to the common good, or unable to act to ensure that the terms of social association are consistent with those assessments:

The Sovereign may well say, I currently will what a given man wills or at least what he says he wills; but it cannot say: what this man is going to will tomorrow, I too shall will it; since it is absurd for the will to shackle itself for the future, and since no will can consent to anything contrary to the good of the being that wills. (SC 2.1.3)

Though the case is here stated in terms of the authority of the people as a collective body, the implications for the right of individual citizens to participate seem straightforward.

What, then, about direct democracy as a particular form of the right to participate and of the organization of consent? Does Rousseau's defense of it operate at the level of principle, say, by establishing an individual right to participate directly in lawmaking? Earlier, I mentioned Habermas's complaint about Rousseau's conflation of a principle of legitimacy with a procedure for organizing consent. If this complaint were well founded, we would expect Rousseau's argument for direct democracy to be an argument of principle. And Rousseau himself suggests that it is, when he asserts that representative government is tantamount to slavery. Because slavery is excluded on principle, as incompatible with our nature as free, moral agents, representation appears to be so excluded as well. Moreover, Rousseau tells us both that the general will cannot be alienated, and that the general will cannot be represented for the same reason that it cannot be alienated (SC 3.15.5), which locates the critique of representation at a level of basic principle.

Still, Rousseau's strongest case against representation is not, I think, best interpreted as an argument of principle, and the strongest case for the assembly form, correspondingly, does not make direct participation a matter of basic right. Although the criticisms of representation are extensions of the argument against slavery, the main concerns are not themselves matters of principle. How so?

Rousseau's condemnation of slavery follows directly from the basic logic of the master and slave relationship. Slavery is illegitimate because it is a defining feature of that relationship that the slave has no will of his or her own; a slave is understood to be an extension of the will of the master.¹⁵ But someone with no will of his or her own cannot be a bearer of obligations (SC 1.4.6). So the notion of a right of slavery—the idea that there are relations of right between master and slave—is incoherent: this is the force of Rousseau's comment that “the right to slavery is null, not only because it is illegitimate, but because it is absurd and meaningless. These words *slavery* and *right* are contradictory; they are mutually exclusive” (SC 1.4.13).

But the same cannot be said of a system of representation. Whatever the vices of such a system, it is simply not part of the defining conditions of representation that a person who is represented has no will of his or her own independent of the will of the representative. If it were, then Rousseau's own account of government would be in trouble, because the government does represent the sovereign (SC 3.15.8). To which it might be responded that government *can* represent the sovereign precisely because government is assumed (at least ideally) simply to *implement* or execute a sovereign will made determinate through legislation ("executive power . . . is nothing but force applied to Law"—SC 3.15.8)—a view of executive power as perfect agent for a legislative principal that Rousseau may have thought plausible because he was imagining a relatively small number of well-defined laws, not the legislative and regulatory profusion of the modern administrative state. In contrast, the sovereign will cannot be represented by lawmakers because legislative representation would assign to the legislators the responsibility for determining the content of the will, with the result that the people would have no will of its own.

This response, however, does not seem compelling, because it understates the complexity of representation. The represented may have views about political questions, and express those views in a variety of ways—including in their choice of representative and periodic reassessments of that choice, as well as in presenting views to elected legislators: the relationship of represented to representative is, in this respect, unlike the relationship of slave to master. Nor need the will of the represented who chooses the representative be a particular will. It is fully consistent with the logic of representation that judgments about who is the best representative express citizen's substantive views, held on reflection, about the representatives' views on the common good and how best to achieve it, and not simply preferences rooted in considerations of rational advantage.

It is not, then, intrinsic to a scheme of legislative representation that those who are represented have no general will: not intrinsic that they lack a concern for the common good, or that that concern fails to guide their political judgments. Of course, it is also true that it is not intrinsic to a system of representation that citizens have a general will: they may

be entirely self- or group-interested, or treat politics as having principally entertainment value. But then the same is true of a system of direct democracy, since in such a scheme it is possible that “the passions . . . not the reason, of the public would sit in judgment.”¹⁶ In such a system, the legislative assembly may be dominated by theatrical display, rhetorical flourish, and group-bargaining.

Of course, it might be argued that a representative scheme is most appealing when citizens are not themselves concerned about the common good. Moreover, such a scheme may be thought to ratify that lack of concern by further deflecting the attention of citizens from politics. By establishing as a principle of public order that some citizens are responsible for attending to matters of common concern, it may encourage others to neglect those common concerns or consider them solely from the standpoint of their particular interests (as a system in which a supreme court is assigned special responsibility for protecting basic rights may lead citizens generally—and legislators in particular—to be less vigilant about defending those rights). Rousseau makes (more or less) precisely this argument. He suggests that the “invention” of systems of deputies and representatives has its roots in the “cooling of the love of fatherland, the activity of private interest, the immensity of States, conquests, [and] the abuse of Government” (SC 3.15.4). And he adds that “[a]s soon as public service ceases to be the Citizens’ principal business, and they prefer to serve with their purse rather than with their person, the State is already close to ruin” (SC 3.15.1). But these contentions rest on political-psychological claims, not to say implausible claims, about the psychological consequences of different forms of political institution. Because they so depend, we do not have a case for direct democracy as a matter of fundamental principle, on a footing with the case against slavery or for popular sovereignty.

Taking these considerations together with Rousseau’s willingness to entertain systems other than direct democracies in his constitutional writings (see above pp. 136–7), I conclude that Rousseau’s case for direct democracy is not best understood as operating at the level of principle: it does not have the same basic structure as his case against slavery. Instead, his assessments of more participatory and more representative systems derive from broadly empirical assessments about the operation and evolution of such systems.

Textual considerations point strongly in the same direction. Thus consider the location (and purpose) of the criticisms of representative government in *The Social Contract*. The criticisms come only at the end of Book 3, on government, after Rousseau has presented his case for the general will's sovereignty, inalienability, and indivisibility; the expression of the general will through the law; the great legislator's role in the initial formation of a general will; and the distinction between legislation and government. Moreover, the most important discussion of representation arises out of concerns about the "death of the body politic" (SC 3.11) as a consequence of the inevitable disposition of government, which is essential to giving effect to the law, to undermine popular sovereignty and the general will's supreme authority: "Just as the particular will incessantly acts against the general will, so the Government makes a constant effort against Sovereignty. The greater this effort grows, the more adulterated does the constitution get, and since there is here no other corporate will to resist the will of the Prince and so to balance it, it must sooner or later come to pass that the Prince ends up oppressing the Sovereign and breaking the Social treaty. This is the inherent and inevitable vice which relentlessly tends to destroy the body politic from the moment of its birth" (SC 3.10.1). In *Letters Written from the Mountain*, he tells a more detailed and forceful story, historically (focused on Geneva) and conceptually, about executive usurpation. The passage is remarkable and reveals, even more than the corresponding remarks in the *Social Contract*, the concerns about the reversal of role between legislative principal and executive agent that animate Rousseau's views about civic vigilance: "What happens to all Governments like yours, Gentlemen, has happened to you. At first the Legislative power and the executive power that constitute sovereignty are not distinct. The Sovereign People wills by itself, and by itself it does what it wills. Soon the inconvenience of this cooperation of all in everything forces the Sovereign People to charge some of its members to execute its wills. These Officers, after having fulfilled their commission, account for it and return to the common equality. Little by little these commissions become more frequent, finally permanent. Insensibly a body forms that always acts. A body that always acts cannot account for each act: it no longer accounts for any but the principal ones; soon it reaches the point of accounting for

none of them. The more active the power that acts is, the more it enervates the power that wills. Yesterday's will is deemed to be today's also; whereas yesterday's action does not dispense from acting today. Finally the inaction of the power that wills subjects it to the power that executes; little by little the latter renders its actions independent, soon it wills; instead of acting for the power that wills, it acts upon it. Then there remains in the State only an acting power, that is the executive. The executive power is only force, and where force alone reigns the State is dissolved. There, Sir, is how all democratic States perish in the end" (LM 238–9).

Presented in response to these concerns about the government's appropriation of the sovereign people's legislative powers, and the attendant "death of the body politic" (SC 3.11), the account of citizen participation—and the unfavorable contrast of representation with direct democracy—is presented as a way to maintain popular sovereignty: Book 3, chapter 12 is called "Comment se maintient l'autorité souveraine," and 3.13 and 3.14 are both called "Suite." Direct lawmaking in an assembly is about preserving sovereignty, about ensuring its stability (SC 3.12–14), not as a defining condition in the conception of sovereignty itself. Why is it a good way? I will come back to this question later. For now, it is enough to distinguish the requirement of the general will's sovereignty and the rights that flow from it, from arguments about the specific forms of participation rights that might help to ensure the stability of the sovereignty of the general will.

Will-Formation

The second element in the theory of institutions concerns the formation of the general will. A social association is regulated by a general will only if it features a shared conception of the common good (Common Good Condition), and citizens take the fact that rules and policies advance the common good as a sufficient reason for supporting them (Priority Condition). Neither the existence of such a conception, nor its role in public deliberation can, however, simply be stipulated; we need to explain how a general will might be elicited. And doing so has two main aspects. We need an account both of the conditions that contribute to the initial formation of such a framework and of those that contribute to its maintenance once it exists.

Rousseau's account of the issue of initial formation relies on the great legislator who forms a people (SC 2.7–8). Here, however, I want to focus solely on questions about the stability of associations which are regulated by a general will, and not with the original formation of such an association. And so, in the discussion that follows I will abstract from the legislator, assume that there is a general will, and consider how that will might be maintained.

The existence of a general will requires that citizens have a capacity to recognize common interests and are motivated to advance those interests. We know from the genealogy of vice that neither the capacity nor the motivation are at all automatic, or part of a normal process of human maturation; we cannot simply assume that, whatever their conditions, people will develop a capacity to recognize the common good or an effective desire to act in pursuit of it. Motivations depend on how people think about themselves and others; but how people think about themselves and others depends on their social circumstances. The problem of sustaining the motivational requirements of a supreme general will needs, therefore, to be addressed institutionally.

We know what the general character of the solution must be: Because having a general will requires giving equal weight to the interests of each citizen, people will only develop a general will if their conditions are fundamentally equal, whatever the differences in the particular circumstances. In the earlier discussion of "motivational complements," I sketched the background conditions of the formation of a general will, indicating how reciprocity, generalized compassion, and the sense of self-worth might all come to support an allegiance to the general will, assuming appropriate institutional conditions—in particular, conditions in which we are regarded as equals and experience others as such. Consider, first, reciprocity:

What fosters the well-being of an individual attracts him; what harms him repels him. This is merely blind instinct. What transforms the instinct into sentiment, attachment into love, aversion into hate, is the intention manifested to harm us or to be useful to us. One is never passionate about insensible beings which merely follow the impulsion given to them. But those from whom one expects good or ill by their

inner disposition, by their will—those who we see acting freely for us or against us—inspire in us sentiments similar to those they manifest toward us. (E 213)

Assume this is right, and that others act with a manifest concern for our good. Then we will develop a concern for their well-being and, from that concern, a willingness to advance their interests as well. What we have here is not simply the behavioral reciprocity that Hobbes defends in his response to the fool, but the formation of a genuine concern for the welfare of another person in response to a recognition of their genuine concern.

Recall now that the society of the general will meets the Common Good Condition, which requires a general willingness to act in ways that advance common interests, and the Priority Condition, which requires giving priority in public decisions to the common good. The account of reciprocity provides the psychological foundations for an explanation of how citizens might come to be moved by reasons of the common good. Extended to include all members of society, reciprocity generates a concern for the good of all. That is, if each other member of a collection of persons displays a concern for my good, then—assuming reciprocity—I develop a concern for their welfare in addition to a concern for my own. But to have a concern for the welfare of each person *is* (abstracting from certain details that are irrelevant here) to have a general will. From the point of view of institutional design, then, the problem is to characterize circumstances in which the psychological mechanism of reciprocity would be, in this way, brought into play.

Turning next to the social aspects of self-love: in conditions in which people are publicly treated as equals, I come to see myself as an equal; if I do, then others affirm my worth when they treat me as an equal; there is no encouragement of the inflamed form of self-love and so no need for some to be treated by others as betters in order for those others to affirm their worth. But when there is a supreme general will each person is in fact treated by others as an equal. So the existence of a general will ordering the terms of social association not only ensures that the more particular interests of each citizen in security and liberty are taken into account in settling the laws, it also provides a way in which the worth of each is affirmed, even in the face of the inevitable differences in the circumstances of different citizens.

And this affirmation of worth supports the requirements of the general will.

Here—rather than at the level of principle—we find a case for Rousseau’s “participatory,” directly democratic forms of collective decision. Given the background assumptions already introduced, the reason is straightforward. Popular assemblies bring citizens together under manifestly equal conditions. In such assemblies “the person of the last Citizen is as sacred and inviolable as that of the first Magistrate, because where the Represented is, there no longer is a Representative” (SC 3.14.1). Operating in circumstances in which citizens have a general will, such popular assemblies bring together citizens who recognize that others take their interests into account. When public deliberation is guided by reasons of the common good, members do “see [others] acting freely for [them].” They reciprocate, and thus form similar sentiments. That is, they form a general will. The experience of being treated as an equal in public arenas, then, leads to the formation of a motivation that expresses this equality.

For public deliberation to support the formation of a general will, however, citizens must assume that their decisions settle the substance of policy, and that it is not, for example, determined by a faction in the population, or by the ability of the wealthy and powerful to control the terms of debate. So considerations about will-formation support popular assemblies, but only on the assumption that the restrictions on inequality, as well as conditions on property rights and associations, are in place as well.

Rousseau’s basic objection to a system of representation, I think, is that such a system does not have the same virtues from the standpoint of motivation formation. Since participation in a representative system is no longer a directly public activity among equals, the conception of citizens as equals and the motivations that are associated with that conception are less powerfully supported, and the considerations that figure in the choice of representatives are more likely to be private or particularistic. “As soon as public service ceases to be the Citizens’ principal business, and they prefer to serve with their purse rather than with their person, the State is already close to its ruin” (SC 3.15.1). Both representative and represented tend to become corrupted, meaning that they subordinate judgments of the common good to particular interests in their political deliberation.

But larger states are inconsistent with direct, assembly democracy of the sort entertained in the *Social Contract*. And in the constitutional writings after the *Social Contract*, Rousseau himself suggested other ways to engender and sustain a general will in states whose size provides an obstacle to legislation through popular assemblies. The essential point is to keep political discussion focused on the substance of proposals. Thus, in his constitutional proposals for Poland, Rousseau proposes to focus public debate on the substance of the common good through frequent meetings of the legislative body, high rates of turnover for representatives, and requiring “representatives to adhere exactly to their instructions, and to render a strict account of their conduct in the Diet to their constituents” (P 201). He contrasts this effort to maintain popular control of legislators, and to keep the focus on substance, with the “negligence, the carelessness, and I dare say the stupidity of the English Nation,” which, “after arming its deputies with the supreme power,” impose no restrictions on what their representatives do with that power “during the entire seven years of their mandate” (P 201).

Further discussion of alternatives to either the direct democratic or mandate system is beyond the scope of the discussion here—though, for example, a strong political-party system that casts political debate in terms of alternative ways of advancing the common good might help serve the integrative and educational functions that Rousseau attributes to the popular assemblies. Here my aim is to emphasize that we should not construe Rousseau as identifying popular sovereignty with a particular way of exercising that sovereignty, that the point of direct democracy is in part to provide at least a partial account of the acquisition of a general will, that the story about acquisition makes much of ensuring a focus on the substance of laws, and that once we have separated these issues we can entertain alternative institutional schemes that satisfy the more abstract requirements that Rousseau endorses, as essential elements in a free community of equals.

Temptation/Assurance

The third strategy of argument arises out of concerns about temptation and assurance. It is the most complicated, since it

directly embodies both rational choice and ethical elements, and so will require more detailed treatment.

Let's first return to the account of a society regulated by a general will. Rousseau supposes that the members of such a society have a general will and a particular will—that they order social states on the basis of (at least) two distinct kinds of consideration. They evaluate them from the point of view of both their own preferences and the common good: this is the force of the Particular Interest and Common Good Conditions. But given these multiple sources of ordering, why expect people to act on their general will—why will that be supreme? Merely to have a general will is not sufficient to ensure that we meet the Priority Condition. Citizens also have particular interests (Particular Interest Condition). So they can have a general will (which is “always constant, unalterable, and pure”), but not act on it because that will is “subordinated to others that prevail over it” (SC 4.1.6; PE 7–8). Recognizing the conflict between general will and particular interests, they subordinate the general will to their particular interests by, for example, withholding contributions to the provision of public goods, even though they recognize and take advantage of the benefits delivered by such goods, or by voting on the basis of their own welfare, or the good of their group, or by seeking to advance their interests by influencing the administrative implementation of the laws. Because the mere existence of a general will, understood as a shared conception of the common good, is insufficient to ensure satisfaction of Priority—to exclude an “inversion” in the order of practical-political reasons—Rousseau must consider ways to encourage people to act on their general will.

I call the problem of avoiding subordination the “temptation/assurance problem” to emphasize the two different perspectives from which the problem arises.

1. We need to avoid arrangements that tempt people to subordinate their general will to their private advantage.
2. We need to assure those who are not tempted that their willingness to cooperate will not be abused: “Either give me guarantees against all unjust undertakings or do not expect me to refrain from them in turn” (GM I.2.10).

Consider these two aspects in more detail.

1. According to the fundamental moral maxim stated in the *Confessions*, people are led into temptation by “situations that put our duties in opposition with our interests, and which promise us good in the bad of others” (C 56). We met with such social structures of temptation in Rousseau’s genealogy of vice: for example, private property and specialization destructively channel the sense of self-worth into inflamed self-love; and from inflamed *amour-propre* comes vice. Temptations can be resisted, and to have a general will is to have some resistance: it involves a willingness to act for the common good, and on obligations associated with common good, even when that is not most advantageous to oneself. But resistance to temptation has limits: that’s why motivational complements are important. And the larger the gain from pursuing one’s own interests (that is, the larger the loss from the point of view of one’s interests to acting on the general will), and the more pervasive the variation between those particular interests and common interests, the more likely are reversals in the proper ordering of reasons—with reasons of the common good subordinated to reasons of personal advantage (SC 4.1.6). According to Rousseau’s great moral maxim, under conditions of temptation, “however sincere our love of virtue may be, sooner or later one weakens, and becomes unjust and wicked in fact, without having ceased to be just and good in the soul” (C 56; SC 2.6.10).

2. Assurance problems arise because a person with a general will is willing to act on reasons of the common good, if *but only if* he/she has some reason to believe that others will also so act. The “independent man” who figures in Rousseau’s first draft of the *Social Contract* emphasizes this point: “Everything you tell me about the advantages of the social law would be fine if while I were scrupulously observing it toward others, I were sure that all of them would observe it toward me. But what assurance of this can you give me . . . ? Either give me guarantees against all unjust undertakings or do not expect me to refrain from them in turn” (GM 1.2.10; SC 1.7.6, 2.6.2). This request for assurance reflects a willingness to treat others justly, to treat them as equals—to act on a general will—provided that there is some guarantee that those others will not take a free ride.

The problems of temptation and assurance arise from conflicts between what right commands and what interest suggests. But they bear on different aspects of that concern. The problem of

temptation arises from a realistic appraisal about the strength of moral motivations in the face of temptations; the problem of assurance goes to the content of morality itself. It reflects the fact that justice commands reciprocity, not saintly self-sacrifice.

But while the problems of temptation and assurance bear different relations to the issue of right and interest, their resolutions are importantly connected. For each individual recognizes that the greater the advantage that others derive from injustice (i.e. from failing to comply with the general will), the greater their temptation to act unjustly. The fact that the system of incentives provides temptation for others undermines my assurance that they will act justly. And if I lack such assurance, then it is unreasonable in turn to expect me to act to promote the common good. By thus undermining assurance, a structure of temptations undermines the supremacy of the general will that is required for the compatibility of freedom and social interdependence. The problem, then, is to "combine what right permits with what interest prescribes, so that justice and utility may not be disjoined" (SC I.I).

This combination of right and interest, of justice and utility, is of course neither natural nor plausibly the product of some spontaneous social ordering. The problem of ensuring their coherence must be addressed institutionally. The structure of incentives induced by institutions needs to be designed to keep the incentives faced by individuals from conflicting too sharply with their views about what is right. The general form of the solution requires that social and political institutions structure choices and incentives such that the conduct that best advances particular interests, given the constraints on advancing those interests imposed by the institutions, is not pervasively and sharply at variance with the conduct recommended by judgments of what is right.

The key to achieving such a solution lies in the fact that judgments of what is right are themselves judgments about what advances common interests: in the society of the general will, the basic standard of political right is given by the conception of the common good. So the institutional problem is to ensure that individuals are in circumstances in which the best strategies for advancing their own particular interests lead them to choose policies that promote common interests: to achieve Bentham's

“artificial identification of interests.” The best way to do that is to structure institutions so that individuals and groups are not typically in a position to design and implement policies which can reasonably be expected to yield benefits for themselves alone and to impose the bulk of them on others. (Consider the difference between telling someone to divide a pie into six pieces and telling them that the piece they get depends on a throw of a die, and telling someone to divide a pie into six pieces, and letting them know in advance that they can pick their piece first.)

How is such design possible? The central feature of well-designed institutions is that they require that “everyone necessarily submits to the conditions which he imposes on others; an admirable agreement between interest and justice which confers on common deliberations a character of equity that is seen to vanish in the discussion of any particular affair, for want of a common interest that unites and identifies the rule of the judge with that of the party” (SC 2.4.7). Under such conditions, each must “think of himself as he votes for all” (SC 2.4.5). Required to live under the same terms that I impose on others, and attentive to my own interests, I am more likely than I would otherwise be to select policies that advance common interests and that, therefore, conform to my general will as well.

Several institutional arguments fit this strategy, including those in support of the rule of law, popular supervision of government, and limits on inequalities of resources and power. Arguments from temptation and assurance are also deployed in opposition to a significant popular role in the administration of the laws. In each case, the aim is to prevent individuals and groups from being in positions in which they are able to design or advance public policies with predictable benefits for themselves alone, or which are especially onerous for others.

Consider first the theory of government. The government’s proper function is to bring the general rules fixed by popular decision to bear on particular cases. So “the Prince’s dominant will [i.e. the will of whoever executes the law] is or should be nothing but the general will or the laws” (SC 3.1.19). Because executive decisions can have such direct bearing on the well-being of individuals, and because government has the power to enforce those decisions, those who exercise executive power will be tempted to subordinate their general will to a particular or

corporate will. They will always be greatly tempted to corruption: to abuse their power to enforce laws by exercising it for their own benefit.

One solution to this fundamental temptation problem is for the people as a group to exercise executive power. But that will not do. Apart from problems of numbers of citizens and other demands on their time, citizens are subject to the same temptations as officials, thus are imperfect agents of the sovereign. Moreover, and perhaps worse, encouraging such temptations might corrupt their legislative will as well: "It is not good that he who makes the laws execute them, nor that the body of the people turn its attention away from general considerations, to devote it to particular objects. Nothing is more dangerous than the influence of private interests on public affairs, and abuse of the laws by the Government is a lesser evil than the corruption of the Legislator, which is the inevitable consequence of particular considerations" (SC 3.4.2). Instead, the people should supervise the government's work, aiming to ensure its accountability. Not governing, thus not applying the law to individual cases where the costs and benefits are more clear, their own temptation to invert the proper order of wills, general and corporate, is reduced (SC 3.4.2). At the same time, by supervising the work of government, they increase the costs to the executive of subordinating the general will to a corporate or particular will, thus reducing its temptation to do so.

Rousseau's "continuous proportion" theory of the proper size of government can also be understood in terms of these concerns (SC 3.1.8–16). According to this theory, population growth requires stronger government, because of the increased demands on enforcing laws that come with more people; but as government grows in strength, the capacity of the people to control the government—to monitor its faithful execution of the laws—must correspondingly grow: "since the expansion of the State offers the trustees of the public authority more temptations and more means to misuse their power, it follows that the more force the Government has to have in order to contain the people, the more force does the Sovereign have to have in its turn in order to contain the Government" (SC 3.1.14). The essential point is that in working out the proper size of Government, we need to take account of incentive alignment problems, in particular to ensure that the government, as agent,

has incentives to act on the wishes of the legislative principal. We are not to assume that political officials, with implementation and enforcement powers, will internalize the norms associated with their role in government and aim faithfully to implement legitimate law.

Consider next the rule of law and the requirements of substantive equality of condition, as ways to address problems of temptation and assurance. The requirement that coercively enforced regulations take the form of laws can be interpreted in either of two ways. On the first interpretation, the requirement of legality is simply a relatively uninformative restatement of Rousseau's theory of law. As the earlier discussion of Hobbes and Rousseau indicated, it is essential to a regulation's *being a law* that it be an expression of the general will: roughly, that we can reasonably interpret the regulation as aimed at advancing the common good and as treating members of the political society as equals. Rousseau's test of legal validity is thus partially content-based rather than, as with the legal positivists, purely source-based. The intuitive idea is that a *legal* regulation is not an arbitrary imposition, but a rule supported by good reasons. To be supported by a good reason, a regulation must bear a plausible relationship to the common good of a society of equals. This feature of Rousseau's view may be masked by his claim that *anything* that the sovereign general will wills is a law. But the concern for the common good is internal to the general will. So this claim simply says that the only conditions on legal validity are those that are internal to the general will, thus associated with a concern for the common good.

On a second interpretation, the requirement of legality is not simply the requirement that regulations bear a reasonable relationship to the common good. Instead, legality serves as an institutional device for producing decisions that conform to the requirements of the common good. An essential element of the rule of law is that the exercise of coercive power is to be confined to the enforcement of *general* rules. The idea of generality is difficult to capture precisely. But on any rendering, there is a distinction between a system of law, with its requirement of generality, and a system of decrees in which coercive power can be applied to individuals without the backing provided by a general rule.

The requirement of a general rule arguably makes the exercise of power more predictable, which has benefits for personal

liberty because individuals are better able to plan their affairs in ways that avoid external interference. But two other consequences of generality play a larger role in Rousseau's account of the rule of law. First, a central evil of social life is the subjection of some to the will of others. The paradigm of subjection is slavery, where the slave is required (by expectation and force) to live by the commands of the master. The republican tradition of political thought focuses on this evil of personal subordination, and identifies liberty with an absence of subjection to the arbitrary will of others. Rousseau's own thought is in part republican, in its association of freedom with non-subordination.¹⁷ In a crisp and forceful statement of this conception, Rousseau tells us that "Liberty consists less in doing one's will than in not being subject to someone else's; it also consists in not subjecting someone's else's will to ours" (LM 260–1). Though this passage talks simply about subjection to a will and not subjection to an *arbitrary* will, I do not think anything is to be made of the distinction. Subjection to someone's will means having to do what the person commands, whatever the content of the command happens to be. Writing in a republican spirit, Rousseau emphasizes the role of law, with its condition of generality of regulation, as protection against subjection to the will of others: "there is no liberty without laws, nor where someone is above the laws. . . . A free people obeys, but it does not serve; it has leaders and not masters; it obeys the laws, but it obeys only the laws and it is from the force of the laws that it does not obey men" (LM 261, 237). "Arbitrary power," in contrast, is "the worst of all disorders" (LM 249). To be sure, "arbitrary power" may here simply mean "power without reason," that is, without appropriate justification, which takes us back to the first interpretation. But in the context—a discussion of legislative power—Rousseau pretty clearly is thinking of a more institutional idea, about the need to cast regulations in a legal form.

Second, and linking law to the common good, it is more difficult to advance one's own interests to the exclusion of the interests of others if one is required to advance those particular interests by choosing a system of general rules that applies to oneself as well as others. If, as Locke said, the rules are "not . . . varied in particular cases," if we have "one rule for rich and poor, for the favorite at court and the country man at plough," if, in particular, legislators "are themselves subject to the laws

they have made," then there is "a new and near tie upon them [legislators], to take care, that they make them [laws] for the public good."¹⁸

By itself, this requirement is too weak to address problems of temptation and assurance. Assuming considerable *de facto* inequality of condition, a requirement that coercive power be used only in support of general rules will not prevent groups from assessing with some confidence the differential bearing of rules and policies on their members. And wealthier and more powerful groups will still be able to design and therefore will be tempted to design rules and policies for their own advantage. In these circumstances the legal equality of citizens "is only apparent and illusory" (SC 1.9.8 n.), and the rule of law provides limited support for the presumption that legislative judgments advance common interests. The requirement of generality of rules needs, then, to be supplemented by requirements of equality of wealth and power. For if groups are relatively equal in power, and if resources are widely distributed, then it is difficult to design policies that predictably deliver particularized benefits and generalized costs. By thus reducing temptations and providing assurance, such arrangements permit public deliberation to proceed in terms of considerations of the common good.

Epistemic Considerations

The fourth strategy of argument addresses the epistemic aspect of the general will. Suppose that citizens do not just have preferences for policies, but that we also have beliefs or opinions about what advances the common good—say, beliefs that honestly differ about how best to promote the common good. Suppose also that people are prepared to act on those beliefs (SC 4.2.8), for example to vote their opinions about the common good, and not simply their preferences. Nevertheless, individuals can have erroneous beliefs about how to advance the common good, and decision rules for aggregating individual judgments can be badly designed. In either case, collective decisions can fail to achieve the shared aim of promoting the common good. "One always wants one's good, but one does not always see it: one can never corrupt the people, but one can often cause it to be mistaken, and only when it is, does it appear to want what is bad" (SC 2.3.1).

The institutional issue is to increase the likelihood that the common good will be advanced if collective decisions are made by individuals who express their different opinions about what will promote common interests. An answer to this problem of sound collective judgment needs to have two components. First, we need to improve the quality (reliability) of individual judgments about what advances the common good. And second, we need to have decision rules such that collective judgments are more likely to be right than individual judgments.

Consider first the problem of decision rules. As I indicated earlier, Rousseau endorses the following view of collective choice: When there is a shared conception of the common good, and public deliberation is guided by that conception, and citizens vote their opinions about the common good rather than their preferences over policies, then the decisions of majorities about which policies to pursue can provide good evidence about which policies are in fact best. I suggested as well that the Condorcet jury theorems speak in support of this view, providing some basis for confidence in majority decisions—on condition, of course, that individuals are good judges. But one cannot simply assume that judgmental competences are fixed and high. Instead, attention must be given to the determinants of the judgmental competence of voters, and in particular, given my concerns here, to the institutional determinants of that competence.

In addressing those institutional determinants, we should notice first that the issue of competence in part overlaps with the temptation/assurance problem. To see that overlap, assume first that each individual is a reasonably good judge of what advances his/her own interests. Assume also that we have solved the temptation/assurance problem. If we have, then we can regard judgments about what advances our own particular interests as reasonable proxies for judgments about what will advance the common good (SC 2.3.3). Therefore individual judgmental competence can be increased by resolving the temptation/assurance problem. Under these circumstances, votes still reflect judgments about whether a proposal “does or does not conform to the general will, which is theirs” (SC 4.2.8). But, on the assumption that the temptation/assurance problem is solved, citizens can treat their judgment about the likely impact on their own situation as evidence for the broader impact.

In addition to its overlap with temptation/assurance arguments, this epistemic strategy of argument obviously supports efforts to ensure that citizens are well informed; the importance attached to political participation, since participation can be a source of information; and the rights of citizens to assemble, since such assembly may serve to enhance individual competence. There are additional features of the institutional view that are supported by epistemic concerns. But their force will be better appreciated if I postpone discussion of them until the end of the next section.

Popular Democracy or Executive Dominance?

Rousseauian institutions, then, help to establish a shared framework for collective judgments and decisions, encourage the stability of that framework over time, and ensure its regulative role in the life of the society of the general will. They give institutional life to the ideal of a free community of equals, in part by fostering allegiances to that community.

But an objection looms. I have not told the full story of Rousseau's conception of political institutions, and, according to the objection, I have abstracted from certain apparently undemocratic elements of Rousseau's political conception, thereby obscuring a central lesson of Rousseau's work—that, as Plato suggested in his analysis of the transition from democracy to tyranny, a populist conception of politics bears an intimate relationship to political authoritarianism.¹⁹

I want now, then, to consider certain elements in Rousseau's view that might be thought undemocratic on any familiar construal of this notion and consider the hypothesis that those elements reflect deep Rousseauian commitments. In particular, I will consider the view that Rousseau's commitment to democratic principles and politics is not as strong as it appears on the surface of the *Social Contract*. More specifically, I want to consider the contention that the *Social Contract* can mislead both about the substance of Rousseau's commitment to democracy and about the principles that underlie that commitment.²⁰

Undemocratic Aspects

Consider a few features in Rousseau's view that prompt concerns about his democratic credentials.

First, Rousseau's abstract claims about the ultimate locus of legitimate authority may mislead us about his views concerning the proper *de facto* distribution of political power. Formally, Rousseau divides political authority between a sovereign legislative people comprising citizens who directly enact laws in a legislative assembly, and a subordinate government, in which the people vest the authority to execute the laws, applying them to cases. But this formal structure arguably stands in tension with Rousseau's description of the real powers of legislative assembly and government. Putting the formal-institutional niceties to the side, Rousseau in fact defends a regime in which political initiative is concentrated in the hands of a largely self-sustaining political elite that not only administers the state, but also sets the political agenda and monopolizes the legislative process in all but its final stage.

In his illuminating account of Rousseau's critique of representation, Richard Fralin argues that Rousseau: (1) imposes considerable restrictions on popular selection of officials, in effect permitting the government to choose its own successor through its control over nominations; (2) rejects popular legislative initiative, placing all such initiative instead in the hands of the government; and (3) bars popular deliberation about the legislation introduced by the government.²¹ Formally an agent of a popular will, the government monopolizes political initiative, and thus plays a central role in shaping the content of the popular will that it subsequently executes. The formally sovereign people, by contrast, are deprived of the resources and political capabilities required for arriving at an independent judgment about the requirements of the common good and the laws and policies that will advance that good.

Second, the limits Rousseau imposes on the power of the assemblies are arguably not mere matters of institutional detail lying at the periphery of his central commitments and concerns. Instead the details reveal in the small his deep ambivalence about the political capacities of ordinary citizens while

also illuminating his underlying conception of the nature of democratic politics and the point of popular participation. As to popular capacities, the people are capable of helping to check the temptations of the government to promote its own corporate will at the expense of the general will. To play this role, citizens must be brought to identify with the community. Paralleling the role that he assigns to festivals in his constitutional proposals for Poland (see P 186–7), Rousseau assigns to citizens assemblies the task of encouraging that identification. Their function, to use the terms set out above, is will-formation. In this educative role, however, political participation is basically “passive,” and not an instrument of popular control of policy. Judith Shklar forcefully summarizes the main idea: “The fixed periodic assemblies of the people, at which they express their sovereignty, are primarily preventive in purpose. Their chief political aim is to halt the all but irresistible tendency of any government to become arbitrary and despotic. Their positive function is symbolic and ritualistic. They actually *do* very little. . . . The very occasion of consenting, the assembly, is a device for keeping their country before their eyes, and their public selves intact. Like the endless ceremonies and festivals that Rousseau urged upon the Poles, the assemblies exist to remind men of their public role.”²²

So—the objection continues—even the most participatory forms of democratic politics do not provide, nor does Rousseau intend them to provide, a public arena in which citizens can debate, judge, and advance their conception of the public good. Instead, the point of popular participation and democratic arenas is to provide an arena in which citizens are instructed in and can internalize through such instruction the understanding of the public good that is formulated by the government and embodied in existing laws.

These objections raise two broad questions:

1. What force is there to the claims about elections, deliberation, and initiative?
2. Insofar as those claims have force, what do they show about the aims of Rousseau’s theory of democracy in particular and his views about political and social institutions generally?

Let’s consider these two questions in turn.

More on Institutions

According to the critic, Rousseau defends an executive monopoly on political initiative. Closer inspection shows that this criticism is not accurate. To appreciate the trouble, let's consider the aforementioned contentions about his views on the conduct of elections, popular deliberation, and popular legislative initiative.

Elections. The election of officials is an exercise of executive or governmental power (SC 3.17.3). This follows from the fact that elections are, in effect, an application of the law. They apply the constitutional law stipulating the form of government to individual cases in that they involve the selection of named individuals for office (SC 3.17.2–3). But asserting that elections are, from a formal-constitutional point of view, an exercise of executive power does not settle the question of what body has the authority to exercise this specific aspect of executive power. In particular, it does not imply that the “magistrate” has the right to choose its own successor.

Part of the difficulty in addressing this question lies in Rousseau's uses of the term “government.” Rousseau uses the term “government” both normatively and institutionally. Used normatively, it refers to the legitimate exercise of executive power. Used institutionally, it refers to a particular body—the “intermediate body” that stands between subjects and sovereign, and that, like all such bodies, has sensibilities, a will, and rights of its own (SC 3.1.5–7, 20). When Rousseau says that the choice of officials is a “function of Government” (SC 3.17.3), it is clear that he is using the term “Government” normatively, indicating that, for the reasons I just mentioned, the choice of officials should be classified as an exercise of executive power.

But this point about classification has no implications at all about whether government in the institutional sense—properly speaking, the magistrate—has the right to exercise this aspect of governmental power. Even in a system with an intermediate governmental body separate from the people (e.g. an aristocratic government), some aspects of governmental power may be reserved by the people. If they are, then the people do not exercise these powers in their capacity as sovereign, but instead as bearers

of certain (reserved) executive powers. In short, Rousseau's identification of elections as a governmental function does not imply that the magistrate has the right to choose its own successor. Neither, of course, does it imply that the magistrate lacks that right. But other aspects of Rousseau's view do carry this implication.

In fact, Rousseau's view about the exercise of electoral power is clear: the power to choose officials must be exercised either by the assembled citizens or their mandated representatives. And this is a permanent feature of any legitimate state, and not simply a requirement on the initial creation of a government. Thus Rousseau holds that meetings of the periodic assemblies "ought always to open with two motions which it should be impossible ever to omit, and which are to be voted on separately" (SC 3.18.6). The first concerns whether or not to change the form of government, the second whether or not to leave the government in the hands of the current officials (SC 3.18.7–8). Because the choice of officials is an act of government, and the exercise of this governmental power must be retained by the popular assembly, every legitimate government is, at least to this extent, democratic.

Still, it might be said that these remarks on the right to elect officials are excessively attentive to issues of form, and insufficiently attentive to the real distribution of power. Even if the people have the right to elect officials, elections may nevertheless consist of choices from a menu of candidates selected by current officials: the executive dominates the choices of officials. But Rousseau excludes this possibility. Such a system of government-managed electoral choices is, he says, "a vain formality without solidity" (LM 249; also 246–7). And in describing the actual "servitude" of the Genevan people, he offers the Genevan system of elections as an example, accusing the Genevans of making "a great show" of a "choice of little importance" (LM 237–8).²³

Popular Deliberation. Does Rousseau reject popular deliberation about legislation introduced by the executive? Here, again, the balance of evidence suggests otherwise.

The evidence of opposition to citizen deliberation comes from an ambiguous passage in the *Social Contract*, where Rousseau emphasizes the efforts made by government to monopolize political discussion. Referring to "the right of voicing opinions, proposing, dividing, discussing [motions]," Rousseau observes

that "the Government always takes great care to allow [the right] only to its own members" (SC 4.1.7). But as I said earlier (pp. 138–9), Rousseau is not *endorsing* this practice of governmental monopoly. Moreover, in the passage in question, he links the right of voicing opinions to "the simple right to vote in every act of sovereignty; a right of which nothing can deprive Citizens," which points more or less unambiguously in the opposite direction.

Furthermore, he elsewhere contrasts the "actual state" of servitude of the Genevan people with their "legitimate state" of freedom and sovereignty (LM 237). In part, the basis of the distinction is that, in their own sovereign assembly, the general Council, the people can only act "when it pleases your Magistrates" and cannot "speak except when they interrogate you" (LM 238). Rousseau objects to these practices. He defends the right of legislative assemblies to determine their own agenda, and to come to their own views about public issues (LM 238, 249–51). It is "contrary to all reason," Rousseau argues, "that the executive body rule the public order of the Legislative body, that it prescribe to it the matters it must take cognizance of, that it forbid it the right of giving an opinion" (LM 251). For this makes the magistrate "master of its Sovereign" (LM 251). Contrasting the earlier Genevan assemblies in which everyone could speak with the situation in the 1760s, Rousseau allows that there was less "public order and decency" in the earlier period. Nevertheless, "the people was free, the Magistrate respected, and the Council was frequently assembled" (LM 251 n.). While acknowledging that a "numerous body needs supervision and order," Rousseau urges that Genevans "not let this supervision and this order overturn the goal of its institution." And he wonders why somber Geneva cannot establish "rule without servitude," whereas Athens, with its much larger assembly of "quick-tempered, ardent, and almost unrestrained Citizens," succeeding in being "Capital of the world" (LM 251–2).

Nor, as we saw in the earlier remarks about private societies, does Rousseau restrict political discussion to legislative assemblies. For example, he defends the Genevan citizenry's right of "deliberating within itself" outside of the legislative assembly (LM 269). Even in the "most absolute Governments assemblies of communities that are not presided over by any Magistrate are permitted" (LM 270). And his description of the Genevan *circles*

in his *Letter to d'Alembert* indicates as well the importance of political discussion outside of the legislative assembly in helping to make "friends, citizens, and soldiers" (LD 105). In praise of the Genevan system of clubs (in which men gather separately from women), he says that in the *circles* men can "devote themselves to grave and serious discourse without fear of ridicule," that they "dare to speak of country and virtue without passing for windbags," and that even if "the turn of conversation becomes less polished" nevertheless, "reasons take on more weight" and "the mind gains precision and vigor" (LD 105). For these reasons among others, the *circles* provide an example of a "practice" that is "linked to the form of government" and that contributes to the preservation of that form (LD 98).

Legislative Initiative. Consider next the governmental monopoly on legislative initiative. Rousseau endorses Genevan practice, and holds that proposals for new laws must come from the executive, from, as he puts it, "the inferior councils" (LM 276–7; also 264–5): "the Law has very wisely provided that the establishment and even the proposing of such [legislative] innovations would not pass without the approval of the Councils" (LM 264). This restriction is important, and in need of explanation. But before proposing such an explanation, I need to enter some cautionary remarks about making too much of the restriction.

In the first place, while Rousseau endorses the limitations on popular legislative initiative in Geneva, he does not say that a broader popular right to initiate legislation would be illegitimate, but rather that it would be unwise, and this for two reasons. First, in smaller states, innovations are typically disruptive, presumably because frequent changes of law deprive the existing laws "of that veneration that time bestows on everything",²⁴ thus decreasing the likelihood that they will be obeyed. And second, permitting a popular legislative initiative, as against an exclusively governmental right, would increase the likelihood of political innovation. To be sure, individual innovations taken separately might be beneficial. But it is reasonable to adopt a general policy of avoiding innovations, and to build that policy into the structure of the institutions. "Aversion for innovations is thus generally well founded" (LM 264), and the (negative) right of the Genevan magistrates to prevent legislative innovation is a reasonable instrument. But while it is generally well founded,

it is not for that reason illegitimate to permit a right of popular legislative initiative.

Furthermore, it is important to be clear about the precise restriction on legislative initiative. Such restriction should not be understood as barring citizens from proposing new laws to the government, or from expressing their views about such laws. Citizens have a duty to express their opinions on these matters (LM 264). Rousseau's claim is that there is wisdom in legislative institutions in which government must endorse legislation before final approval by a popular legislative assembly (LM 264–5).

Finally, Rousseau's restriction on specifically legislative initiative does not reflect a general opposition to popular political initiative. On the contrary, he argues passionately and at length for a popular right of remonstrance—that is, rights of the people, as supreme authority, to pursue inquiries into the government's execution and interpretation of the laws without prior initiation or approval by the government. The second part of *Letters Written from the Mountain* is devoted to distinguishing the “negative right” of the government to block legislative innovation, which is reasonable, from the alleged right of the executive to block popular review of the government's interpretation and implementation of the law, which Rousseau condemns. “The Legislator, existing always, sees the effect or the abuse of its laws: it sees whether they are followed or transgressed, interpreted in good or in bad faith; it watches over it; it ought to watch over it; that is its right, its duty, even its sworn oath. It is this duty it fulfills in Remonstrances, it is this right, then, that it exercises; and it would be against all reason, it would even be indecent, to wish to extend the negative right of the [small] Council to that object” (LM 265). And Rousseau proposes a variety of methods through which the people might ensure that its remonstrances are heard.

The conditions that surround the restriction on legislative initiative go some way to blunting the force of that restriction. Still, it is not trivial, particularly in a view that emphasizes the sovereignty of the people and the expression of that sovereignty through the direct, popular exercise of legislative power. And this brings me to the second issue: What, if anything, does the restriction on legislative initiative tell us about Rousseau's conception of democratic order? Does it reveal that Rousseau

aims effectively to eliminate popular political initiative and, more broadly, that he construes popular participation principally as a form of socialization and public celebration?

Democracy at the Right Speed

I have already provided the basis for an answer to this question by emphasizing the *system* of institutional conditions in which political participation is set, and the broader framework of argument that figures in Rousseau's theory of institutions. The restrictions on initiative are only one component of a political-institutional scheme in which the people have final legislative authority, rights to oversee the administration and interpretation of the laws, the power to decide if they want new officials and if they want to alter the form of government, and in which the people control the government's revenues. These powers are not the equivalent of a right of legislative initiative. They do, however, enable the people to impose costs on a government that consistently resists proposing legislation that follows popular concerns. Since the government does not control the electoral process or the conduct of political debate, and the performance of the government is subject to regular review by the sovereign people, the failure to propose legislation for which there is broad support increases the likelihood that it will be turned out. In short, the restriction on initiative is highly qualified.

But these qualifications may serve simply to underscore the question: Why propose the restrictions at all?

The principles of institutional design I outlined earlier suggest an answer. Perhaps the restriction on initiative can be defended by reference to epistemic considerations, as *a* way (of course, there are other ways) to encourage considered popular judgment about whether a new legislative proposal is likely to advance the common good. Consider, in particular, states in which citizens do not devote all or most of their time to politics: what Rousseau knew to be true about Geneva (LM 292–3) and thought would be true of any society without slaves (SC 3.15.9–10). Speaking to Genevans, he says the “situation demands maxims peculiar to you. Not being as idle as the ancients Peoples were, you cannot ceaselessly occupy yourselves with the Government as they did: but by that very fact that you can less constantly keep watch over it, it should be instituted in such a way that it might be easier

for you to see its intrigues and provide for abuses" (LM 293). In particular, some sort of initial review of the merits of legislative proposals is needed. Rousseau supposes that the review should be carried out by an executive, subject to the political constraints I have sketched: for example, if the government does not propose legislation that is widely desired, then the people can get rid of it. This power is, in effect, a roundabout, potentially costly, and clumsy way to initiate legislative changes.

But this slower method might bring gains in deliberativeness. It could be expected to encourage reflection on the reasonableness and importance of the changes, reflection that could reasonably be expected to improve popular judgments about legislative changes, and thereby help to sort good proposals from bad. By thus promoting more sustained reflection on the reasonableness of changes, the obstacles to initiative would discourage people from legislating their passing concerns, and improve popular judgmental competence. Understood this way, the restriction on initiative is Rousseau's response (or part of his response) to a classical criticism of democratic politics: that democracy is the rule of disorderly and fluctuating popular passions.²⁵ Rousseau, if I have him right, aims to meet this criticism, but without sacrificing the ideals of popular self-government that animate his theory.

A Democrat After All?

Taking these considerations together, I conclude that Rousseau's constraints on legislative initiative are not best understood as betraying a hidden desire to promote a monopoly on independent political action by a self-sustaining political elite. Nor should we interpret those restrictions as evidence that Rousseau thinks of political participation as a form of passive moral education, a kind of public festival and performance aimed at securing the internalization, celebration, and legitimacy of the public values articulated by a political elite.

To be sure, Rousseau recognized that popular judgments can be unsound—as unsound as the judgments of political elites and of the magistrates who predictably aim at political aggrandizement. But he also valued broad political participation. He

trusted the powers of ordinary citizens to assess the virtues of alternative conditions of association. And he sought to find institutions that would enhance popular judgments and thus help preserve, against the inevitable countervailing pressures, a free community of equals: a fragile achievement, he thought, but an important one and not beyond our reach. What we learn from reflection on Rousseau's political philosophy is not a hidden, dark secret of democratic thought: that a self-proclaimed commitment to popular self-government masks enthusiasm for a directive authority with responsibility for tutoring the people in the exercise of a rhetorically exalted collective authority. Instead, we see a profound formulation of the ideal of political self-government, rooted in a conception of moral freedom, and a reasonably realistic appreciation of the complex and demanding conditions of institution and conviction required to make that compelling ideal real.

Notes

INTRODUCTION

1. My notes from Rawls's lectures suffered water damage at some point, and I had to throw them away. The presentation of Rousseau in the published version of Rawls's lectures on the history of political philosophy is substantially different, and comes from lectures prepared 15–20 years later. See John Rawls, *Lectures on the History of Political Philosophy*, ed. Samuel Freeman (Cambridge, Mass.: Harvard University Press, 2007), 191–250.
2. For Rousseau's impact on Kant, see Ernst Cassirer, *Kant's Life and Thought* (New Haven: Yale University Press, 1981), 89, and in general, 84–90.
3. Georg Wilhelm Friedrich Hegel, *Lectures on the History of Philosophy*, vol. 3 (London: Routledge, 1892), 402.
4. I am grateful for the wonderful advice (editorial and substantive) that Charles Beitz provided in his capacity as the journal's book review editor.
5. On Rousseau's views about music, including the important debate with Rameau on harmony and melody, and on the relative merits of French and Italian music, see the illuminating introduction to volume 7 of *The Collected Writings of Rousseau*, trans. and ed. John T. Scott (Hanover, NH and London: University Press of New England, 1998).
6. The details and the distinctive features of his sexism have been a subject of important studies by Susan Moller Okin, *Women and Political Thought* (Princeton: Princeton University Press, 1979), pt. 3; Joel Schwartz, *The Sexual Politics of Jean-Jacques Rousseau* (Chicago: University of Chicago Press, 1984); and Anne L. Harper, "The Family and the State in Rousseau's 'Emile, Or On Education,'" unpublished dissertation, University of Michigan, 1987.
7. Frederick Neuhouser, *Rousseau's Theodicy of Self-Love: Evil, Rationality, and the Drive for Recognition* (New York: Oxford University Press, 2008).
8. Leo Damrosch, *Jean-Jacques Rousseau: Restless Genius* (Boston and New York: Houghton-Mifflin, 2005), p. xi.
9. Rawls, *Lectures on the History of Political Philosophy*, 193–4.
10. The general line of interpretation here derives from the account of Rousseau in Kant and Cassirer. See esp. Immanuel Kant, "Speculative Beginning of Human History," in *Perpetual Peace and Other Essays*, trans. Ted Humphrey (Indianapolis: Hackett, 1983), esp. pp. 53–5; Ernst Cassirer, *Question of Jean-Jacques Rousseau*, trans. Peter Gay (Indianapolis: Indiana University Press, 1963).

CHAPTER 1

1. According to *Emile*, “one is more free under the social pact than in the state of nature” (E 461).
2. Thomas Hobbes, *Leviathan*, ed. Richard Tuck (Cambridge: Cambridge University Press, 1991), 120, hereafter cited as Lev. Here and throughout, my discussion of Hobbes, and of contrasts between Hobbes and Rousseau, draws on my “Protection for Obedience: Hobbes’s Theory of the State” (unpublished).
3. See e.g. P 239–48, which discusses new forms of social mobility and gradual political emancipation in Poland, with peasants “becom[ing] free men and Citizens” (P 246). And see as well the discussion of the right of remonstrance and the importance of reestablishing periodic meetings of the general Council in Geneva, in LM 270–81.
4. To say that the conception of the common good is shared is, of course, not to say that citizens agree that existing laws are the best ways to advance that conception. See SC 4.2 on why citizens who comply with laws they disagree with are nevertheless obeying themselves.
5. To clarify: the community is free not in virtue of being independent—free from domination by another community—but in virtue of the members being fully autonomous.
6. According to Judith Shklar, Rousseau had a utopian’s indifference to charges about a lack of realism. See *Men and Citizens: A Study in Rousseau’s Social Theory*, 2nd edn. (Cambridge: Cambridge University Press, 1985), esp. chap. 1. I disagree, and am moved by, among other things, the detail in his constitutional writings, especially in his *Letters Written From the Mountain*. Moreover, while I would not wish to hang the large interpretive issue on a single passage, I am struck by that in *Rousseau, Judge of Jean-Jacques*, Rousseau says of Jean-Jacques: “Deluded by the ridiculous hope of making reason and truth triumph at last over prejudices and lies, and of making men wise by showing them their true interest, his heart—excited by the idea of the future happiness of the human race and by the honor of contributing to it—dictated to him a language worthy of such a great undertaking” (R) 131, emphases added). I also do not know of any clear affirmation of the utopian indifference to realism that Shklar attributes to him.
7. See generally Lev., chaps. 10, 11, 13, 17. On the passions as sources of irrational conduct, see Hobbes’s distinction between “multiplying glasses” and “prospective glasses” at Lev., p. 129, and *De Homine*, chap. 12, sec. 1, in *Man and Citizen: Thomas Hobbes’s “De Homine” and “De Cive”*, ed. Bernard Gert (New York: Anchor Books, 1972).
8. Thus Hobbes’s emphasis on fear of punishment, which serves to “tame pride”—to “bridle men’s ambition, avarice, anger, and other passions” (Lev. 96)—and not principally to concentrate the mind or to elicit an “extraordinary use of reason” (Lev. 206), or in some other way to psychically transform natural individuals into self-disciplined subjects who are able to control their “perpetually incumbent and pressing” passions (Lev. 206).

9. See Hobbes's derivation of laws of nature 8–12, Lev. 107–8. The ninth law, proscribing prideful conduct, is especially important.
10. Juergen Habermas, *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*, trans. William Rehg (Cambridge, Mass.: MIT Press, 1996), 7.
11. See Joshua Cohen, *Philosophy, Politics, Democracy: Selected Essays* (Cambridge, Mass.: Harvard University Press, 2009).

CHAPTER 2

1. John Locke, *Second Treatise*, para. 17, in *Two Treatises of Government*, ed. Peter Laslett (Cambridge: Cambridge University Press, 1988).
2. *Ibid.*, para. 18.
3. Aristotle, *Politics*, trans. Benjamin Jowett, bk. 2, chaps. 2–5, in *The Complete Works of Aristotle*, ed. Jonathan Barnes, vol. 2 (Princeton: Princeton University Press, 1984).
4. Aristotle, *Politics* 1261a18.
5. *The Republic*, ed. G. R. F. Ferrari, trans. Tom Griffith (Cambridge: Cambridge University Press, 2000), 462A/B; also 423D.
6. *Ibid.*, 462B, 464A.
7. Allan Bloom, "Rousseau's Critique of Liberal Constitutionalism," in Clifford Orwin and Nathan Tarcov (eds.), *The Legacy of Rousseau* (Chicago: University of Chicago Press, 1997), 157.
8. *The Question of Jean-Jacques Rousseau*, 52.
9. Emile Durkheim, *Montesquieu and Rousseau* (Ann Arbor: University of Michigan Press, 1960), 98.
10. John Rawls, *A Theory of Justice*, rev. edn. (Cambridge, Mass.: Harvard University Press, 1999), 109–12.
11. *Ibid.*, 281.
12. The theological background of the notion of a general will in the idea of universal grace underscores the need for a non-utilitarian, aggregative interpretation of the common good. For a highly illuminating account of that background, see Patrick Riley, *The General Will Before Rousseau: The Transformation of the Divine into the Civic* (Princeton: Princeton University Press, 1986).
13. A number of commentators correctly emphasize that Rousseau's talk about the common good should be explicated in terms of the interests of the individual members of an association. See Brian Barry, "The Public Interest," *Proceedings of the Aristotelian Society* 38 (1964), 9–14; W. G. Runciman and Amartya Sen, "Games, Justice, and the General Will," *Mind* 74 (1965), 554–62; Hilail Gildin, *Rousseau's Social Contract* (Chicago: University of Chicago Press, 1983), 54–6.
14. See Joshua Cohen, "Structure, Choice, and Legitimacy: Locke's Theory of the State," *Philosophy and Public Affairs* 15/4 (Autumn 1986), 301–24.

15. See David Gauthier, *Morals By Agreement* (Oxford: Oxford University Press, 1986).

CHAPTER 3

1. See generally Lev., chaps. 16, 17, 26.
2. Habermas, *Between Facts and Norms*, chap. 3; "On the Internal Relation Between Law and Democracy," in *The Inclusion of the Other: Studies in Political Theory* (Cambridge, Mass.: MIT Press, 1998), chap. 10.
3. A more precise statement of this Generalized Condorcet Theorem and a summary of related jury theorems can be found in Bernard Grofman, Guillermo Owen, and Scott Feld, "Thirteen Theorems in Search of Truth," *Theory and Decision* 15 (1983), 261–78. On the relevance of the Condorcet theorems to a Rousseauian view, see Brian Barry, "The Public Interest," 9–14; Bernard Grofman and Scott L. Feld, "Rousseau's General Will: A Condorcetian Perspective," *American Political Science Review* 82 (June 1988), 567–76; and the contributions by Jeremy Waldron, David Estlund, Bernard Grofman, and Scott Feld to "Democratic Theory and the Public Interest: Condorcet and Rousseau Revisited," *American Political Science Review* (Dec. 1989).
4. See Riley, *Will and Political Legitimacy*, chap. 4. Riley claims that there is a conflict between Rousseau's "morality of the common good" and his endorsement of the voluntarist conception of legitimacy associated with the social contract tradition.
5. The idea of a self-effacing theory comes from Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), secs. 9, 17.
6. For a more complete discussion, see the account of the "Natural Bases of Conflict," in my "Protection for Obedience" (unpublished).
7. See *Critique of Practical Reason*, in Immanuel Kant, *Practical Philosophy*, trans. and ed. Mary J. Gregor (Cambridge: Cambridge University Press, 1996), 162–4.
8. For illuminating discussion, see Christine M. Korsgaard, *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), chap. 6 ("Morality as Freedom").
9. I am very grateful to Andrew Williams for raising the concern that I address in the next few paragraphs.

CHAPTER 4

1. This chapter is based on Joshua Cohen, "The Natural Goodness of Humanity," in Andrews Reath, Barbara Herman, Christine Korsgaard (eds.), *Reclaiming the History of Ethics: Essays for John Rawls* (Cambridge: Cambridge University Press, 1997).

2. For criticism, see Noam Chomsky, *Reflections on Language* (New York: Pantheon, 1975), 128–34.
3. Cassirer, *Kant's Life and Thought*, 89.
4. See his *Jean-Jacques Rousseau: Transparency and Obstruction*, trans. Arthur Goldhammer (Chicago: University of Chicago Press, 1988), 20.
5. *Cartesian Linguistics* (New York: Harper and Row, 1966), 59–72.
6. Kant's account of evil has broadly Rousseauian contours, though Kant differs from Rousseau about the possibility of *explaining* moral evil. See *Religion Within the Bounds of Mere Reason*, in Immanuel Kant, *Religion and Rational Theology*, trans. and ed. Allen W. Wood (Cambridge: Cambridge University Press, 1996). According to Kant, we are naturally endowed with "predispositions to humanity," a form of self-love that leads us to make "comparison with others" and from which "originates the inclination to gain worth in the opinion of others" (p. 75). Though focused "originally" on "equal worth," this inclination is corrupted by social cooperation (pp. 129–30), which changes it from a concern to be on an equal footing to "an unjustified craving to win it [superiority] for oneself over others" (p. 75). And this desire for advantage leads in turn to the "vices of culture"—for example, envy, greed, spitefulness, and the lust for power. So far Kant has simply borrowed from Rousseau. But Rousseau sometimes appears to think that an explanation of this corruption—of the emergence of what I will be calling the inflamed form of *amour-propre*—also suffices to explain vice, perhaps because he supposes that conscience is simply motivationally "weaker than the excited passions" (B 29); given a corrupt social order, then, people in it "must of necessity be what they are" (RJ 176–7). To be sure, Rousseau thinks it possible to take special measures both to diminish the unjustified craving to win superiority and strengthen the sense of duty associated with conscience, enabling a person better to resist the passions. *Emile* and *Social Contract* describe two ways to encourage such strength, and both underscore (with Kant) that a sense of duty is needed to avoid vice, once human powers are developed. But despite Rousseau's emphasis on the essential human importance of the power to choose, his distinction between the exercise of that power and natural mechanisms, and his insistence on our capacity to choose to resist inclinations (D2 141; E 280–1, 293), his explanation of vice is more naturalistic than Kant's, and does not develop the claim that choice underlies human corruption. Here Kant disagrees. To be sure, he accepts that the weakness or "frailty of human nature" (*Religion*, 77) is the first phase of vice, that the corruption of self-love makes it more difficult to throw off "the dominion of evil," and that "as far as we can see" the only path to avoiding such corruption and so to establishing the "sovereignty of the good principle" is to construct an "ethical community" (p. 130). Still, Kant emphasizes that a person with corrupted predispositions can resist their demands: we can only regard a greedy man, for example, as responsible for his conduct if we believe that such resistance is possible. In short, Kant distinguishes sharply between the corruption of our natural predispositions and the emergence of moral evil. While the corruption of

self-love has social causes, that corruption does not suffice to explain evil; instead, moral evil requires a choice (ultimately mysterious) to subordinate moral considerations to self-love.

7. I take this distinction from Rousseau's account of "perfectibility" (D2 141, 148, 159).
8. On love of others, and the role of valuations in such love, see D2 165.
9. The remarks that follow draw on Kant's remarks on the predispositions to humanity in *Religion*, 75, 129–30; John Rawls's discussion of self-respect and envy in *A Theory of Justice*, secs. 67, 81; and esp. N. J. Dent's *Rousseau* (Oxford: Blackwell, 1988), chaps. 2, 4.
10. See Kant's remarks on the predisposition to humanity, a comparative self-love expressed in the concern to "gain worth in the opinion of others" *Religion*, 75.
11. To what extent do people in socially subordinate positions internalize ruling conceptions of their "inferior" nature (or people in socially dominant positions embrace the regnant view of their "superior" nature)? For skeptical views about such internalization, see e.g. Orlando Patterson, *Slavery and Social Death* (Cambridge, Mass.: Harvard University Press, 1982); James Scott, *Domination and the Arts of Resistance: Hidden Transcripts* (New Haven: Yale University Press, 1990), esp. chap. 4.
12. Hegel's account of lordship and bondage suggests a further problem for inflated *amour-propre*: that it is directly self-defeating to depend for my sense of self on the recognition of others whom I denigrate, a "recognition that is one-sided and unequal," so that "the truth of the independent consciousness [of the lord] is accordingly the servile consciousness of the bondsman." See G. W. F. Hegel, *Phenomenology of Spirit*, trans. A. V. Miller (Oxford: Oxford University Press, 1977), 116–17. I do not find a view of this kind in Rousseau, but see Dent, *Rousseau*, 63–4.
13. I follow Dent, *Rousseau*, 52–6.
14. The trouble comes from differences presumed to be natural. Rousseau has no objection to taking pride in one's virtues, because virtues are achievements, not endowments: "The good man can be proud of his virtue because it is his. But of what is the intelligent man proud?" (E 245).
15. Rousseau's view of compassion and its diverse forms may be obscured by the discussion of compassion in the *Discourse on Inequality*. There, Rousseau associates natural compassion with a more or less mechanical reaction to suffering—the "pure movement of Nature prior to all reflection" (D2 152, emphasis added). But his *Essay on the Origin of Languages* and *Emile* present a different and more plausible picture: "Pity, although natural to the heart of man, would remain eternally inactive without the imagination that puts it into play. How do we let ourselves be moved to pity? By transporting ourselves outside of ourselves; by identifying with the suffering being ourselves and identifying with a being who suffers." But, he continues, "this transport presupposes acquired knowledge." And because it does, "He who has never reflected cannot be clement, or just, or pitying—no more than he can be wicked or vindictive" (OL 306). *Emile* states the

- cognitive preconditions of pity in similar language (E 222–3). There, too, he recognizes that compassion—like reciprocity or the concern to be treated with respect—can be part of our natural endowment, even if its expression is only triggered by circumstances.
16. In the *Second Discourse*, Rousseau adds the natural aversion to the suffering of others.
 17. It is not obvious why the natural desires should be so closely tied to needs. Of course, we could not survive as a species if we did not desire what we need. But there is no parallel problem with natural desires for goods that, while unnecessary for survival, are not detrimental to it—say, natural curiosity, or a natural desire for companionship. Rousseau offers this reason in *Emile* for tying natural desires to natural needs: “nature, which does everything for the best . . . gives him with immediacy only the desires necessary to his preservation and the faculties sufficient to supply them” (E 80). If we were endowed with desires that outstripped our natural powers, that would have been for the worse, since it would have been a source of unhappiness. But why suppose that nature does everything for the best? The doctrine of natural goodness does not require such a strong assumption.
 18. It is clear from the context that Rousseau is concerned with the “impetuous ardor” that men direct to women. See D2 156.
 19. See Kant’s discussion of how to reconcile the apparently conflicting strands of Rousseau’s view in his “Speculative Beginning of Human History,” 54–5.
 20. Starobinski suggests that Rousseau was simply torn between Innocence/Vice and Innocence/Virtue—between “an optimistic and a pessimistic version of the myth of origin.” *Transparency and Obstruction*, 15.
 21. See e.g. Seneca, *On Anger*, in *Seneca, in Ten Volumes*, vol. 1 (*Moral Essays*) (Cambridge, Mass.: Harvard University Press, 1928), 119.
 22. As a general matter, I will focus on the vice rather than the unhappiness. But the two are closely connected. True happiness consists in an equilibrium of desires and capacities (E 80). The vices are rooted in a desire for advantage over others, and an insistence that others acknowledge us as their betters. They make us miserable—irritated or discontented—because desires rooted in vice necessarily outrun our capacities.
 23. On Hobbes, see D2 159–60; on the Augustinian view, see B 937–8. Arthur Melzer contrasts Rousseau’s views with Platonist, as well as Hobbesian and Augustinian, conceptions. See his *The Natural Goodness of Man: The System of Rousseau’s Thought* (Chicago: University of Chicago Press, 1990), 17–23. According to Melzer, Platonism endorses a dualistic conception of the soul, with one part fixed as “selfish” and “irrational.” Human virtue, then, requires a constant struggle against part of our own nature. And he claims that Rousseau rejected this dualism and “inaugurated the great moral revolution” that emphasizes the unity of the soul and proposes to replace “wisdom and self-control with the new ethic of sincerity and spontaneity” (p. 22). This revolution represents a continuation of the Stoic tradition, “carrying on their argument against Plato and others in support of the

natural unity of the soul" (p. 21). Starobinski, too, suggests Rousseau's affiliation with Stoicism. See *Transparency and Obstruction*, 28, 37. This association strikes me as mistaken. To be sure, Rousseau agrees with the Stoic view that the passions are based on socially formed beliefs. But the role of the sense of duty in the final stages of Emile's education reveals Rousseau's distance from Stoicism. In explaining why Emile must leave Sophie, Emile's teacher rejects the Stoic ideal of extirpating the passions: "It is not within our control to have or not to have passions" (E 445). Moreover, once the passions have been awakened, goodness demands virtue, which depends in turn on a sense of duty and strength of will. "I have made you good rather than virtuous. But he who is only good remains so only as long as he takes pleasure in being so. Goodness is broken and perishes under the impact of the human passions. . . . Who, then, is the virtuous man? It is he who knows how to conquer his affections; for then he follows his reason and his conscience; he does his duty." He continues: "Now be really free. Learn to become your own master. Command your heart, Emile, and you will be virtuous." Elaborating the conception of command, the teacher says: "It is an error to distinguish permitted passions from forbidden ones in order to yield to the former and deny oneself the latter. All passions are good when one remains their master; all are bad when one lets oneself be subjected to them. . . . [W]hat is forbidden to us by conscience is not temptations but rather letting ourselves be conquered by temptations. It is not within our control to have or not to have passions. But it is within our control to reign over them. All the sentiments we dominate are legitimate; all those which dominate us are criminal" (E 444–5; see also 446, 473; RJ 158). And in the continuation of this passage, he explains that to "dominate" our sentiments is to follow "the law of duty." Rousseau's motto for *Emile* comes from Seneca's *De Ira*, but we find nothing in Seneca's treatise corresponding to Rousseau's account of "reigning over passions" and resisting temptations by following the law of duty. See Seneca, *On Anger*, 106–356. Cassirer is more nearly right when he says that "Rousseau's ethics is not an ethics of feeling but the most categorical form of a pure ethics of obligation (*Gesetzes-Ethik*) before Kant." See Cassirer, *Question of Jean-Jacques Rousseau*, 96. But Cassirer neglects issues of motivational strength, and so omits any account of the complexities (explored below, pp. 122–7) of ensuring a fit between the general will and our affections. On the Stoic idea of extirpating the passions, see Martha C. Nussbaum, *The Therapy of Desire: Theory and Practice in Hellenistic Ethics* (Princeton: Princeton University Press, 1994), chaps. 10, 11. I am grateful to Martha Nussbaum for discussion of the themes in this note.

24. *Religion*, 75.

25. How much inequality, and of what sort? Rousseau's scattered remarks do not add up to a sustained answer. See below, p. 140.

26. The person who does what is right without having to resist inclinations is at best "good without merit." Genuine virtue depends on following the

- sense of duty in the face of conflicting temptations. See E 473; SC 1.8.1; RJ 158.
27. Rousseau's institutional strategies to ensure congruence between motives and requirements of the common good are discussed below, pp. 152–64.
 28. Recall that the “whole of human wisdom in the use of the passions” is to keep a “sense of the true relations of man” and then to “order all the affections of the soul according to these relations” (E 219).
 29. See Immanuel Kant, *The Metaphysical Principles of Virtue: Part II of the Metaphysics of Morals*, trans. James Ellington (Indianapolis: Bobbs-Merrill, 1964), §§34–5.
 30. *Ibid.*, §35.
 31. Melzer presents an illuminating discussion of four arguments for the doctrine of natural goodness—introspective, psychological, social, and historical arguments. And he mentions in passing (p. 30 n. 1) a fifth, metaphysical argument. My rationale for the doctrine is specifically moral, and seems different from any of the five views he considers. Melzer, *Natural Goodness*, pt. 2.
 32. The Profession of Faith, as I am interpreting it, anticipates Kant's conception of reasonable faith. See, e.g. the “Dialectic of Pure Practical Reason,” in *Critique of Practical Reason*, 226–58. For a moving statement of the animating moral idea behind the idea of reasonable faith—a statement that powerfully reveals the deep affinity of Kant and Rousseau—see the concluding paragraph in “The Canon of Pure Reason,” in the *Critique of Pure Reason*, trans. Norman Kemp Smith (London: Macmillan, 1929), A831–B859. On the connections between the Vicar's Profession of Faith and Kant's moral religion, see Dieter Henrich, “The Moral Image of the World,” in *Aesthetic Judgment and the Moral Image of the World: Studies in Kant* (Stanford: Stanford University Press, 1992), 3–28.
 33. Immanuel Kant, *On the Old Saw: That May Be Right in Theory But It Won't Work In Practice*, trans. E. B. Ashton (Philadelphia: University of Pennsylvania Press, 1974), 77–8.

CHAPTER 5

1. Rousseau officially confines the term “democracy” to political regimes in which executive power is in the hands of all citizens. See SC 3.3.2. Outside of the *Social Contract*, however, he commonly uses the term for all systems with popular sovereignty, whatever the constitution of the executive. See e.g. LM 240, 257–8. Here I will be following the latter usage.
2. Juergen Habermas, “Legitimation Problems in the Modern State,” in *Communication and the Evolution of Society*, trans. Thomas McCarthy (Boston: Beacon Press, 1975), 185–6.
3. The sketch omits the tribunate, dictatorship, and censorship discussed in SC 4.5–7. For discussion, see Gildin, *Rousseau's Social Contract*, chap. 6.

For reasons that I discuss on pp. 13–14, I also will not consider the role of the legislator. These elements of Rousseau's view are important and interesting, but a discussion of them would not change the basics of my account here.

4. In PE 12, Rousseau argues against direct democracy. In his constitution for Corsica, as with Poland, Rousseau does not require direct assembly democracy. See P 197–202; C 285–6. In LM, he shows some uncharacteristic (for Rousseau) appreciation of the virtues of the English constitution. See LM 246, 251–2, 288–91. For insightful discussion, see Richard Fralin, *Rousseau and Representation* (New York: Columbia University Press, 1978). Fralin is attentive to the different forms of legislative authority, but suggests (p. 184) that Rousseau did not think that there is any significant difference in outcomes between a scheme in which voting is conducted in a single assembly for all citizens, and a system in which subsets of the population meet in separate assemblies, those subsets choose delegates to a national assembly, and give instructions to their deputies who are bound by the majority decision of the electors. Whatever Rousseau may have supposed, this conclusion is incorrect, even if the local assemblies are of equal size. Suppose the citizens divide into local assemblies of equal size. In the local assemblies they select representatives to a national assembly. These representatives are each given one vote and are required to vote the decision of the local assembly. Then even if both local and national assemblies require a two-thirds majority to enact a law, laws can pass that are supported by a minority of the citizens (44.4% is sufficient, i.e. two-thirds support in two-thirds of the assemblies, and no support elsewhere).
5. They “separate power from the Right” in that, on the one hand, the *Règlement* recognized the right of the Genevan bourgeoisie to criticize the decisions of the government (the so-called “right of remonstrance”), but on the other hand Article 25 of the *Règlement* denied them the power to exercise that right by preventing them from assembling. But, he asks, “Why deprive it [the Bourgeoisie] of the means of deliberating within itself” (LM 269–71). See also the notes for *Letters Written from the Mountain* by Jean-Daniel Candaux, in Jean-Jacques Rousseau, *Œuvres complètes*, vol. 3, (Paris: Gallimard, 1964), 1698–1700.
6. On the virtues and vices of “multi-faction” democracies, see Robert Dahl, *Dilemmas of Pluralist Democracy: Autonomy vs. Control* (New Haven: Yale University Press, 1982), esp. chap. 3. In several places Rousseau endorses a system of “partial societies” more strongly than he does in the *Social Contract*. Such an endorsement is suggested by the account of the *circles* in LD (see below, pp. 171–2), and is provided in a discussion of Sparta and Rome which indicates that both had “many electoral divisions and private societies [*sociétés particulières*]” and that these encouraged a “love of country.” See FP 542.
7. In *Corsica*, for example, Rousseau proposes not “to destroy private property absolutely, since that is impossible, but to confine it within the narrowest possible limits; to give it a measure, a rule, a rein which will contain, direct,

- and subjugate it, and keep it even subordinate to the public good. In short, I want the property of the state to be as large and strong, that of the citizens as small and weak, as possible" (Cor. 317).
8. Regulation of the distribution of wealth might, for example, proceed through a tax system in which "Someone who has only the bare necessities should not pay anything at all; taxation on someone who has superflux may, if need be, go up to the full amount that exceeds his necessities" (PE 31). On the close connections between power and wealth, see D2 171; C 327–8.
 9. See John Stuart Mill, *Considerations on Representative Government* (Buffalo: Prometheus Books, 1991), chap. 2; Rawls, *A Theory of Justice*, sec. 69.
 10. For an insightful discussion of Bentham's view, see Elie Halévy, *The Growth of Philosophical Radicalism*, trans. Mary Morris (London: Faber and Faber, 1928), and Ross Harrison, *Bentham* (London: Routledge, 1983). In chap. 5, Harrison provides an illuminating discussion of Bentham's "duty and interest junction principle." Also see chap. 3 of John Stuart Mill, *Utilitarianism* (Indianapolis: Bobbs-Merrill, 1957). Mill's idea of a natural sense of human unity presents an alternative to Bentham's idea of an artificial identification of interests, and suggests a way that individuals might come to identify more directly with the principle of utility as a standard of right conduct. The principle of utility would not, then, be an "external norm" in the sense described below.
 11. Consider, for example, Rawls's conception of a "well-ordered society," as a society whose members share a conception of justice, which is "normally effective": it "enables them to understand and apply the publicly recognized principles of justice, and for the most part to act accordingly as their position in society, with its duties and obligations, requires." John Rawls, *Justice as Fairness: A Restatement*, ed. Erin Kelly (Cambridge, Mass.: Harvard University Press, 2001), 9.
 12. Rousseau says that this "maxim" clarifies his own peculiarities. He has sought, he says, to avoid situations in which he would be tempted to wish harm to others. Describing it as "true philosophy," he has recommended it, he says, in all his later writings.
 13. On the rights of citizens in areas of security and personal liberty, see SC 2.4.10, and the discussion of punishment in LM 280.
 14. Rousseau also affirms, as I said earlier, a right to religious liberty within the limits set by the dogmas of civil religion: that God exists, that there is an afterlife, that the soul is immortal, that the just are rewarded and the wicked punished, that the social contract and the laws are sacred, and that there is to be no toleration for the intolerant (SC 4.8.33). While these dogmas are religious in content, they have a secular justification, founded on the claim that the rejection of these doctrines is bound to undermine public order (SC 4.8.31–2; LM 134–65).
 15. I do not mean that slaves are in fact simply extensions of the will of their masters, but only that they are so represented in the normative and legal accounts of slavery. On the conflict between that representation and the

- realities of slavery, see Joshua Cohen, “The Arc of the Moral Universe,” *Philosophy and Public Affairs* 26/2 (Spring 1997), 91–134.
16. James Madison, *Federalist* 49, in Alexander Hamilton, James Madison, and John Jay, *The Federalist Papers* (London: Penguin, 1987).
 17. On freedom as non-subordination, see Philip Pettit, *Republicanism: A Theory of Freedom and Government* (Oxford: Oxford University Press, 1997); see also Friedrich Hayek, *The Constitution of Liberty* (Chicago: University of Chicago Press, 1960), 11–12.
 18. Locke, *Second Treatise*, paras. 142, 143.
 19. See *Republic*, 562 ff.
 20. The view that I will discuss is developed in broad outlines in Shklar’s account of the “politics of prevention” in *Men and Citizens*; Fralin’s *Rousseau and Representation* fills in the institutional details. While Shklar and Fralin diverge on some points, the basic lines of argument are similar. I focus here more on Fralin because he has more institutional detail. My main aim, however, is not to criticize the alternative view, but to explore further the general structure of argument I have already presented by showing how additional elements of Rousseau’s view can be accommodated within it.
 21. See Fralin, *Rousseau and Representation*.
 22. *Men and Citizens*, 20.
 23. Fralin argues that this represents a change in Rousseau’s views after the *Social Contract*, a change motivated by the condemnation of his work in Geneva, which made him even more sensitive to the dangers of the power of magistrates. His sensitivity there is very great: see the powerful statement at LM 238–9. But once the confusion about executive power in *Social Contract* itself has been eliminated, there is no reason to believe that the *Social Contract* endorses the practice of vesting the power of nomination in the government, and therefore no evidence for the change of view.
 24. Madison, *Federalist* 49.
 25. See Plato, *Republic*, 557–8, 562 ff.; *Lev.*, 131–2.

Bibliography

- Althusser, Louis, *Politics and History*, trans. Ben Brewster (London: New Left Books, 1972).
- Aristotle, *Politics*, trans. Benjamin Jowett, in *The Complete Works of Aristotle*, ed. Jonathan Barnes, vol. 2 (Princeton: Princeton University Press, 1984).
- Barry, Brian, "The Public Interest," *Proceedings of the Aristotelian Society* 38 (1964), 9–14.
- Bertram, Christopher, *Rousseau and the Social Contract* (London: Routledge, 2004).
- Bloom, Allan, "Rousseau's Critique of Liberal Constitutionalism," in Clifford Orwin and Nathan Tarcov (eds.), *The Legacy of Rousseau* (Chicago: University of Chicago Press, 1997), 143–67.
- Cassirer, Ernst, *Kant's Life and Thought* (New Haven: Yale University Press, 1981).
- *Question of Jean-Jacques Rousseau*, trans. Peter Gay (Indianapolis: Indiana University Press, 1963).
- Chomsky, Noam, *Cartesian Linguistics* (New York: Harper and Row, 1966).
- *Reflections on Language* (New York: Pantheon, 1975).
- Cobban, Alfred, *Rousseau and the Modern State* (London: Archon, 1964).
- Cohen, Joshua, "The Arc of the Moral Universe," *Philosophy and Public Affairs* 26/2 (Spring 1997), 91–134.
- "The Natural Goodness of Humanity," in Andrews Reath, Barbara Herman, and Christine Korsgaard (eds.), *Reclaiming the History of Ethics: Essays for John Rawls* (Cambridge: Cambridge University Press, 1997).
- *Philosophy, Politics, Democracy: Selected Essays* (Cambridge, Mass.: Harvard University Press, 2009).
- "Protection for Obedience: Hobbes's Theory of the State" (unpublished).
- "Structure, Choice, and Legitimacy: Locke's Theory of the State," *Philosophy and Public Affairs* 15/4 (Autumn 1986), 301–24.
- Dahl, Robert, *Dilemmas of Pluralist Democracy: Autonomy vs. Control* (New Haven: Yale University Press, 1982).
- Damrosch, Leo, *Jean-Jacques Rousseau: Restless Genius* (Boston and New York: Houghton-Mifflin, 2005).

BIBLIOGRAPHY

- Dent, N. J., *Rousseau* (Oxford: Blackwell, 1988).
- Derathé, Robert, *Jean-Jacques Rousseau et la science politique de son temps* (Paris: Vrin, 1950).
- Durkheim, Emile, *Montesquieu and Rousseau* (Ann Arbor: University of Michigan Press, 1960).
- Fralin, Richard, *Rousseau and Representation* (New York: Columbia University Press, 1978).
- Gauthier, David, *Morals By Agreement* (Oxford: Oxford University Press, 1986).
- Gildin, Hilail, *Rousseau's Social Contract* (Chicago: University of Chicago Press, 1983).
- Grofman, Bernard, Guillermo Owen, and Scott L. Feld, "Thirteen Theorems in Search of Truth," *Theory and Decision* 15 (1983), 261–78.
- Grofman, Bernard, and Scott L. Feld, "Rousseau's General Will: A Condorcetian Perspective," *American Political Science Review* 82 (June 1988), 567–76.
- Habermas, Juergen, *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*, trans. William Rehg (Cambridge, Mass.: MIT Press, 1996).
- "Legitimation Problems in the Modern State," in *Communication and the Evolution of Society*, trans. Thomas McCarthy (Boston: Beacon Press, 1975), 185–6.
- "On the Internal Relation Between Law and Democracy," in *The Inclusion of the Other: Studies in Political Theory* (Cambridge, Mass.: MIT Press, 1998), chap. 10.
- Halévy, Elie, *The Growth of Philosophical Radicalism*, trans. Mary Morris (London: Faber and Faber, 1928).
- Harper, Anne L., "The Family and the State in Rousseau's 'Emile, Or On Education,'" unpublished dissertation, University of Michigan, 1987.
- Harrison, Ross, *Bentham* (London: Routledge, 1983).
- Hayek, Friedrich, *The Constitution of Liberty* (Chicago: University of Chicago Press, 1960).
- Hegel, Georg Wilhelm Friedrich, *Lectures on the History of Philosophy*, vol. 3 (London: Routledge, 1892).
- *Phenomenology of Spirit*, trans. A. V. Miller (Oxford: Oxford University Press, 1977).
- Hendel, Charles W., *Jean-Jacques Rousseau Moralizer* (Indianapolis: Bobbs-Merrill, 1934).

BIBLIOGRAPHY

- Henrich, Dieter, "The Moral Image of the World," in *Aesthetic Judgment and the Moral Image of the World: Studies in Kant* (Stanford: Stanford University Press, 1992), 3–28.
- Hobbes, Thomas, *Leviathan*, ed. Richard Tuck (Cambridge: Cambridge University Press, 1991).
- *Man and Citizen: Thomas Hobbes's "De Homine" and "De Cive"*, ed. Bernard Gert (New York: Anchor Books, 1972).
- Kant, Immanuel, *Critique of Practical Reason*, in Immanuel Kant, *Practical Philosophy*, trans. and ed. Mary J. Gregor (Cambridge: Cambridge University Press, 1996).
- *Critique of Pure Reason*, trans. Norman Kemp Smith (London: Macmillan, 1929).
- *The Metaphysical Principles of Virtue: Part of the Metaphysics of Morals*, trans. James Ellington (Indianapolis: Bobbs-Merrill, 1964).
- *On the Old Saw: That May Be Right in Theory But It Won't Work In Practice*, trans. E. B. Ashton (Philadelphia: University of Pennsylvania Press, 1974).
- *Religion Within the Bounds of Mere Reason*, in Immanuel Kant, *Religion and Rational Theology*, trans. and ed. Allen W. Wood (Cambridge: Cambridge University Press, 1996).
- "Speculative Beginning of Human History," in *Perpetual Peace and Other Essays*, trans. Ted Humphrey (Indianapolis: Hackett, 1983).
- Keohane, N. O., *Philosophy and the State in France* (Princeton: Princeton University Press, 1980).
- Korsgaard, Christine M., *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), chap. 6.
- Locke, John, *Second Treatise*, in *Two Treatises of Government*, ed. Peter Laslett (Cambridge: Cambridge University Press, 1988).
- Madison, James, *Federalist 49*, in Alexander Hamilton, James Madison, and John Jay, *The Federalist Papers* (London: Penguin, 1987).
- Maistre, Joseph de, *Against Rousseau*, trans. Richard A. Lebrun (Montreal: McGill-Queen's University Press, 1996).
- Masters, Roger, *The Political Philosophy of Rousseau* (Princeton: Princeton University Press, 1968).
- Melzer, Arthur, *The Natural Goodness of Man: The System of Rousseau's Thought* (Chicago: University of Chicago Press, 1990).
- Mill, John Stuart, *Considerations on Representative Government* (Buffalo: Prometheus Books, 1991).
- *Utilitarianism* (Indianapolis: Bobbs-Merrill, 1957).

BIBLIOGRAPHY

- Miller, James, *Rousseau: Dreamer of Democracy* (New Haven: Yale University Press, 1984).
- Neuhouser, Frederick, *Rousseau's Theodicy of Self-Love: Evil, Rationality, and the Drive for Recognition* (New York: Oxford University Press, 2008).
- Nussbaum, Martha C., *The Therapy of Desire: Theory and Practice in Hellenistic Ethics* (Princeton: Princeton University Press, 1994).
- Okin, Susan Moller, *Women and Political Thought* (Princeton: Princeton University Press, 1979).
- Parfit, Derek, *Reasons and Persons* (Oxford: Oxford University Press, 1984).
- Patterson, Orlando, *Slavery and Social Death* (Cambridge, Mass.: Harvard University Press, 1982).
- Pettit, Philip, *Republicanism: A Theory of Freedom and Government* (Oxford: Oxford University Press, 1997).
- Plato, *The Republic*, ed. G. R. F. Ferrari, trans. Tom Griffith (Cambridge: Cambridge University Press, 2000).
- Rawls, John, *Justice as Fairness: A Restatement*, ed. Erin Kelly (Cambridge, Mass.: Harvard University Press, 2001).
- *Lectures on the History of Political Philosophy*, ed. Samuel Freeman (Cambridge, Mass.: Harvard University Press, 2007).
- *A Theory of Justice*, rev. edn. (Cambridge, Mass.: Harvard University Press, 1999).
- Riley, Patrick, *The General Will Before Rousseau: The Transformation of the Divine into the Civic* (Princeton: Princeton University Press, 1986).
- *The Will and Political Legitimacy* (Cambridge, Mass.: Harvard University Press, 1982).
- Runciman, W. G., and Amartya Sen, "Games, Justice, and the General Will," *Mind* 74 (1965), 554–62.
- Schwartz, Joel, *The Sexual Politics of Jean-Jacques Rousseau* (Chicago: University of Chicago Press, 1984).
- Scott, James, *Domination and the Arts of Resistance: Hidden Transcripts* (New Haven: Yale University Press, 1990).
- Seneca, *On Anger*, in *Seneca, in Ten Volumes*, vol. 1 (*Moral Essays*) (Cambridge, Mass.: Harvard University Press, 1928).
- Shklar, Judith, *Men and Citizens: A Study in Rousseau's Social Theory*, 2nd edn. (Cambridge: Cambridge University Press, 1985).
- Spink, J. S., *Jean-Jacques Rousseau et Genève* (Paris: 1934).
- Sreenivasan, Gopal, "What is the General Will?," *Philosophical Review* 109/4 (2000), 545–81.

BIBLIOGRAPHY

- Starobinski, Jean, *Jean-Jacques Rousseau: Transparency and Obstruction*, trans. Arthur Goldhammer (Chicago: University of Chicago Press, 1988).
- Strauss, Leo, *Natural Right and History* (Chicago: University of Chicago Press, 1953).
- Viroli, Maurizio, *Jean-Jacques Rousseau and the "Well-Ordered Society"*, trans. Derek Hanson (Cambridge: Cambridge University Press, 1988).
- Waldron, Jeremy, David Estlund, Bernard Grofman, and Scott Feld, "Democratic Theory and the Public Interest: Condorcet and Rousseau Revisited," *American Political Science Review* (Dec. 1989).

This page intentionally left blank

Index

- amour de soi*
see self-love
- amour propre* 101, 103–4, 107, 109,
112, 115, 119–20, 126, 158
- Aristotle 34 n. 3, 34 n. 4
- assemblies
legislative 137, 171
popular 133–4, 155–6
- associations 76, 139, 139 n. 6, 155
- Augustine 114
- authoritarianism 1
political 166
- authority
legitimate 25, 68, 133–4
political 3, 14, 75, 80, 86, 167
sovereign 48, 66, 138
supreme 51, 58
- autonomy
see self-legislation
- Beaumont, Archbishop of Paris 97,
114
- Bentham, J. 141–2, 159
- Bloom, A. 34–7
- Cassirer, E. 35–7, 114 n. 23
- Chomsky, N. 99
- civic unity 21, 34, 35–6, 38, 40, 55,
57, 72–3
complete 34–5, 40, 55
unity through integration
see complete
unity through ordering 40, 55–8
see also Plato
- civic virtue 5, 113
- civil religion 21, 46–7, 49, 147
religious toleration and 44–9
- cognitive powers 106–7, 111, 118
- common good, the 15, 40, 55, 68
baseline for 49–53
distributive/aggregative
constraint 41–2
equality/common good
constraint 42–3
individual independence and 44–9
self-development and 44
- Common Good Condition (GW2) 40,
73, 83, 140, 152, 154
- communitarianism 5, 21–2, 94–6
- compassion 98, 101, 104–7, 125–7
- Condorcet jury theorems 78, 78 n. 3,
165
- conflicts of interest 27, 116
- consensus 3, 68–73
see also political disagreement
- continuous proportion theory (size of
government) 161–2
- Damrosch, L. 8
- democracy
democratic institutions, theory
of 24
democratic lawmaking 134
depth of Rousseau's commitment
to 167–76
direct 1, 16, 20, 133–4, 136,
148–50, 156
democracy requirement,
minimum 137–8, 169–70
public deliberation in
see public deliberation
- Dent, N.J. 4, 101 n. 9
- Durkheim, E. 35–7
- duty, sense of 18, 88, 113, 122–5,
127, 145
- elections 169–70
- executive power 136–8, 149, 160–1,
169–70
see also government
- Fralin, R. 136 n. 4, 166 n. 20, 167, 170
n. 23
- freedom 11, 21, 24–5, 27–33, 91–4
as non-subordination 163
inalienable 29–32
moral 12–14
- free community of equals 10, 16, 20,
58, 59, 84, 98, 131, 146
- fundamental problem, the 2, 11,
15–16, 20, 24, 32–3, 84
solution to 84–96

INDEX

- general will, the 15–16, 53, 62, 66, 77
 common good and 15–16, 40–54
 exclusive direction by 34, 38
 procedural interpretation of 75–7
 rights and
 see rights
 society of, four defining conditions
 (GW1–GW4) 34–58
 supreme direction by 33–4, 37–8,
 51, 53, 58, 68, 82, 85, 91–4
- government
 government accountability 138,
 161
 government, theory of 160
 governmental power, delegation
 of 137
see also executive power
- Habermas, J. 20, 67, 133–4
 Hegel, G.W.F. 1, 102 n. 12
 Hobbes, T. 11–13, 18, 29, 33, 38, 52,
 60, 62–8, 86, 87, 94, 109, 114, 154
 good laws 64–5, 68
 just laws 64
 Leviathan state 13, 17
 self-effacing Hobbesianism 87–96
 supreme particular will 64
 theory of human nature 114–15
 theory of sovereignty and law 66
- human nature 19, 25, 99–100, 111,
 113–14
- inequality 2, 53, 113, 116–17, 120–1,
 140, 155, 164
- institutional design 109, 110, 126–7,
 141–5, 174
 ethical conception of 142–3
 rational choice/incentive alignment
 conception of 141–2
- institutional possibility
see realism
- institutions
 epistemic reasons for 164–6
 principles and 146
 temptation/assurance and 156–7,
 165–6
 theory of 140, 143, 152, 174
 will formation and 152
- interdependence
 social 25–7, 32, 93, 108
- Kant, I. 1 n. 2, 2, 98, 116, 125 n. 30,
 130
 on explaining evil 100 n. 6
 moral law 91
 reasonable faith 129 n. 32
- language 5, 99, 108
- law
 Hobbes and Rousseau on 62–8
 lawmaking 132, 136
 rule of 135, 160, 162–4
 theory of 62–8, 162
- legislative initiative 167, 169, 172–5
 executive role in 172–3
- legislative power 135–8
- Leviathan state
see Hobbes
- Locke, J. 29, 33, 86, 163 n. 18
 Lockean social contract 52
- majority rule 7, 77–8, 79–80, 137,
 165
- Marx, K. 1, 8
- Melzer, A. 114 n. 23
- Mill, J. S. 141 n. 9, 141 n. 10
- motivations
 human 106–10, 130
 motivational complements 123–4,
 126, 145, 153, 158
 motivational possibility
see realism
 social bases of 109
- natural goodness of humanity 16,
 19–20, 97–100, 105–6, 113
 brutishness/virtue 111
 doctrine of 4, 111–12, 122, 129–30
 innocence/vice 111, 112
 innocence/virtue 111
 natural innocence 111
- Particular Interest Condition
 (GW1) 34, 36, 145, 157
- pitié*
see compassion
- Plato 34, 39–40, 55, 166
- political autonomy 13, 18, 73, 80–2
see also self-government/
 self-legislation

INDEX

- political disagreement 5, 21, 69–73,
 79–80, 82
 see also consensus
- political participation 20, 147,
 148–52, 155, 166, 168, 174–5
- political right, principles of 24, 33,
 91–2
- Priority Condition (GW3) 62, 141,
 144, 152, 154, 157
- private property 119–20, 139–40,
 140 n. 7, 158
- Profession of Faith 129
- public deliberation 69, 70, 76–7, 127,
 155, 170–2
 no-communication condition 76
- public opinion 107
- Rawls, J. 1–2, 5, 7 n. 9, 33, 39–40,
 42, 51, 141 n. 9
- realism 14, 16, 19, 20, 23, 132
 institutional possibility 19, 132
 motivational possibility 16,
 18–19, 104, 122, 128, 130
- Reasonable Confidence Condition
 (GW4) 58, 73, 141
- reciprocity 125–6, 153–4
- Reformation Christianity 44–5, 47
- representation 148–9, 151, 152, 167
- rights 82–3, 146–8
 assembly 139
 basic 146
 participation 147
 property 139–40, 155
- self-government/self-legislation 14,
 16–17, 23, 45–6, 73–5, 84, 86,
 131, 175–6
 see also political autonomy
- self-interests 87–9, 94
- self-love 11–12, 25, 32–3, 42, 84,
 86–7, 98, 101, 106, 116–17,
 120–1, 124–7, 130–1, 154,
 158
- self-preservation 25, 63
- self-regard/self-worth 98–105,
 120–2, 124–5, 127, 158
- Seneca 113 n. 21, 114 n. 23
- Shklar, J. 16 n. 6, 168
- slavery 148, 150, 163
- social compact/contract 22, 33, 38,
 42, 44, 51, 68–9, 83–4, 86–7,
 89–95
- social solidarity 5, 21–2, 96, 135
- sovereignty
 see authority, general will, Hobbes
- Starobinski, J. 98 n. 4, 112 n. 20
- suffrage 137
- unanimity
 see consensus
- vice, a genealogy of 113, 115,
 129–30, 143, 153, 158
- will of all 25–7
- will-formation 152–6, 168