

Catherine Wilson

MORAL ANIMALS

IDEALS AND CONSTRAINTS IN MORAL THEORY



OXFORD

Moral Animals

Ideals and Constraints in Moral Theory

In *Moral Animals* Catherine Wilson develops a theory of morality based on two fundamental premises: first that moral progress implies the evolution of moral ideals involving restraint and sacrifice; second that human beings are outfitted by nature with selfish motivations, intentions, and ambitions that place constraints on what morality can demand of them. Normative claims, she goes on to show, can be understood as projective hypotheses concerning the conduct of realistically-described nonideal agents in preferred fictional worlds. Such claims differ from empirical hypotheses, insofar as they cannot be verified by observation and experiment. Yet many, though not all, moral claims are susceptible of confirmation to the extent that they command the agreement of well-informed inquirers.

With this foundation in place, Wilson turns to a defence of egalitarianism intended to address the objection that the importance of our non-moral projects, our natural acquisitiveness and partiality, and our meritocratic commitments render social equality a mere abstract ideal. Employing the basic notion of a symmetrical division of the cooperative surplus, she argues that social justice with respect to global disparities in well-being, and in the condition of women relative to men, depends on the relinquishment of natural and acquired advantage that is central to the concept of morality.

Moral Animals will spark fresh debates within philosophy and across the social sciences.

This page intentionally left blank

Moral Animals

Ideals and Constraints in Moral Theory

CATHERINE WILSON

CLARENDON PRESS · OXFORD

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 Bangkok Buenos Aires Cape Town Chennai
Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata
Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi
São Paulo Shanghai Taipei Tokyo Toronto

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

© Catherine Wilson 2004

The moral rights of the author have been asserted

Database right Oxford University Press (maker)

First published 2004

First published in paperback 2007

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 this same condition on any acquirer

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloging in Publication Data

Data available

Typeset by SPI Publisher Services, Pondicherry, India

Printed in Great Britain

on acid-free paper by

Biddles Ltd., King's Lynn, Norfolk

ISBN 978-0-19-926767-5 (Hbk.) 978-0-19-922809-6 (Pbk.)

1 3 5 7 9 10 8 6 4 2

To my maternal grandparents,
Harry Helson (1898–1977) and Lida G. Helson (1900–1979)
in memoriam

This page intentionally left blank

Preface

The aims of this book are first, to furnish a foundation for moral theory that is independent of any particular set of moral commitments and second, to defend a particular version of egalitarianism on that foundation. Though meta-ethics and political philosophy can be and are often treated independently, there is a reason for offering a content-neutral theory of moral judgement and moral practice along with a defence of particular normative claims. The most powerful arguments against egalitarianism in contemporary moral theory gain much of their force from the ostensibly non-normative theories of the place of the self in the world and the allied accounts of the nature of moral judgement that frame them.

Moral judgements, according to the descriptive theory advanced here, form a subset of normative judgements. Unlike aesthetic and non-moral practical judgements regarding what ought to be done, they reflect the endorsement of advantage-reducing rules on the part of those who assert them. Moral rules are rules, one might say, for not getting ahead. Morality is the system laid down to compensate for the wear and tear that is the unavoidable by-product of our ordinary strivings, through the imposition of certain sacrifices and deprivations.

This might seem puzzling. Though the fiercer and darker aspects of morality were emphasized by anthropologists earlier in the century, what might be termed Freudian pessimism has lost ground to a conception of morality as a source of human flourishing. The motivation behind this equation is understandable; the prohibitory taboos of our ancestors are viewed with scepticism if not dismay, and there can be no doubt that our opportunities and well-being depend on the moral behaviour of others towards us, especially their veracity, impartiality, and benevolence. Yet, the relationship between morality and flourishing is mediate and qualified, not direct and unqualified. Observing the norms of finance, cookery, decorating, and intimate relationships helps us to live good human lives, while meritocratic institutions enable us to parlay our talents and attributes into wealth and influence. Morality, as Kant speculated, is for something other than worldly success, though it is not necessarily for the expression of our rationality or the use of our noumenal wills. Platitudes regarding human

flourishing obscure much of what is interesting and difficult in morality and in moral theory. How many of us can claim that the breaking of an agreement, or some show of partiality, or some occlusion of the truth, has never helped us to carry on with our lives as we wished to, and that moral considerations have never held us up?

Humans are disposed to invent, observe, and enforce advantage-reducing rules of varying degrees of stringency, and they hold high status in the hierarchy of social norms. Yet, as we might expect given their origins, there is considerable disagreement as to how far agents can reasonably be required to restrain and suppress the operation of their natural and acquired powers for their own enjoyment and benefit. In my local culture, for example, we agree that one may not walk into unlocked houses and make off with people's television sets, but we do not agree on whether the manufacturers of television sets may set their wages at whatever level they find to be most profitable. We agree that the well-off have some responsibility for the sick, poorly educated, and demoralized members of the underclass, but we disagree on how far their needs should cut into our enjoyments. We believe that persons should enjoy the attentions of one spouse at a time, no matter how many others they could attract and maintain with money or savoir-faire, but we disagree over their entitlement to non-marital friendship or adventure. Different codes enjoin different degrees of advantage renunciation on members of the communities bound by them or individuals who subscribe to them. Morality is, in this respect, scalar.

First-order moral argument is sometimes aimed at establishing what to do when advantage-reducing rules conflict in a moral emergency. The obligation to do all that one can to save a life may conflict with the duty not to prolong someone's suffering by employing showy medical expertise. But first-order arguments are often addressed simply to the question how moral to be. Moral rules are such that we often feel burdened by them, resist them, and produce what are often plausible justifications for our non-compliance, even when there is no emergency and no conflict between competing obligations. Strict adherence to a principle of veracity can be highly disadvantageous to an agent; loyalty to difficult and demanding friends can prove exhausting; and requests from worthwhile charitable organizations can be irritating. We are faced, in other words, with the problem of exigency. When can aesthetic, prudential, or simple hedonic considerations justify an exemption from an obligation that has been assumed by an agent, or that is held to be generally binding? Are we really required to act as the

famous modern moral theories, such as Kantianism and utilitarianism, say we must?

The argument that the burdens these theories impose—including the burden of submitting all one's proposed courses of action to the test of an impersonal theory—are too great for human beings as they are constituted by nature is frequently cited as a defeater of their seemingly exigent requirements, particularly with regard to issues of social justice. Weaker aspirations with respect to socio-economic and sexual equality have been a striking feature of recent prescriptive moral theory and the reasoning behind this lowering of demand levels has been set out with formidable intelligence in books and articles published over the last two decades and is documented below. No contemporary moral theorist has presented these meta-ethical and substantive issues with greater force and clarity than Bernard Williams, who died as this book was undergoing its last set of revisions.

Williams's meta-ethical scepticism with regard to moral realism and demonstrable obligations is defended here as well founded. To a large extent, it is up to each of us how moral we want to be and what sacrifices we are willing to incur. Moral theorizing is constrained by what we want, now, not what our ideally rational selves ought to want, and by what we find it easy and difficult to do. Nevertheless, it is possible to preserve a good deal more of the revisionary content of the famous modern moral theories than Williams and other critics believed to be possible.

To meet the sceptic's objection to the very idea of an obligation that could be independent of an agent's motivational state, I offer a modal theory of moral judgements that bypasses the question whether moral judgements or prescriptive theories can be true. The assertion that an action in a given context is obligatory has both representational and conative content. The representational content of a moral judgement is given by an idealized moral world. Roughly, to assert that action *ACT* is obligatory in circumstances *c* is to claim that, in a morally good world otherwise like ours, agents all do *ACT* in *c*. A satisfactory theory of morals is a representation of an ideal world that is, all things considered, with respect to its advantage-reducing behaviour, preferable to rival worlds instantiating different behaviour. Though we have no direct access to ideal worlds, only to our own, imperfect one, moral judgements are in principle confirmable. Theorists with different prescriptive commitments disagree on what things are like in a good world. They advance their favoured candidates, projecting paraworlds, fictional worlds for which both verisimilitude and moral goodness are implicitly claimed,

and argue that they are the best. Though the account offered here has significant conceptual connections with both modern contractualism and modern consequentialism, it is formally distinct from them both.

One is obliged to do what a theory says, to the extent that it has been confirmed, but no further. Unconfirmed moral judgements are mere authored norms, with few or many adherents. The confirmation of a moral claim requires only the agreement of reasonable, competent, well-informed judges as to betterness relations between worlds, not agreement as to what the criteria of betterness are. Analogously, confirmation in physics requires agreement by competent judges about a physical phenomenon, not about confirmation theory, a technical branch of philosophy of science or statistics. This simplification of the prescriptive theorist's task ought to be welcome, for it is no easy matter to gain agreement on betterness relations between paraworlds.

Later chapters discuss the problem of exigency in terms of the subjective costs to agents of conforming to particular rules or policies. Heavy subjective costs tend to disqualify policies, but counterweight principles tend to override agent's concerns about costs to themselves. Prescriptive moralists avail themselves of the argument from heavy costs to justify lower demand levels and counterweight principles to justify higher demand levels than an assumed set point. Acceptable moral rules need not be universal and can be relativized to particular reference classes. But prescriptive proposals, even if they arise from within particular cultural settings and reflect the concerns of creatures known to be partial to themselves and to kith and kin, presuppose a detached perspective. There is an anonymity requirement on moral theorizing, a distinct intellectual pursuit with its own methodology that is different from the activity of merely deciding what one is going to do. The requirement implies that the endorsement and propagation of norms that differentially serve the interests of the particular reference class that endorses and propagates the norm qualifies as ideology, not moral theory proper. For, in virtue of knowing that we have powerful interests in how things go for us as individuals, we know about ourselves that we are disposed to look for compromises between moral formulas of obligation and self- or class-interest and that we tend to seek exceptions to prima facie obligations in the form of exemptions and privileges. The anonymity requirement carries no implication to the effect that agents in our world exist in a state of empirical equality. Indeed, the chief reason for adopting it is the observation that they do not.

With this meta-ethical framework in place, I turn to a discussion of social equality. The presupposition in force is that conditions of social dominance in which some members of human societies have worse lives—less access to resources, more anxiety, less leisure, worse health—are rooted in our primate heritage and are exacerbated by technological progress. Human beings are inclined to coerce others and to take advantage of their labour when they are able to do so and both the descriptive and prescriptive sectors of moral theory must build on this assumption. My argument is that morality steps in where nature and the marketplace fail. The existence of moral practices and motivations, in other words, presupposes, not a condition of natural equality, as Hobbes imagined, but a condition of natural and acquired inequality, in which agents possess, temporarily or over the long term, natural or situational advantages, including superior strength, intelligence, knowledge, beauty, alliances, power, or wealth.

To wear down the intuition that moral agents exist in a state of natural equality, I employ two characters, A_1 and A_2 , who engage in various transactions. They are equals in their enjoyment of basic human goods and states, but one of them is *primus inter pares*. A_1 and A_2 cooperate for Hobbesian reasons—because conflict is expensive and they want to increase their productive capacity—but their decision to cooperate rather than compete does not make their relationship morally adequate. The initial moment of cooperation announces the beginning of their moral problems, as our interdependency has multiplied ours.

Where social dominance once depended on ferocity, charisma, birth, or alliance, alone or in combination with one another, modern institutions reward competence at specialized tasks with power. Presumed competence is associated in modern societies with the differential enjoyment of authority, prestige, and well-being. Some degree of variance in well-being produced by meritocratic sorting is, I try to show, defensible. Worlds that reward meritorious performance are better than similar worlds that allocate surplus resources according to other protocols. Yet existing distributive systems fall well short of what can be considered just. For, in the first place, large sectors of humanity do not participate in these meritocratic systems. Second, while the tendency of modern institutions to understand merit as specialist competence, rather than as ferocity, charisma, birth, or alliance, points to the role of moral influences that moderate crude advantage-taking, meritocratic systems can remain undermoralized. The modern market economy, and the relationships of employer and employee, investor

and worker, husband and wife to be found within it, represent the modification by degrees of the earliest urban societies founded on two principles: the agricultural, building, and craft labour of large numbers of slaves of both sexes, and the domestic labour of nearly all women. The increase in circulating wealth and in the organization of productive power has a seemingly intrinsic tendency to increase inequality between classes and nations, and between men and women. It is naive to maintain that observed high variance with respect to well-being is the product of a carefully contrived and well-monitored utilitarian plan to improve the status of the worst-off, and that it is simultaneously the by-product of a well-functioning merit-reward system. It is simply the condition we have inherited, modified, and succeeded in partially moralizing.

The last three chapters are concerned with the fair division of the cooperative surplus and focus on the question how much variance in well-being is morally tolerable. They are linked with the descriptive account of the earlier chapters by the premiss that to have a moral concern is to be willing to accept a reduction of advantage to benefit another, and by the premiss that theory choice cannot reflect one's actual situation. A morally good world, it is argued, exhibits variations in well-being at the margins—with respect to access to the doubtful and speculative, but not possession of the known and necessary components of well-being. Statistical equality of outcomes is further defended as the only plausible test of fair procedures. The last chapter returns to the sociobiological themes of the opening to consider male–female relations, including love, as morally relevant phenomena. The strengths and weaknesses of the argument from heavy costs, as it has been advanced in recent years against the demand for greater social equality between men and women, are assessed in a way that I hope will encourage philosophers and social theorists to investigate more thoroughly the relationship between the constraints allegedly imposed by nature in our actual world and our sense of what is morally right.

Acknowledgements

Work on this manuscript was supported by the Canadian Social Sciences and Humanities Research Council. Many colleagues have discussed its contents with me and offered important criticisms, including Sam Black and Virginia Held. Scott Anderson, David Braybrooke, Avi Cramer, Edward Halper, Cynthia Holder, Peter Vranas, and David Zimmerman commented acutely on individual sections, repairing many deficiencies and errors, as did a number of anonymous referees. David Donaldson and Mukesh Eswaran of the Department of Economics at the University of British Columbia, and Sarah Hrdy of the Department of Anthropology at the University of California, Davis have contributed generously of their expertise. I am especially grateful to Husain Sarkar, who read an entire draft, offering detailed suggestions for improvement, and to my students, Johnna Fisher and Tim Christie, for proofreading and pointed queries. Mohan Matthen, always quick to see the shape of a problem, has been a valued interlocutor throughout the writing.

Permission to reprint previously published material is acknowledged with thanks from the following sources: Kluwer Publishing for 'The Role of a Merit Principle in Distributive Justice', *Journal of Ethics*, 7 (2003), 1–38 and the University of Calgary Press for 'The Biological Basis and Ideational Superstructure of Morality', in Richmond Campbell and Bruce Hunter (eds.), *Naturalized Moral Epistemology*, *Canadian Journal of Philosophy suppl. vol.* (2000), 211–44.

This page intentionally left blank

Contents

1. Morality as a System of Advantage-Reducing Imperatives	I
1.1. A Platform for Human Morality	2
1.2. The Demarcation Problem	8
1.3. Are Advantage-Reducing Imperatives Natural?	16
1.4. Morality and Hypermorality	27
2. Paraworlds and Confirmation	35
2.1. Moral Belief-Sets and Theory <i>M</i>	36
2.2. Authored and Unauthored Norms	38
2.3. The Form of a Moral Theory	43
2.4. Moral Theories and Paraworlds	48
2.5. How Remote are Paraworlds from the Real World?	56
2.6. Relativism	62
3. Limits on Theory I: Costs to Agents	70
3.1. Exigency in Moral Theory	71
3.2. The First-Person Standpoint	77
3.3. The Argument from Heavy Costs	85
4. Limits on Theory II: Immanent Standpoints	95
4.1. Immanentism	97
4.2. Fast and Slow Paraworlds	101
4.3. The Disqualification Thesis	104
4.4. Opacity and Disqualification	113
4.5. In Defence of Theory	119
5. The Anonymity Requirement and Counterweight Principles	126
5.1. The Anonymity Requirement	127
5.2. The Partiality Exemption	142
5.3. From Theory to Practice	147
5.4. Counterweights to the Argument from Heavy Costs	151

6. The Division of the Cooperative Surplus	164
6.1. Is Variance a Moral Concern?	166
6.2. Procedural Theories of Justice	174
6.3. Basic and Symmetrical Cooperation	179
6.4. Equality	189
6.5. Justification and Consent	195
6.6. Immanentism and the Argument from Inevitability	202
7. The Role of a Merit Principle in Distributive Justice	212
7.1. Two Distributive Norms	213
7.2. Instrumental Considerations regarding Merit	228
7.3. Detecting and Assessing Merit	236
7.4. Objective Deprivation and Thresholds	242
7.5. Statistical Equality of Outcomes Required	248
8. Moral Equality and ‘Natural’ Subordination	254
8.1. Male–Female Relations in Moral Philosophy	255
8.2. Are Women Objectively Deprived?	257
8.3. Some Favoured Explanations for Female Subordination	266
8.4. The Argument from Heavy Costs	277
8.5. Recursive Effects of Social Judgement	283
8.6. Policies for Equality	288
8.7. Love as a Morally Relevant Phenomenon	293
<i>Bibliography</i>	302
<i>Index</i>	313

Morality as a System of Advantage-Reducing Imperatives

The theory of morals divides into the descriptive theory of moral phenomena and moral judgement and the prescriptive theory of what we ought, morally, to do. Before treating of moral judgements in the abstract and addressing particular questions of right and wrong, it will be useful to give some attention to moral phenomena, descriptively considered.

The concept of morality is related to but not coextensive with the concepts of care and mutual assistance, both of which have their place in the non-human world. Yet more seems to be required for moral observance than the occurrence of actions-that-benefit-another or actions-that-benefit-the-collective. The care of the crow or the wolf for its offspring benefits them, as the warning call of the goose benefits the flock, without being moral. Ants and bees live in family groups whose members must interact with one another to reproduce, to feed, and to defend themselves and their young, but their cooperation is no more moral than is the symbiosis of tree and vine.

Nevertheless, the altruistic and cooperative behaviour exhibited by social animals has a precursor relationship to human morality. The psychological platform that enables an animal to suppress or moderate its impulses—especially its aggressive and proprietary impulses—is a necessary underpinning for morality as we understand it. However sophisticated or complicated by conditions and exceptions our moral beliefs are, whatever rationale in terms of long-term happiness and general flourishing we provide for them, and however great the satisfactions of morally appropriate behaviour may be, moral emotions and practices involve some degree of repression. An appreciation of this fundamental point is important for progress in the

prescriptive sector of the theory of morals as well as in the descriptive sectors.

1.1. A Platform for Human Morality

Consider the behaviour observed in modern social primates such as baboons and chimpanzees.¹ These animals have distinctive personalities and recognize each other as individuals. They know who their children are even after they are grown, and they have friendships and enmities. Their behaviour is characterized by patterns of loyalty, reciprocity, and revenge for injury or betrayal. The animals form coalitions and may come to one another's defence, but they also refuse at times to assist each other when help would be useful. They compete with one another, chasing and biting each other, snatching each other's food, or displacing one another from desirable resting places. They also retaliate against such interference and attack strangers. Both males and females—but principally females—look after the welfare of infants; there is also occasional infanticide by males, as well as loss of infants through bad mothering or carelessness.

The animals take an interest in the condition of their own and each other's skin and hair. They take turns grooming each other and can treat each other's splinters and abscesses with some success.² Aggressive interactions between males and males and between males and females are triggered by feeding competition, or represent redirected aggression towards another animal. Sometimes one animal harasses or attacks another for no evident reason.³

In these animals, biological flourishing is compatible with and perhaps depends on a combination of benign neglect, help, especially where the effort may be repaid in the future, and harm, especially where successful retaliation is unlikely. They are neither primarily

¹ These details are drawn from M. R. A. Chance and Clifford L. Jolly, *Social Groups of Apes, Monkeys and Men*; Barbara Smuts, *Sex and Friendship in Baboons*; and Wolfgang Koehler, *The Mentality of Apes*.

² Koehler describes how a chimpanzee removed a splinter from Koehler's own finger 'by two very skillful, but somewhat painful squeezes with his fingernails; he then examined my hand again very closely, and let it fall, satisfied with his work'. *Mentality of Apes*, 321–2. Koehler observes further that '[if] one is on friendly and familiar terms with an ape who has been injured—say by a bite—one can easily induce the creature to extend the injured limb or surface for inspection, by making the expressive sounds which indicate sorrow and regret, both among us and among the chimpanzees'. *Ibid.*

³ Barbara Smuts, *Sex and Friendship in Baboons*, 90 ff.

selfish, nor primarily altruistic, neither exclusively partial to kin, nor indifferent to kinship relations. They are all these things, under different conditions, and different individuals possess these traits and dispositions in different measures. They react and respond to opportunities, threats, or changes in circumstances according to their present moods and temperaments, the perceived configuration of the situation, and their past relationships with others. What an animal does may not be the right solution to its immediate problem from the Darwinian point of view. The decision to stay and fight rather than to flee may result in death; the decision to mate now might result in its having no offspring who survive to maturity. Over the long run, however, the combination of personality traits and reactive habits, as these are distributed amongst individuals in an existing species, is adapted to the most frequent and the most critical situations they face.

There is little reason to ascribe moral beliefs or moral agency to animals that behave in this flexible manner. Only to the anthropomorphic eye are there paragons and reprobates amongst them. The animals cooperate—sometimes. Their cooperation is advantageous to them as individuals and to their kin—usually. They are aware of each other's needs, emotions, and intentions—to some extent. And human observers can easily develop affectionate relationships with individual animals. Yet there is no reason to call their animal society a moral one. This is not because the animals do not have language. For even if their behaviour was accompanied by verbalizations describing, sincerely or insincerely, their actions and intentions, this would not indicate that they had placed themselves under the particular restraints of morality. Nor is its absence explained by the animals' inability to ascribe mental states to others.

Missing from their orientation towards the social world is an interest in regulation as such. There is a certain kind of thought about themselves that the animals do not have, the thought that social interactions require the inhibition of spontaneous impulses, whether these impulses involve aggression or assistance. They may seek on occasion to control the social behaviour of others, breaking up fights or engaging in jealous interventions, and they may suppress their own reactions at times, but it cannot be said of them that they regard the whole field of social interactions as susceptible of moulding and

determination by themselves as agents. Analogously, it might be said that chimpanzees do not have aesthetic beliefs or engage in aesthetic practices, even if they draw or paint when given materials and opportunity, or decorate their bodies by draping them with ropes and branches. For they do not see the surfaces of the world—walls, containers, expanses of skin—as objectionably bare and as calling for remedial action.

When Hume⁴ traces the origins of morality to a natural disposition to perform just and benevolent actions, to approve just and benevolent actions in others, and to attribute merit to those who perform them, he expresses the view that morality is not exemplified simply in the performance of actions that happen to benefit others, but requires a social system that regards actions as items for judgement and criticism. A Humean might nevertheless protest against the claim that morality presupposes reflective awareness of social interaction as a field requiring control of natural tendencies by arguing as follows: Suppose we were to happen on a group of social creatures somewhat like humans who possessed speech and reason. Relationships between members of the group appeared to be friendly and affectionate, characterized by mutual assistance and devoid of the conflict, physical aggression, and psychological provocation for which primate societies, including human societies, are noted.

Suppose these creatures were articulate and explained to us that their benevolent actions flowed from their sympathetic identification with the needs of others. Would we not recognize this society as a moral one, even if its members were not conscious of any struggle to regulate their behaviour and that of others? The Kantian position is that there is no morality in this culture, in so far as its members act from inclination, not from a sense of duty. Nor would their acting from a sense of duty render them moral, according to Kant, if dutifulness were simply a special moral emotion unrelated to thoughts expressible as universal imperatives. As a culture might lack painting or theatre, and yet be attractive for other reasons—the extensiveness of its mathematical thought or its melodious songs—the one described lacks morality and is appealing for other reasons. Whatever the formal and substantive weaknesses of Kant's moral

⁴ David Hume, *A Treatise of Human Nature*, II. III. ii. 500.

theory may be—and I defer their consideration for later—his position captures the sense in which morality is an imposition that is not only an outgrowth and an expression of our natural dispositions but exists as a corrective to them.

Humans walk upright, talk, laugh, share food, care for their offspring for many years, and use their hands for constructive purposes including building, writing, drawing, and calculating. Their fondness for normative rules—for doing things in the right way, often in exactly the right way—is manifest in all their activities.⁵ Whether we are aware of them and can articulate them or not, our behaviour and our productions are constrained by internalized canons of appropriateness, decency, taste, and civility that forbid us certain actions that we could easily perform and that deem worthless certain products that we could easily fashion and display. Normative statements concerning what is fitting, good, meet, appropriate, and right to do are asserted, inculcated, followed, and enforced, and they are also scorned, ignored, contested, and evaded. Norms may be explicit and general; into this joint category fall the international codes of conduct pertaining to war, the actions by the commanders of ships on the high seas, and the agricultural regulations of large countries. Or they may be tacit and restricted, like the telephone protocols followed by a group of small-town teenagers or the haircut norms of a group of businessmen. They may be explicit and restricted or tacit and general; there are norms establishing what it is fitting to eat at different times of the day and on different holidays, what we talk about and what words we use, how we greet people, and how we manoeuvre our bodies through the world. We scan for infractions of the rules of fittingness and goodness, comment upon them, and punish them, even if the punishment is only adverse criticism and the rule-breaker is unaware that he is a subject of critical gossip.

The distinction between a species-specific behavioural regularity and a widely followed norm is imprecise. The habit of eating within an hour of arising in the morning and eating again at midday is partly a physiological requirement for active diurnal creatures, partly a convention. Exclusive pairings between males and females are natural

⁵ Allan Gibbard refers in this connection to our 'broad propensities to accept norms, engage in normative discussion, and to act, believe, and feel in ways that are somewhat guided by the norms one has accepted'. *Wise Choices, Apt Feelings*, 27.

for humans, but marriage is a norm that adds extra rigidity to the typical pattern. Deviance from a species-specific behavioural pattern tends, however, to reduce the biological fitness of a creature, through the working of what Bentham termed a natural sanction, while intentional or unintentional deviance from a norm may actually enhance biological fitness but has the potential to call forth a social sanction.⁶ Overeating shortens life and reduces reproductive opportunities, but bigamy might well increase both, though in many countries it is punished by law. The rules humans collectively invent and propound, and to which they try to hold others, extend beyond what is necessary either for biological survival or for the persistence and flourishing of communities. If wealthy businesspersons in Canada, unlike Italian aristocrats of a former era, eschew the wearing of ruby pendants, this is not because the practice is biologically dysfunctional or intrinsically disruptive.

The liking for norms and the pleasure taken in moulding thinking and acting so that it operates within constraints is evident in the great human interest in games, in which we take part cheerfully despite what is often a virtual certainty of losing. Economists like to present us as chiefly motivated by the desire to obtain preferred goods through the acquisition of exchangeable currency, but no rational person would accept the offer of a pile of gold on condition that he abstain from all normatively structured activity. Even those who enter lotteries in the vain hope of obtaining a pile of gold seem to take their chief pleasure in picking their numbers according to some system. Cognitively, we are equipped to follow rules, and affectively we are equipped to enjoy following them, and it is not fanciful to think that the ability to master phonological and grammatical systems is somehow connected with a broader facility with rules. Young animals play, and perhaps they use rudimentary rules or could be taught to use them, as some chimpanzees can be taught, with effort, to use sign language. Human children have a broader aptitude for learning new routines and seem to enjoy the constrained behaviour involved in dancing, singing, and drawing, as well as in talking. They grow up into such norm-governed activities as proving theorems, making

⁶ Jeremy Bentham, *Introduction to the Principles of Morals and Legislation*, 27 ff.

artistic representations, and creating and participating in elaborate bureaucracies and administrative hierarchies.

Rules encompass regulations, norms, idiosyncratic personal rituals, and social conventions. Any rule can be asserted in the imperative voice: ‘You! Do (not) x ’.⁷ The imperative form of any rule is convertible into a declarative form as a value judgement (‘It is wrong (unseemly, inappropriate, immoral, indecent, incorrect, illicit . . .) to/not to x ’). Not only do humans proclaim and observe rules, they reflect on them and theorize falsely and truly about their rules. They make certain assumptions regarding them—for example, how frequently certain rules are likely to be broken—and decisions about what to do about it when they are. Some rules are known to need strict enforcement, others are not; some rules are believed to apply universally, while others are believed to apply only to members of one community or class.⁸ The logic of rules is non-monotonic; rules admit of exceptions, and exceptions to rules admit of exceptions in turn. Nevertheless, almost all cultures believe that there are some rules that admit of no exceptions and that bind categorically.⁹ And they may give them supernatural or at least supramundane significance, insisting, for example, that certain rules were issued by a god, or are observed by an immaterial substance resident within us, or that infractions of important rules are automatically lethal for the rule-breaker, or shameful to his dead ancestors.

Formulas of obligation—statements of the form ‘I (you, he, she, it, one, we, they) ought to (should, must) do such-and-such’, uttered aloud, written down in books, implied or precisely articulated in public discourse—are expressions of social rules and are ubiquitous in both their hypothetical and their so-called categorical forms. Rules stating moral obligations are an interesting and problematic subclass. There is a greater tendency to regard moral norms and requisites as issuing from a transcendental source and as commanding universal human agreement in principle than there is to regard prudential and aesthetic rules as transcendental or universal. It is often said that moral

⁷ That moral rules are imperatives backed up by reasons was a major theme of R. M. Hare’s work; see *The Language of Morals*, ch. 1 and his retrospective *Sorting out Ethics*, 12 ff.

⁸ Robert B. Edgerton, *Rules, Exceptions and Social Order*, 221 ff.

⁹ *Ibid.* 254.

rules take precedence over other rules and other considerations. But which rules are moral rules?

1.2. The Demarcation Problem

The theoretical question what makes a given rule a moral rule—in virtue of what perceived properties are those who treat it as a moral rule doing so?—is different from the question whether anyone does or everyone should endorse the rule. We can agree that ‘Doctors should not assist their patients to commit suicide’ is a moral rule, rather than a rule of etiquette, even if we think it is a bad rule or that it ought to be disregarded under specific conditions. We can agree that ‘Protect your eyes when looking directly at the sun’ is a prudential, not a moral rule, even if we think that it is a good rule that all sighted creatures ought to obey. It is difficult, however, to specify the *topic* of moral rules, what they seek to regulate, in a way that is non-committal as between moral theories and that does not import prescriptive considerations into a descriptive task. Though we can sort rules into the categories of manners, dress codes, aesthetic guidelines, professional protocols, game-specific rules, practical injunctions, and moral imperatives, it is surprisingly difficult to articulate the criteria employed in making such discriminations. The demarcation problem is not solved by appeal to content. Both moral rules and taboos are largely concerned with prohibitions involving sex, killing, and kinship obligations. And certain conceptions of virtue or upright living are difficult to distinguish from specifications of elite manners.

It is sometimes said that moral rules are concerned with how to behave or how to live, but this specification is vague. To be told that morality contributes to human flourishing, or upright and decent living, is not to be informed. All rules—the rules of chess, the rules of warfare—instruct us about how to behave in various situations, and both aesthetic and prudential rules (Don’t mix plaids and stripes! Save your money! Wear a seatbelt!) tell us how to behave and how to live. There is a wealth of information available from decorators, psychologists, nutritionists, and government agencies on how to flourish as a human being. And to be told that morality is concerned with

minimizing suffering is to be misinformed. The injunction against using your hairdryer in the bathtub is not a moral rule, and the acceptance of a moral rule may even imply that more rather than less pain is morally meet or fitting; suicide and indifference can eliminate it. It might be suggested that moral rules in some way prescribe non-interference with others, but this does not differentiate them from certain rules of commerce and sport.

Asked to itemize their moral beliefs—the declaratives corresponding to moral imperatives—most respondents will produce a list of actions to be eschewed, including violations of contract, gratuitous cruelty, deception, fraud, injury, and insult. The current moral literature offers many examples of allegedly objective moral truths. Most of these examples concern the wrongness of harming animals or children, or torturing people, or engaging in genocide. Leaving aside for now the question of the truth status of judgements of wrongness, it is evident that these beliefs concern actions that are forbidden. Rules are plausibly seen, as David Braybrooke suggests, as ‘in origin physical blocking operations that prevent people from acting in ways prohibited, or, better, systems of blocking operations’.¹⁰ They are the verbal analogues of pinning someone’s arms behind his back. Moral rules tend to be formulated as ‘Thou shalt nots’. They mandate a sacrifice of opportunities for gratification; they deny a permission to act in a careless or indifferent way in pursuit of one’s self-interest. By extension, they may involve a sacrifice of opportunities deemed symbolically representative of gratification or regarded as likely precursors or empirical signs of the enjoyment of such gratification.

The high degree of confidence in the correctness of moral judgements relating to harm to children and animals can be explained by the supposition that they correspond to highly presentable samples of moral rules. This suggests the following semi-essentialist hypothesis: *Moral rules are restrictive and prohibitory rules whose social function is to counteract the short- or long-term advantage possessed by a naturally or situationally favoured subject.* A morality, in short, is a system of compensatory or advantage-reducing imperatives that correspond to moral judgements. It follows that a social rule that commands the harming of children or animals cannot be a moral rule, and that the

¹⁰ David Braybrooke, ‘The Representation of Rules in Logic and their Definition’, in Braybrooke (ed.), *Social Rules*, 3–20.

corresponding judgement cannot be a moral judgement. The claim ‘You should torture children if you gain satisfaction from doing so’ is not an example of a moral judgement that happens to be false. Whether the statement is false or lacking in truth value altogether, it is not a moral judgement at all.

Further, moral rules are those concerned with the adjustment of perceived situational balance of power. At the most basic level, they regulate aggression and the appropriation of goods; they protect the physically weaker members of the group against the strong and agile. Moral wrong is liable to occur wherever persons stand in relationships of unequal social power, whether the inequality is temporary or long-term, circumstantial or based in endowments. Without expropriating tangible property or inflicting visible corporeal damage, the powerful can influence our well-being by withholding information or encouraging false beliefs, by removing or failing to provide opportunities, or corrupting our relationships with others.¹¹ The duties considered to form the core elements of morality, to avoid interfering with people’s possessions, to refrain from exercising lethal force, to tell the truth, to keep promises and perform contracts, even when it would be easy and profitable not to, reduce the advantages of those who observe them. Even the duty to assume responsibility for oneself and to refrain from being a burden on others after sizing up their probable willingness to help falls under the proposed characterization.

‘Respect your contracts’, according to the hypothesis, is a moral rule that aims to prevent the stronger party from walking away from it because a contract no longer suits him. A_1 in observing the rule vis-à-vis A_2 makes things worse for herself by keeping to her side of the bargain. ‘Don’t steal’ prevents light-fingered A_1 from taking advantage of inattentive A_2 , though she loses what is perhaps a rare opportunity to gain thereby. ‘Eat no meat’ prohibits capable hunters or consumers from taking advantage of vulnerable edibles, at their own nutritional expense. ‘Take care of your own children’ prevents parents from leaving helpless infants to the kindness of strangers, even if the costs to the parents are heavy. ‘Share your wealth’ prevents tenacious A_1 from holding on to resources for life that needy A_2 does not have, though A_1 is thereby deprived of many pleasures. The

¹¹ See J. Harvey, *Civilized Oppression*, esp. chs. 3, ‘Having the Upper Hand’, and 4 ‘On the Receiving End’.

overall function of moral rules in the social economy is to serve as a brake, not just on our emotions or our inclinations, where the latter are viewed as non-rational velleities, but also—to some extent—on our intelligence, competence, and social forcefulness. It is for this reason that their alleged requirements are perceived as difficult and their justification as problematic. Moral rules are concerned with the regulation of actions that can broadly be described as *self-interested*, as aesthetic rules are concerned with the regulation of appearances, and prudential rules are concerned with maintaining health, wealth, and reputation. Moral wishes are just those wishes amongst all the regulatory wishes we have (such as the wish that more or fewer people would wear shorts) that are concerned with limiting the physical and psychological damage individuals can do to one another in pursuit of their own interests or the interests of their party, class, nation, or tribe. Harms resulting from negligence and indifference, as well as harms resulting from the desire to exploit or injure, can be understood as the effects of self-interest in this sense.

The most succinct attempt to characterize morality in the abstract is perhaps John Stuart Mill's discussion in the last chapter of *Utilitarianism*, and it is useful to hold his characterization up to the definition just sketched. Mill defined justice and injustice as notions pertaining to the upholding of legal rights, the award of goods and the imposition of evils according to desert, the maintenance of contracts, the avoidance of partiality, and the furtherance of equality except where expediency required inequality.¹² Contraventions of justice, he thought, involved 'two things—a wrong done, and some assignable person who is wronged'. To what he regarded as the mandatory *duties* of justice, Mill added the optional *virtues* of generosity and benevolence to make up the subject area of morality, succumbing to the temptation to mix descriptive and prescriptive considerations.¹³ If the distinction between duties and virtues is set aside, Mill's characterization might be understood as follows: morality prohibits certain wrong actions and prohibits inaction in the face of unfortunate states. The wrong actions concerned are not merely impractical or unaesthetic; they are typically actions performed with the intention of benefiting or maintaining the status of an advantaged party that exact some costs

¹² J. S. Mill, *Utilitarianism*, in *Collected Works*, x, 241–4.

¹³ *Ibid.* 247.

from or that fail to improve the status of a disadvantaged party; negligent actions, though not performed with the intention of harming or refusing help, are morally culpable when they betray self-interest.

The claim that moral rules are advantage-limiting or advantage-reducing is 'semi-essentialist': Moral rules occupy a sector of the normative realm, just as sofas and chairs occupy sectors of the category 'furniture'. As there are 'good' and 'less good' exemplars of sofas and chairs, as well as items that are intermediate between 'sofa' and 'chair', so there are good and less good exemplars of moral rules, as well as rules that are intermediate between prudential and moral rules, or rules of decorum and moral rules. It might be urged at this point that preventing wrong being done by an advantaged agent to another is not the unique aim of morality and that Mill's characterization too misses some of its central elements. The maintenance of personal dignity and integrity and the prevention of personal suffering, it might be insisted, are elements of the influential Stoic tradition and have every right to be considered as principal moral objectives. To meet the objection that the characterization offered is unduly narrow, I shall present some historical evidence that the reduction of advantage and the prevention of 'transitive' harms to the weak are ancient and universal features of what are agreed to be moral codes, and that other features are more or less peripheral.

Funerary inscriptions from the Egyptian Old Kingdom of the Third Millennium BCE provide some insight into the value systems of ancient people. One typical inscription praises the deceased for his or her refusal to kill, rob, commit adultery, trespass, execute ritual impurities, blaspheme, slander, cheat, and neglect the gods.¹⁴ Another cites the deceased's veracity, accuracy and fairness in speech, rescue of the weak, feeding and clothing of the hungry and naked, burial of the poor, furnishing of transportation, honouring and pleasing of parents, and assistance to widows, orphans, and lost strangers.¹⁵ Praiseworthy characteristics seem to fall naturally into two categories. Personal righteousness is exemplified in the failure to execute ritual impurities, blaspheme, or neglect the gods, and in the performance of parent-pleasing activities and observance of measured language. By contrast,

¹⁴ Scott N. Morschauser, 'The Ideological Basis for Social Justice/Responsibility in Ancient Egypt', in K. D. Irani and Morris Silver (eds.), *Social Justice in the Ancient World*, 106–7.

¹⁵ *Ibid.* 106.

the deceased's rescue and charity activities, as well as his or her restraint with respect to robbing and killing, seem to belong to another order of goodness that is specifically moral. Several ancient codices prescribe kindness to animals; opposition to circuses was even a feature of Stoicism. Other prescriptive texts from the ancient world describe an ideal condition in which no one stands to gain or to maintain an advantage through the deprivation or suffering of another.¹⁶

Or consider two familiar sets of norms, the commandments of the Old Testament and those of the New Testament. The Old Testament rules, the prohibitions on murder, swearing, adultery, and false witness, and the command to honour one's parents, appear to have little in common; personal-righteousness rules and power-restraining rules are bundled together. By contrast, many of the commandments of the New Testament offer variations on a single theme, the partial or total renunciation of advantage, or even the inversion of the relative advantage possessed in some situation by A_1 with respect to A_2 . The rules that one ought to divide one's cloak in two and give half to the beggar, love one's enemies, and respond to aggression by turning the other cheek, are exemplary moral rules. The New Testament is accordingly a source of excruciatingly, even perversely advantage-reducing imperatives.¹⁷ By contrast, the Old Testament imperatives are a mixed bundle: the prohibitions against murder and false witness are clearly advantage-reducing, but the other commandments appear to be composites in which morality, taboo, and status considerations are mingled in the formulation of the rule.

The hypothesis nevertheless faces several criticisms:

First, it might be objected that many moral rules are not compensatory or advantage-reducing, and many advantage-reducing or compensatory rules are not moral. Under the first category, one might propose such widely accepted norms as 'Debtors should repay their debts' or 'Talent should be recognized and rewarded'. The repayment rule seems to take further from the weaker party and the reward rule to give further to the advantaged party.

One response to this objection is that the cited rules can but do not always function as moral rules. Repayment and reward rules may

¹⁶ K. D. Irani, 'The Idea of Social Justice in the Ancient World', in Irani and Silver (eds.), *Social Justice*, 5.

¹⁷ v. Friedrich Nietzsche, *The Genealogy of Morals*, tr. Walter Kaufmann and R. J. Hollingdale, 34.

function as pragmatic rules that are beneficially observed in societies that attach great importance to converting natural resources into commodities for human use. They may resemble in this regard the rule ‘Interest rates ought to be raised to control inflation’. The latter is usually a good rule for keeping economies on track, but it is not a moral rule. Under other interpretations, the repayment rule and the reward rule can be construed as having moral content. The first may be understood as an instance of the ‘Keep your contracts’ rule that prohibits disaffected contractors from walking out whenever they can do so with impunity. In so far as a given debtor has the funds or a reasonable prospect of acquiring them and can repay the money, he is in the advantaged position and should make restitution. The reward rule may be understood as requiring that sacrifices not be in vain. In so far as a meritorious person has endured hardship, he should be compensated. ‘Share your candy with your friends’ may function as a moral injunction not to tolerate the relative deprivation of others, or it may be a prudential recommendation for achieving popularity. Nothing precludes a given social rule’s having a dual significance.

A second objection to the hypothesis is that there can be no cross-culturally valid characterization of a moral rule. Different cultures, it might be argued, group their rules governing social conduct and personal dignity together in various ways, and the designation of some subset of them as that culture’s moral rules must follow the culture’s own discursive practices. To designate a rule as moral is to single it out as an especially important personal conduct rule, and it is not up to us to say whether the naming taboos or clothing regulations of another culture are genuinely moral or belong in the same category with prohibitions on theft and murder. A culture might prescribe a set of what its members designate as ‘E-rules’ that enjoin fidelity to promises and generosity towards friends, and that require revenge for all insults, as well as the adoption of a distinctive mode of a dress associated with special personal dignity and authority such as the wearing of white shifts and the carrying of a small ceremonial knife. There is no fact of the matter, according to the critic, about whether these are all moral rules or not, and the decision whether to translate the foreign term ‘E-rule’ as ‘moral rule’ ought not to depend on the similarity or dissimilarity of the E-rules to some prototype in the mind of the translator.

The objector may point out that even in our own culture the term 'moral' is used in a broad sense and suggest that it is unacceptably revisionary to propose that it refers principally or centrally to rules concerned with advantage reduction. Some moral imperatives, she will insist, forbid an agent to debase himself or waste his talents, or bid him refrain from taking into his body or his mind substances, thoughts, or images of an impure or polluting nature. Kant's rules that one should not use other people as playthings even when they are agreeable to it, nor allow oneself to be so used, nor sell parts of one's body such as one's teeth for profit fall into this category. Many people consider recreational drug-taking and bioengineering to pose, each in their own way, serious moral problems. Yet the alleged wrongdoings of the weekend hallucinator or the professional cloner do not lend themselves easily to our analysis of moral rules as advantage-reducing. Revisionary definitions, the objection might continue, may be called for in the exact sciences, but they have no place in philosophical inquiry, which must be concerned with the common understanding of terms. If a significant number of people describe cloning, and other alterations of organic bodies, as moral issues—not merely a set of prudential, aesthetic, etc., issues, or as expressions of a worry about the taboo status of simulacra, impure hybrid 'mixtures', or artificial life and experience—they must really be moral issues.

One way to meet this objection is to insist that those who assert that cloning is immoral are speaking or writing in an unusual dialect. Prohibitions on pornography-consumption, drug-taking, or gene-splicing, it might be further argued, do not constitute moral rules; they are assignable to the neighbouring category of restraining usage taboos applying to objects belonging, or in this case conceived as belonging, to a sovereign entity, oneself, or perhaps God. Another way of meeting the objection is to point out that some notions of duties to oneself, or perhaps even to nature, have the proposed moral marker to some degree.¹⁸ The belief that the consumption of pornography is contrary to morality is doubtless influenced by non-moral ideas about dignity and integrity, some of them superficial, others arguably profound. Yet it may also be rooted in the idea that the

¹⁸ Freud describes such taboos in 'Taboo and the Ambivalence of Emotions', in *Basic Writings*, 828 ff. 'To touch is the beginning of every act of possession, of every attempt to make use of a person or thing.' *Ibid.* 833.

activity involves the exercise or maintenance of situational advantage by one or another party. The sense that the vending of surgically extracted organs is contrary to morality may be based on non-moral ideas regarding the unseemliness of contact between personal and foreign body parts; again these may be foolish worries or not. This impression may also however reflect worries about the temptation to victimize helpless or needy persons that the profitable vending of organs would awaken. By contrast, it may be difficult to see the individual addict or the addicted sector of the population as members of an advantaged class; only if such persons are conceived as escaping ordinary demands and responsibilities and as letting others down can any moral significance be attached to their actions.

The intuition that drug-taking and cloning are activities with moral implications has another basis as well. For most moralists who are concerned with them are worried about the effects of these activities on others besides their perpetrators who may be affected by them. They are simultaneously worried about whether they would be, *ex officio*, harming anyone in arguing for their prohibition, and about the justifiability of interfering forcibly with other people's preferred activities when one is in a position to do so. There is an implicitly dyadic structure in most moral controversies involving what might at first appear to be purely self-regarding activities.

The semi-essentialist need have no objection to including rules mandating, say, women's haircovering, as moral rules, provided they are not taken without further explanation to be examples of central or focal moral rules and provided their advantage-reducing feature can be made apparent. The position that the *best* examples of moral rules we possess are haircovering rules, anti-intoxication rules, and rules proscribing the making of impure mixtures, such as tomatoes with the genes of fish, has little to recommend it.

1.3. Are Advantage-Reducing Imperatives Natural?

Lucretius regarded the first humans as fiercely amoral individualists: 'They could have no thought of the common good, no notion of the mutual restraint of morals and laws. The individual, taught only to live and fend for himself, carried off, on his own account such prey as

fortune brought him.’¹⁹ These creatures mellowed, he thought, into members of a community. ‘[N]eighbors began to form mutual alliances, wishing neither to do nor to suffer violence among themselves. They appealed on behalf of their children and womanfolk, pointing out with gestures and inarticulate cries that it is right for everyone to pity the weak.’²⁰ From our current perspective, Aristotle was wrong about the autochthonous status of moral and political institutions, right about the social propensities of human beings. Lucretius, in turn, was wrong about the solitary and amoral nature of the first men, but perhaps right in viewing morality as a form of mollification.

Working in the Lucretian tradition, theorists have produced quasi-anthropological accounts of the origins of morality that describe the passage from advantage-seeking to altruistic behaviour, emphasizing the roles of reason, fear, and pity in the transition. Hobbes’s account in Part One of his *Leviathan*²¹ is perhaps the most celebrated use of a naturalistic platform, and it furnishes a model for contemporary accounts based on the Lucretian assumption of a pre-existing state of war or mutual indifference. There are several ways to interpret Hobbes’s story. Historians read it as a novel defence of absolute monarchy. Game theorists read it as an account of the discovery of the rationality of interpersonal cooperation. Whether or not a special relationship between monarch and subjects is conceived as its necessary condition, the realization that the cessation of mutual hostilities is the better strategy for those who want to maximize their happiness and security announces the initial moment of moral reflection. ‘Hobbes’s Theorem’, as it might be called, is that the addition of moral regulation to a world increases its hedonic content, the amount of pleasure, comfort, and happiness it contains. And although the theorem does not appear to be true for arbitrarily large increments of morality, it is clearly true for increments up to some level.

Hobbes’s notion that cooperation reduces deprivations and produces a surplus of human good is mirrored in contemporary accounts that substitute the cunning of nature, or blind natural selection under conditions of environmental scarcity, for strategies consciously chosen in a state of anxious competition. From the perspective of evolutionary theory, the members of a single species are distinct

¹⁹ Lucretius, *On the Nature of Things*, v. 958 ff.

²⁰ *Ibid.* 1018 ff.

²¹ Thomas Hobbes, *Leviathan*, 104 ff.

individuals engaged in a competition for reproductive success.²² The persistence of a trait is like the rational solution to a problem posed by the ambient environment, which may include the presence of conspecifics who share the trait, in so far as traits that are not solving the problem tend to be extinguished by variant traits that solve it better. The disposition to behave morally, to be vigilant about moral infractions in others, and to punish them, could then be explained as an evolutionarily stable strategy, an *ESS*, if we suppose that humans lacking these traits fare poorly in the struggle for existence.²³ A trait it would be beneficial to evolve might seem equivalent to a policy it would be rational to choose.

The view that human morality is simply an *ESS* is tempting but clearly inadequate. Hobbes's notion that a mutual non-aggression pact will be rational for all his warriors to agree to rests on his assumption that they are all equal in the degree of force they can individually exercise and desire to exercise.

Nature has made men so equal in the faculties of the body and mind as that, though there be found one man sometimes manifestly stronger in body or of quicker mind than another, yet, when all is reckoned together, the difference between man and man is not so considerable as that one man can thereupon claim to himself any benefit to which another may pretend as well as. For as to the strength of body, the weakest has strength enough to kill the strongest, either by secret machination or by confederacy with others that are in the same danger with himself.²⁴

The equality assumption is basic to the ordinary sociobiological framework. Natural selection reduces the frequency of deleterious traits and increases the frequency of advantageous traits. It is a uniformity-inducing process, offset by the tendency to variation.

This is not to say that *ESSs* require homogeneity down to the level of individuals; Maynard Smith has pointed out that they may be instantiated in distributions of traits in polymorphous populations,

²² This is oversimplified, but if an individual organism is considered for the purposes of the discussion as the bearer of a trait, no harm is done.

²³ If I is an *ESS*, then 'if almost all members of a population adopt I, . . . the fitness of these typical members is greater than that of a possible mutant'. John Maynard Smith, *Evolution and the Theory of Games*, 14.

²⁴ Hobbes, *Leviathan*, 104–5.

and Robert Trivers takes stable strategies to correspond to the frequency with which individual members display certain forms of behaviour. 'Hawkish' and 'dovish' tendencies can be modelled in terms of patterns that are resistant to self-extermination and invasion alike.²⁵ This point is important, for in any real-world population of social animals, some are cleverer, or stronger, or more ruthless, or more attractive than others; others are correspondingly dimmer, weaker, gentler, less charismatic, and less able to form alliances. The latter are not 'less fit'. Their own hidden-from-view mosaics of physical and psychological traits serve them just as well in the struggle for existence, though not all their genes will appear with the same frequency in later generations. Yet the former can dominate the latter.

Dominance cannot confer a heritable selective advantage on an animal exercising it, for the advantaged trait would spread through the population and there would be no animals to submit.²⁶ It could be 'accidental': A certain distribution of individual traits within a group may be stable without its being the case that any one of the polymorphisms confers an advantage. Some people have exceptionally long, slender fingers, but, as this trait has not as far as we know increased its frequency, it cannot be supposed to confer a special selective advantage, and dominance might be a trait of this sort. More plausibly, dominance and submission correspond to instructions animals heed when responding to other animals that happen to be larger or smaller, fiercer or more pacific than themselves; they may be relational strategies for getting along in such mixed environments.

Assume a population that is linearly ordered with respect to size and ferocity. Then the rule 'Always defer to a larger animal and seize resources from a smaller animal' is an *ESS* that will sort a population into a dominance hierarchy. Provided the animals do not encounter one another so frequently that the smaller fail to survive and reproduce, and provided conditions of great scarcity

²⁵ Maynard Smith, *Evolution and the Theory of Games*, 16–17; Robert Trivers, 'The Evolution of Reciprocal Altruism'.

²⁶ Dominance is relatively easy to explain on the hypothesis of group selection; Trivers suggests that it prevents mass extermination in times of food shortage. This explanation might be reconciled with classical Darwinism on the assumption that members of strongly hierarchical social groups tend to be closely related to one another.

do not intervene, there is no reason such a population of unequals cannot persist indefinitely. The existence of rigid social hierarchies is consistent with the uniformity-inducing tendencies of natural selection and, accordingly, with the artificial, reason-driven evolution of social norms. According to the hypothesis above, the domination of individuals by other individuals, by coalitions, and by institutions is typically what morality seeks to prevent. Stable situations and procedures that have arisen through the historical interactions of individuals cannot therefore be assumed to meet the tests of moral adequacy that culture has at the same time produced.

The metaphysical postulate of the moral equality of human beings furnishes a striking contrast to the variance observed in the degree of social power they exercise. Natural equality is a reality in the following sense: from the knowledge that an animal is dominant in its social group, or a person in hers, we cannot infer that it or she is more intelligent, healthier, more resistant to disease, a superior parent, or that its or her genes are better. Each of us incorporates a mosaic of traits that the rigours of the environment have failed thus far to eliminate, and it makes no sense to describe one existing person as a better product of evolution than another even though not all of everyone's genes will maintain their current frequency in future populations.

The 'fitness' of the biologist therefore has only the sketchiest relationship to the ordinary-language notion of fitness as strength, good looks, etc.²⁷ Nevertheless some humans—whose genes may be slated for extinction—are just now a great deal more successful than others when it comes to attracting admiring attention to their persons, displacing others and appropriating their resources, and collecting a disproportionate number of suitors, mates, or followers. Some people give way easily, are content with modest amounts of everything, do not seek to influence the behaviour of others or recruit them as clients, and try to stay out of the limelight. To what extent should this variance be considered 'natural?' We do not know and cannot easily determine what the original social system of humans was where dominance and subordination are concerned. The primate literature emphasizes the variety of social systems that are to be found

²⁷ On vernacular v. predictive fitness, see Mohan Matthen and André Ariew, 'Two Ways of Thinking about Fitness and Natural Selection'.

in the 200 or so species studied.²⁸ Some species whose habits are believed to be close to those of early man, such as the savannah baboon, are strongly hierarchical, though the closest primate relatives of humans are not agreed to be so.²⁹

The problem of dominance behaviour—the disposition to push others around, to appropriate their resources, injure their offspring, and interfere with their lives—is actual and cannot be considered to have been solved by the evolution of cooperation, by pretheoretical agreement amongst persons equally capable of exercising social force. Yet human morality is a system of norms that limit the expression of dominance. It is made possible by the human capacity for generalization and abstraction, and the capacity to conform to rules that subjects can learn, internalize, and teach. But why do we have such a system? Why do we try, in the name of morality, to reduce and compensate for natural inequalities that permit dominant individuals to accumulate advantages at the expense of weaker ones? Evolutionary ethics does not give a satisfactory answer to this question.

Lucretius and Rousseau, it might be noted, deploy a somewhat more complex model than does Hobbes in their accounts of the evolution of morality. They too posit a renunciation of aggression for mutual benefit, but Lucretius is the first to insist on the importance of pity, an emotion felt by the strong towards the weak, as a moral emotion. Perhaps he sees the appearance of pity as symptomatic of the softening of human ferocity that cooperative behaviour and the production of a cooperative surplus bring in their wake. Or perhaps he sees pity as an indispensable motivator for the rational decision to cease hostilities. Even if these two conceptions, the rational and the sentimental, can be grasped as mutually reinforcing, they are separable. Cooperation between two strong animals can increase their joint advantage. To be sure, coalitions of the weak can be dangerous to the strong, and cooperation with the weak can be beneficial to the strong. Nevertheless, when the strong act from fear, or to increase the advantage they obtain from the weak, they act from an altogether different motive than when they act out of a concern for justice or for the welfare of the weak. The appearance of pacific and even benevolent relations may arise in either case, but it

²⁸ Joan B. Silk, 'Primatological Perspectives on Gender Hierarchies'.

²⁹ Peter van Sommers, *The Biology of Behaviour*, 151.

is only the sentimental motive that can be termed moral. The concessions made to an underclass by a set of revolution-fearing oligarchs are not moral concessions.

Moral rules accordingly have two puzzling features. First, they forbid us to use our intelligence to analyse a situation with an eye to determining what could be done for our own advantage, given the weakness or unpreparedness of others, and to muster whatever social power we antecedently possess to serve our self-interest. The formula of obligation 'Other humans ought not be killed' is a blocking rule, a prohibition that instructs A_1 not to kill A_2 , even when A_1 is irritated by A_2 's presence, when A_1 is stronger or wiler than A_2 and could succeed in doing so with impunity, and when A_1 would have a better life were A_2 out of the picture. Second, they imply the suppression of our well-honed discriminatory abilities. The formula of obligation 'Care for your existing children' impresses on us that we ought not to abandon even one of them, even if, after rationally sizing up matters, we realize that we could raise more children in the long run by leaving one particularly troublesome one exposed to the elements. The formula implies that we should care for our children, whoever they are, and not increase our personal consumption by depriving any of them of necessities, or fob off their care on others in an exploitative manner.

Hans Kummer observes that moral formulas are characterized by their wide scope. 'As regards killing, respect of possession, or false information, they tend to prescribe the same course of conduct in (nearly) all situations and toward (almost) all conspecifics. Advanced codes include nongroup members, alien races, and even all animate beings among the favoured.'³⁰ As Edward Westermarck pointed out, in tribal societies and in the ancient world, the stranger was regarded as someone to whom the concepts of the sanctity of life and property did not apply, or did not apply as strictly as the prohibitions against harming fellow citizens.³¹ The universal or 'overgeneralized' formula rarely makes an appearance outside literate societies in which it is easy to issue broad proclamations.

³⁰ Kummer, 'Analogies of Morality among Nonhuman Primates', 43. As Jack Goody remarks, 'Enshrined in the written word, passed down from century to century, the generalized, decontextualized statement becomes the touchstone of moral rationality. It implies that all men should be treated in the same way, that status, relationship, age, and sex are irrelevant in making judgments about the conduct of mankind.' 'Literacy and Moral Rationality', 161.

³¹ Westermarck, *Ethical Relativity*, 199–200.

Consider a staggered set of policies *D*, *H*, *L*, and *Q*. Each corresponds to a set plan of action and reaction that might correspond to the overall policy of a hypothetical organism. *D* below corresponds to a fully rational strategy, one an animal should follow if it is aware of its own attributes and the attributes of its fellows and has no interest in them other than as a means to an end, the end being its own survival and maximal reproduction.

D: Act always to maintain or promote your own interests, e.g., by consuming all health-enhancing resources, harming competitors, and removing obstructions to your reproductive success.

D may very well incorporate some limiting clauses. It does not command an animal to consume all possible resources (it might die of a surfeit), or to eliminate all competitors (this might be self-defeating), or to produce as many offspring as possible (in that case, none might be viable). But programmes such as ‘Kill occasionally’ or ‘Kill very troublesome individuals’ or ‘Kill if you can do so without detection’ can enhance an individual’s chances of survival, and may be part of a given species’ typical repertoire, expressing themselves in response to certain types of cueing. *D* allows for altruistic actions, so long as there is a net pay-off to the altruistic agent. It tells the animal to consume, harm, and reproduce only to the extent that this is biologically advantageous for it. An animal set to operate according to *D* need not get stuck in a Prisoner’s Dilemma: It can simply make a guess, however wild, about what its partner is likely to do and act accordingly.

Policy *H*, however, restricts the creature further:

H: Act always to maintain or promote your own interests, according to formula *D*, except when you cause substantial harm to others.

H leaves it open to what extent the organism may harm others to pursue large or small gains for itself. But, unlike *D*, it is recognizable as a moral norm in restraining the unlimited pursuit of self-interest.

H is not a rational policy for a purely self-interested being to adhere to. For if some prohibition against harming another actually serves my interests, it is already provided for under *D*. How then could even a weakly limiting policy like *H*, one that a purely rational

self-interested being would not accept, take hold in a species like ours? The moral ascent from *D* to *H* is a small step, but one nevertheless difficult to explain. For, since *D* and *H* really enjoin different patterns of behaviour, it would seem that a *D*-following individual can always successfully invade a population of *H*-followers, and that *H*-followers will fare poorly as intruders in a pre-existing society of *D*-followers.

To be sure, behaviour that formerly increased biological fitness in the technical sense—murderous retaliation, infanticide—may no longer be tolerated once *H* has taken hold, and may decrease the individual's chances of survival and reproduction. What was formerly an *H*-policy is now a *D*-policy. But how did *H* take hold in the first place?

Kummer characterizes advantage-reducing rules as a 'reversion to nonopportunism', noting that this reversion seems difficult for a naturalist to explain, since evolution would seem to favour the development and use of our intelligence and behavioural flexibility. Why, he asks, does nature 'allow a superimposed cultural development to undo just that achievement and, as it were, regress to rigid behavioural rules'?³² Kummer's answer is that the exercise of too much practical intelligence renders social animals too dangerous to one another. It brings about a situation in which, as Hobbes suspected, too many resources are wasted monitoring the behaviour of conspecifics, detecting their stratagems, and trying to defend oneself by planning one's next pre-emptive strike:

Man as a hunter, forager, or toolmaker can hardly be too clever or versatile; he will always benefit from more foresight and a greater arsenal of selfish schemes. But man as a social companion can be too astute even in his own interest. Every being can become more dangerous simply by becoming less predictable, regardless of whether it appears in the role of an enemy, a competitor, or a cooperator. Cooperation requires that each participant be able to predict the other's actions. . . . Thus behaving in a predictable way increases the benefits and alleviates the risks of social life. If we assume that man's preadaptations offered no way to evolve a brain that was shrewd with tools, predators, and prey, but simple and predictable in dealing with his companions, the evolution of moral capacities might have been the adaptive answer: a selective suppression of shrewd flexibility in the social context.³³

³² Hans Kummer, 'Analogues of Morality', 43–4.

³³ *Ibid.*

Compensatory mechanisms are a common feature of biological systems; nature likes to build layers of controllers and releasers that regulate the behaviour of a system at a lower level. The original problem still remains, however. If the 'reversion' to non-opportunism is still in accord with policy *D*, the ascent to *H* has not been explained. If non-opportunism involves an ascent to *H*, the invasion problem has not been solved. It is still unclear how a more predictable, less dangerous animal's genes can invade a population of social animals whose behaviour is a function of the four variables discussed above, and essentially improvisatory. Just how can the rigid type drive the more flexible type to extinction?

One possibility is that differing cultures in separated groups can develop by chance and thrive in differing proportions, just as different physical characteristics can arise by chance in non-interbreeding subpopulations of a single species. Group selection could occur without violating basic Darwinian principles if by-chance rigid populations of related animals left more offspring than by-chance flexible populations.³⁴ Alternatively, rigidity and moderation of advantage-taking dispositions might be an individual defence strategy evolved in response to punishment and reward by other animals. Robert Trivers notes that punishment of 'cheaters' sometimes appears to be out of proportion to their offences, but that 'since even small inequities repeated many times over a lifetime may exact a heavy toll in inclusive fitness, selection may favour a strong show of aggression when the cheating tendency is discovered'.³⁵ An emotional commitment to justice, rather than maximization of immediate pay-off, may be a good long-term strategy.

It is inconceivable that natural selection alone is responsible for moral phenomena such as the existence in some minds of a belief in universal human rights. There is simply no plausible account on which the existence of such a belief, or its necessary conditions and accompaniments, could have increased the number of offspring its possessor left.³⁶ Nevertheless, once the process of moral ascent has got off the ground by taking hold in some minds, further elaborations

³⁴ As envisioned by Elliott Sober and David Sloan Wilson in *Unto Others: The Evolution and Psychology of Unselfish Behavior*, ch. 4, 132 ff.

³⁵ Robert Trivers, *Social Evolution*, 388.

³⁶ This is agreed to by Thomas Huxley, Richard Dawkins, and G. C. Williams; citations can be found in Flack and de Waal, 'Any Animal Whatever', 1-2.

through cultural innovation and transmission may come readily. A few steps up from the vague and somewhat weak policy *H*, we arrive at the quite strenuous and precise Lockean proviso *L*:

L: Act always to maintain or promote your own interests, according to formula *D*, so long as you do not thereby worsen another's situation, except to avoid worsening your own.

According to David Gauthier, the proviso transforms the Hobbesian state of nature into the 'productive natural condition' of Locke.³⁷ It is rational on his view for all self-interested beings to accept it, whether or not they are moved by sympathy or fellow feeling. This follows, however, only on the Hobbesian assumption of natural equality. Consider a group of rational self-interested beings devoid of other motives who differ in their individual endowments of size and ferocity, and who are aware of their own attributes and the attributes of the others. For the strong, in this case, the weak are among the natural resources available to them; for the weak, the strong are natural predators they must evade. Suppose the group has the opportunity to write down one fundamental rule that will license and constrain their behaviour. It is not rational for anyone in the group to agree to *L*.

Depending on how *L* is interpreted, both the strong and weak have reasons for refusing to endorse it. The strong may believe correctly that the weak are unable to make things worse for them. They will complain that the policy benefits only the weak and that it is not in their rational self-interest to subscribe to it. The weak, however, may interpret it as favouring the strong. For suppose their only chance to acquire some essential good is by taking it from the strong. They are prohibited from doing so if this would worsen the condition of the strong, even if the need of the weak for the good is substantial. Suppose that the strong control the land, enjoying more of its resources, and make a small profit from their ownership of a well. The weak cannot ask for more water at a lower price if their condition, already bad, is not worsened if they do not acquire it. The weak have no reason to endorse *L* if it is their last chance to establish a basic policy.

If there is no policy that both the strong and the weak must agree on when the strong regard the weak as environmental resources for

³⁷ David Gauthier, *Morals by Agreement*, 208. The proviso was discussed earlier by Robert Nozick, *Anarchy, State and Utopia*, 175 ff.

their use, then, in so far as morality is concerned with the regulation of harm, morals cannot spring from the foundation of rational self-interest by agreement. This is not to say that *L* might not be an important policy rule that is widely accepted for the regulation of certain types of behaviour in contexts where there is agreement that failing to better someone's situation is tantamount to worsening it, or where the strong are uncertain of their hold over the weak, or even where predation is considered a salutary process. *L* might govern commercial transactions, for instance, or relationships that fall within the purview of a law offering equal protection.

Continuing with the process, we can go on to write a series of even more limiting—and ever less rational, or universally appealing—policies, arriving finally at *Q*:

Q: Act always to maintain or promote your own interests, according to formula *D*, unless by your doing so another's condition is rendered worse than it would be by your not so acting, or not improved.

Q forbids me even maintaining my present advantages if I can assist someone else by renouncing them.

The adoption of a *Q*-policy will likely militate against the practitioner's biological success as well as his enjoyment of consumption opportunities. 'Kill no one', 'Do not eat animals', and 'Never lie' may represent fatal strategies for individuals. By eating nothing that had ever been alive except fruit fallen from the tree, one might compromise one's reproductive health. By refusing to practise infanticide and continuing to feed infants during a short period of harsh conditions, one might exterminate one's whole tribe. However, even profoundly inhibited behaviour is compatible with survival and the most exigent systems can sustain themselves. *Q*-policies can flourish when they are restricted to a subculture and new recruits can be attracted to them, or when they gain such currency that there are few threats to them from selfish invaders.

1.4. Morality and Hypermorality

Policies *D–Q* are meta-rules, not rules of morality. They correspond to programmes that might be imagined as governing hypothetical

organisms. Morality intensity increases as we move up the register from *D* to *Q*. Morality is importantly scalar, and this feature can be seen from a comparison of the intensity levels of various degrees of prohibition pertaining to certain categories of action.

We might think in this regard of human behaviour as a sphere pierced by numerous axes, each marked by positions located at points some distance out from the centre.³⁸ At the centre of the sphere, behaviour is plastic and situationally elicited, subject only to a few inhibitions. There is altruistic care as well as harassment, cooperation as well as competition. We cannot write down the specific patterns of behaviour that characterize this centre, because we cannot return to the moment when it began to be replaced or overpainted by post-Darwinian limiting rules and restraints. Nevertheless, we have a general idea what we would write down if we were asked to articulate, as a set of formulas of obligation or, in Braybrooke's terms, 'blocking rules', the rules at the core falling under policy *D*. We would write down some prohibitions against incest, against eating certain unclean or poisonous foods, and against killing one's own children. Some distance out, where our creatures begin to think and write their own rules, we can see these prohibitions undergoing a process of generalization. A prohibition against killing kin, which is fitness-reducing, may be elaborated into a rule against killing unrelated tribespersons, consistent with policy *H*. A prohibition against eating rotten food, which is unwholesome, may be extended into a rule against eating totem animals. A prohibition against interfering in established consortships, which is dangerous, may be extended into a rule asserting the sanctity of marriage consistent with policy *L*. These prohibition extensions need not be functional, as their predecessor prohibitions were. Their arrival on the scene signals that humans now see social behaviour as a field for control and for decorative elaboration. Compensatory rules can be pushed to extremes that can be described as 'hypermoral'.³⁹

Warlike communities that keep the peace amongst themselves but fight their close neighbours occupy a position close to the centre on

³⁸ The notion that morality is scalar has been expressed by Michael Slote, though in connection with a version of consequentialism. See his *Common-Sense Morality and Consequentialism*, 80 ff.

³⁹ The term 'hypermoral' was coined by the German sociologist Arnold Gehlen in *Moral und Hypermoral*, 146 ff. Gehlen held Nietzschean views that are not under consideration here.

the external aggression axis; 'just war' theorists are positioned somewhat further away, and pacifists a good deal further away, at what we might call the 'hypermoral edge'. Practitioners of infanticide under conditions of social stress or upheaval are near the centre, those who would never consider undergoing an abortion are further away, and those who abjure contraception on the grounds that it injures life are on the hypermoral edge. Omnivorous humans are close to the centre, tribes that avoid eating their totem animal are some distance out, and those sects whose veneration for life extends so far that they eat only fallen fruits are on the periphery. Political groups that deny the appropriateness of any distribution from families to strangers belong near the centre; communists are at the edge. A hypermoral position is one that implies a profound renunciation of advantage, a refusal to be pragmatic, or to deploy all the resources one has at hand, whether natural or cultural, native or acquired, to come out ahead, or even to equalize an unequal situation. The claims that one should turn the other cheek, that evil should be repaid with good, that it is always better to suffer than to do evil, all represent hypermoral positions. Their proponent flatly rejects the suggestion that he should employ certain of his capabilities to worsen the situation of another. The seeming obtuseness of the pacifist in her indifference to the appeal to be flexible and appreciative of varying circumstance is a token of her hypermoral commitment.

Is there any historical evidence, one might wonder, for the hypothesis that the development of moral ideation and its associated discourse is linked with a reduction in the combination of flexibility, pragmatism, and fierceness displayed by our ancestors? Both Plato and Aristotle appear to regard the establishment of appropriate dominance hierarchies as fulfilling the requirements of justice. Nevertheless, they are concerned that the sorting-out of the population be accomplished by some means other than the ascension of the physically stronger or of charismatic 'new men'; the more reasonable, or intelligent, or the well-born, they think, should rule the others, and this notion displays the elements of moral thinking in politics.

Readers of the chronicles of ancient history are struck by the emotional unpredictability of powerful rulers, their devotion to the 'cult of frightfulness' prescribing rape, execution, and the torching of villages and cities. They bring home the point of the philosophers'

critique of untutored power. The old chronicles describe situations in which ambition comes into full play in the absence or impotence of countervailing moral ideation. They show us that the non-lethality contract longed for by Hobbes, and supposed by some evolutionary theorists actually to have come into force, was superseded by the re-emergence of lethal motives when humans abandoned life in small groups of related individuals for life in anonymous cities and city states surrounded by 'enemies'. The behaviour of our ancestors, if Homer and Herodotus are good guides to it, was highly responsive to context, clever, and inventive. If rational self-interest were our only value, it would command our unqualified admiration. But it does not.⁴⁰ Politics and some aspects of human relations seem in former times to have been driven by a flexible, situation-responsive opportunism that has become unusual if not unthinkable today.⁴¹

In private as well as in public life, moral vigilance, overtly peaceable relations, and the enforcement of regularity of conduct have increased.⁴² Modern people even seem to have fewer personal enemies than ancient people did, as Kenneth Dover pointed out in his study of Greek popular morality.⁴³ National leaders still commit atrocities, passively approved by their henchmen, but they no longer commit them as openly, and if the claim that there has been moral progress since ancient times tends to draw sceptical frowns, we should nevertheless acknowledge that the world of Hume's Stuart rulers was already a different world from that of the ancients. We increasingly place a value on protracted intramental deliberation as a precondition of issuing clear signals and on behaving in a consistent manner.⁴⁴

⁴⁰ Herodotus gives many examples of creative problem-solving amongst the ancients. He tells the story of the Egyptian king Sesostris, invited with his family to a banquet by his brother. As they were leaving, their host set fire to the path. 'As soon as Sesostris realized what was going on, he turned to his wife because he had brought her along with him too, and asked her advice. She suggested that he have two of his six sons lie down over the flames and act as a bridge across the fire, so that the rest of them could walk on them and escape. Sesostris did this, and although it resulted in two of his sons being burnt to death, this made it possible for their father and the others to escape.' Herodotus, *The Histories*, tr. Robin Waterfield, 135.

⁴¹ Douglas and Isherwood tell us that during the Hundred Years War in Bordeaux in the 13th cent. most of the 'confused crisscross of lords . . . sniffed the wind, weighed the risks of a change of allegiance, and passed from one camp into another, trafficking in loyalty. Raymond IV, Vicomte de Fronsac, owner of a river fortress, changed sides five times from 1336-1349.' Mary Douglas and Baron Isherwood, *The World of Goods*, 34.

⁴² As Gehlen argues, morality leads to a 'stabilisation of the inner life' so that it is not ruled by affective impulses or subject to psychologically costly and inefficient reflection. *Moral und Hypermoral*, 97.

⁴³ Kenneth Dover, *Greek Popular Morality in the Time of Plato and Aristotle*, 181.

⁴⁴ Dover reports various ancient sayings to the effect that the wise man readily changes his mind, *ibid.* 122. Their meaning is that deliberation, not erratic behaviour, is praiseworthy.

While too extreme a reduction in versatility and intelligent concern for one's own advantage is not always advisable, persons of strong moral fibre tend on the whole to rigidity.

To summarize, human interactions rarely concern persons equal with respect to their persuasive powers or the degree of force they are able to exercise. Because the exercise of morality does not tend to the advantage of the strong, it is surprisingly difficult to explain how moral practices, as opposed to prudentially concessionary tactics, can arise. It is not explanatory to say that morality reflects the fact that all humans are equal in their moral personhood, if not in the degree of social dominance they can exert. The emergence of the notion of equal moral personhood in the face of the manifest empirical inequality of human beings with regard to the degree of social force they can exercise and the proportion of the cooperative social product they are able to control is a moral phenomenon needing explanation. While relative social equality prevails in hunting-and-gathering societies, status and prestige are not evenly allocated. When money and technology facilitate the accumulation of advantage, the difference between one person and another with respect to possession of or access to the components of well-being is marked. The weakest lack the strength and resources to oppose the strongest. They may also lack the guile or the affiliative ability that Hobbes was confident would compensate for a lack of physical strength.

The proposal that moral rules are advantage-reducing imperatives confirms the historical link between our current moral practices and the prelinguistic, protomoral, self-suppressing behaviour and the punitive retaliation or immobilization attempts directed towards social 'offenders' by our non-human ancestors. The earliest moral rules, those that prohibited treacherous attack and murderous retaliation, and that structured property relations and marital attachments, were prohibitions and permissions whose truth or falsity never came into question. Though we now worry more than our ancestors did about justification, the modern institution of morality is still a system of prohibition-and-permission rules comprising positive obligations, such as duties of assistance and duties to perform contracts, and prohibitions on negligence and indifference. Our dispositions, both innate and learned, to interfere with others, to control their behaviour, and to suppress antisocial tendencies in ourselves are the source

of moral normativity. Freud, who noted the depth, universality, and suprarationality of mechanisms of suppression, suggested that the restrictive rules observed in non-literate cultures might ‘throw light on the dark origin of our own “categorical imperative”’.⁴⁵

The imposition of prohibitions, he maintained, is associated with sanctity and power, and prohibitions that have no obvious function or rationale are found in many ancient traditions. Their variability from epoch to epoch and place to place may contrast markedly with the beliefs of rule followers that absolute right and wrong are at issue, absolute danger and perfect safety. ‘Taboos’, according to Freud, concern ‘actions for which there [exists] a strong desire. . . . [Those who observe them] assume an *ambivalent attitude* toward their taboo prohibitions; in their unconscious they would like nothing better than to transgress them but they are also afraid to do it’.⁴⁶ Freud’s suggestion is worth pursuing. Of course the class of taboo rules is much wider than the class of what we are prepared to recognize as moral rules. Many reported taboos seem to concern actions that nobody could experience a powerful temptation to perform and that are hard to see as the objects of deep ambivalences.⁴⁷ A given taboo may not even have a distinctive social function. A. R. Radcliffe-Brown pointed out that any system of codes whose observance produces anxiety and relief and requires mastery by members of the society enhances group understanding and solidarity.⁴⁸

We tend to think, when moved more by biological than by anthropological considerations, that moral rules, unlike taboos, must contribute to human flourishing or at least to human survival, but functionalism is an unwarranted conclusion. If at least some moral rules are the residuals of ancient taboos, they may not have either a strict biological or a distinctive social function. The propensity to

⁴⁵ Freud, ‘Taboo’, 824. Samuel Scheffler pursues briefly the idea that the impression of inexorability of morality derives from an unconscious fear of punishment, in *Human Morality*, 80 ff.

⁴⁶ Freud, ‘Taboo’, 831.

⁴⁷ e.g., the high priest of Jupiter in Rome, according to J. G. Frazer, ‘was not allowed to ride, or even to touch a horse, nor to look at an army with arms, nor to wear a ring which was not broken, nor to have a knot on any part of his garments; he might not touch or even mention by name a goat, a dog, raw meat, beans, and ivy; his hair could only be cut by a freeman and with a bronze knife; . . . and his hair and nails when cut had to be buried under a lucky tree.’ J. G. Frazer, article ‘Taboo’, *Encyclopedia Britannica*, vol. T–Z, p. 13.

⁴⁸ A. R. Radcliffe-Brown, *Taboo*, 39.

impose harsh and complex requirements on oneself and others indicates that natural selection has not formed us to like only what is good for us as living organisms.⁴⁹ Assuredly we enjoy nutritious foods and protected dwelling places, and we dislike poisonous snakes. All the same, a culture may insist that its members undergo certain painful rituals in the name of honour and sanctity. Circumcision, body piercing, and tattooing are biologically bad for us, for they may result in infections, scarring, and permanent impairment. Yet natural selection does not weed out these decorative practices as fitness-reducing. Humans have surplus resources to draw on that enable them to override, and alternative sources of reward that enable them to ignore, considerations of inclusive fitness when inventing and imposing rules. It is possible that all moral rules modern agents regard seriously have been purified of counter-functional elements and that they contribute to human welfare but there is no particular reason to assume that this is so. Hobbes's Theorem, that the hedonic content of a society is a direct function of its moral regulation, holds only for certain values. After a certain point, one cannot make people any happier or more secure by imposing additional regulations on them.

Two important consequences follow from the conceptual gap between the existence of normative institutions and their justification. First, what is regarded as morality in many cultures, including our own, may embody regulations that are trivial or even harmful to human organisms even if they serve to promote group solidarity. This theoretical possibility should encourage us to look at claims for the intrinsic morality or immorality of well-entrenched practices with a critical eye. Second, even acceptable moral regulations may be biologically or socially counter-functional. A rule may be, from the perspective of prescriptive morality, a good rule, even if it does not prolong the life or reproductive capacity of the persons subject to it, and even if the regulation is exigent in a sense still to be explored.

As 'self-enforcing' norms that do not require the application of external sanctions, moral rules are at once natural, non-natural, and counter-natural. They are natural in the sense that human beings

⁴⁹ Cf. Michael Ruse's claim that 'natural selection has made us in such a way that we enjoy things which are biologically good for us and dislike things which are biologically bad', in *Taking Darwin Seriously*, 236.

everywhere have a tendency to construct and enforce them, just as they have a tendency to construct languages and to insist on their proper usage. They are non-natural in the sense that their specific forms are local, like languages, and do not correspond to species-specific dispositions. Finally, they are counter-natural in the sense identified by Kant and Freud. They are difficult impositions that may make large demands on our capacity for self-awareness and social monitoring and that call for the suppression of instinctive or spontaneously arising desires at the same time as they focus attention on them.

Paraworlds and Confirmation

As R. M. Hare observes, moral decisions involve the application of principles to particular situations.¹ Moral questions of the form ‘How should I behave when . . . in light of . . .?’ or ‘How should we all behave when . . . in light of . . .?’ presuppose that we have an interest in evaluating moral policies and in soliciting agreement and approval. These assorted interests present us with the problem of moral self-positioning, a problem that is constantly solved in practice, even without conscious reflection. We form opinions about what we ought morally to do, about what our acquaintances ought to do, and about what perfect strangers such as government officials and film stars ought to do. Some of us attempt to intuit moral reality on a proposition-by-proposition basis. Others guide their conduct according to principles derived from well-reasoned texts. Others go along with what their parents used to say or with what respected or irresistible authorities have laid down as right and good. Most of us draw on all these sources somewhat haphazardly in everyday life.

A given group or a given person may come to situate itself or himself close to the self-interested centre or towards the outer regions of hypermorality on any moral issue. A religious sect requiring hypermoral lifelong monogamy might be opposed to redistributive taxation. Some vegetarians are not monogamous, as surprising as this might seem, and many defenders of abortion rights are vegetarians. There are even just-war-theorist-vegetarians! The better-known moral theories comprise particular formulas of obligation and rules for generating further formulas, together with reasons for preferring them to alternative rules and formulas.²

¹ Hare, *Language of Morals*, 56.

² Gibbard remarks on these norms and their bearing on the search for belief coherence in *Wise Choices, Apt Feelings*, 168.

What are the characteristics of a moral theory and how are its particular propositions evaluated? This chapter advances the view that moral judgements can, in principle, be confirmed with the help of the apparatus of fictional worlds—paraworlds. Few moral judgements are plausibly regarded as already confirmed and most moral claims are addressed to what might be called ‘practical indeterminables’. Nevertheless, to just the extent that a proposition of moral theory is confirmable, it is binding on agents generally.

2.1. Moral Belief-Sets and Theory *M*

The totality of moral permissions, obligations, and restrictions that *N* accepts or endorses is the basis of the judgements $j_1, j_2, j_3 \dots$ that *N* offers or could offer about the rightness or wrongness of actions, policies, or situations. In so far as *N* never thinks about and is not asked to pronounce on many actions, policies, and situations, though she can do so if prompted, perhaps after a long period of reflection and deliberation, her system contains some moral rules only implicitly. Her beliefs may not be constrained by consistency requirements. The set of formulas of obligation and permission (formulas of obligation for short) that she endorses, together with the judgements she makes or would be disposed to make on the basis of these formulas, constitutes her moral belief-set.

Establishing the contents of this set for anyone, including oneself, is not a simple task, for the notion of ‘acceptance’ or ‘endorsement’ covers a range of attitudes and dispositions to act and to judge. Initially presented moral convictions can be thrown into disarray by skilful questioning. Asking a subject whether it is wrong to eat the flesh of mammals, to withhold medical treatment from newborns with such-and-such life-threatening handicaps, to provide narcotics and stimulants to addicts free of charge, or to tell a lie to avoid wounding the *amour-propre* of an author, would likely elicit a number of expressions of uncertainty as well as some confident judgements about right and wrong. Novel moral problems can present themselves suddenly and unexpectedly, and there is no reason to insist that an agent’s response to a query belonged to her system antecedently. Moral avowals do not always correspond to the principles that

would be imputed to the subjects who produce them on the basis of observation. The notion of an implicit moral system suffers from a certain vagueness concerning what is to count as an internalized rule. However, anyone who has at least one moral belief has a moral belief-set, and most people have a number of moral beliefs. Any vagueness that attaches to the notion of a moral belief-set must infect the very notion of a moral belief.

Local moralities are sets of formulas of obligation that are employed by a social unit to regulate, or to try to regulate, its own affairs. The included formulas are believed to be right or proper; they may be itemized and recited, or referenced in contexts of explanation and justification. The social unit may consist of a person, a tribe, an occupational group, a subculture, or another entity that can use the terms 'I' or 'we'. A local morality comprises the rules of conduct and deportment that can be articulated by members of the unit in question as obligations, permissions, and prohibitions; it can be construed as analogous to one of the dialects and idiolects that make up human language. A local morality prescribes responses to instances of transgression and to habitual transgressors. Its codes are supported by positive and negative sanctions, including praise and reproach, reward and retaliation, the bestowal of medals and the communication of thanks, by shaming, imprisonment, and ostracism. Local moralities are supplied with incentives and deterrents. Admonitory songs and stories predict the likely outcomes of certain actions, whether heartbreak or bliss, eternal disgrace or posthumous fame.

The notion of a local morality, whether it is considered under the aspect of a set of rules to which individuals strive to conform or a set of beliefs about right and wrong, is necessarily vague. The inherent difficulties of attributing sets of moral beliefs to individuals or determining which rules they are following are compounded by the difficulties of attributing sets of any type of belief or rule to entire groups or to persons who endure over time. The notion of a local morality is not altogether elusive, however, and anthropologists, sociologists, novelists, and feature writers are able to grasp and articulate important elements of local moralities without pretending to be able to deliver them whole.

An intuitive understanding of local morality is a basis for understanding what a critical morality might be. Kant thought of local

morality as a pastiche of religion, prejudice, sentiment, and folklore and he sought to replace community standards of conduct with universal standards to be adopted and observed by all humans. Philosophers are expert in subjecting to critical scrutiny ordinary notions of reasonable and contextually appropriate behaviour, and they frequently find that local systems come up short. There may be much genuine knowledge and useful insight contained in them, but their contents often appear to need both winnowing and supplementation. And just as the beliefs of some communities, such as ancient soothsayers, are not only factually incorrect but also unhelpful in achieving control over future events, some local moralities appear to be subject to both representational and practical failure. The position that a critical morality ought to replace local moralities need not be ethnocentric. Its proponents are likely to maintain that many of the obligations enshrined in the morality of educated Westerners need winnowing too, while many real obligations are yet unrecognized and unacknowledged by us.

Call M the set of formulas of obligation that partially incorporates, but at the same time wholly supersedes, local moralities in being the right set—the set that ought to comprise one’s normative beliefs. Moral nihilists believe that M is the empty set; one ought to have no moral beliefs. Sceptics are uncommitted to the emptiness or non-emptiness of M . Everyone who is not a nihilist or a sceptic believes that M contains some formulas of obligation, even if they are not sure what actually belongs in M . But what could make one set of formulas of obligation the right set? Moral realists have what at first appears to be a clear and good answer to this question: M is the set of moral truths and associated imperatives. One should believe what’s true and act accordingly. This answer proves to be remarkably unhelpful.

2.2. Authored and Unauthored Norms

It is often urged that nothing short of an acceptance of moral realism-with-bivalence can capture our understanding of ourselves as persons who strive to correct their moral beliefs and who are able to deepen their understanding of morality by experience and reflection and

through critical, belief-altering tests of moral principles. It is frequently suggested that accounts that portray morality as a system of elective and elected social rules imply that tolerance for every kind of barbarism is inevitable or even mandatory. The notion of a moral standard independent of human preferences that demands to be met, is different, it is claimed, from the notion of an action-guiding rule.³ Just as there are norms of logic that all discoursing creatures are bound by, whether they decide to heed them or not, and norms of prudence that constitute reasons for action for all long-lived creatures, no matter how recklessly they choose to behave, there are allegedly moral norms that constitute constraints on and reasons for action for everyone. To fail to acknowledge that ethical as well as logical norms exist independently of human agents is allegedly to reveal oneself as epistemologically deficient, as lacking an understanding of what moral knowledge is. By distinguishing between authored and unauthored norms, however, we can supply the realist with all that she demands without the absurdities of realism-with-bivalence.

According to the realist, when deciding what I may eat, I should decide to draw the line at shrimp rather than at flounder if and only if it is true that one is not permitted to eat shrimp but is permitted to eat flounder; and I should designate abortions after the fourth month of pregnancy as morally wrong if and only if it is false that a person may undergo or perform an abortion after the fourth month of pregnancy. And just as I ought to believe all and only true empirical statements, provided I have the storage capacity to do so, I should adopt all and only true moral beliefs, provided I have the capacity to do so.

Moral realism is consistent with the view that the notion of a prohibitory rule is central in moral theory. Anthropologically established rule-priority as alleged in Chapter 1 is not the basis of the criticisms that follow. For the imperative, optative, and declarative forms of normative statements are related. The issuing of a command to an agent implies a wish that an act be done or not done. The wish that an act be done or not done implies the existence of a thought that

³ v. Stephen Darwall, *Philosophical Ethics*, 64.

the action is good to do or bad to do.⁴ The directed-to-someone imperatives:

- (1) Everyone: Don't eat animals!
- (2) Police: Don't torture!

correspond to the generalized optatives:

- (1') Would that no one ever ate animals!
- (2') Would that the police refrained from using torture!

and to the moral beliefs:

- (1'') It is morally wrong (for anyone) to eat animals.
- (2'') It is morally wrong (for the police) to use torture.

The question of the truth status of moral judgements, therefore, cannot be dismissed. Nevertheless, whether or not some moral judgements such as 1'' and 2'' are true, the truth of moral realism—understood as the strong claim that every moral judgement is either true or false, that moral statements are subject to the law of bivalence—is neither a necessary nor a sufficient condition for the establishment of a critical morality. Nor is a belief in, or doctrinal commitment to, moral realism either a necessary or a sufficient condition for advancing good moral theories.⁵

One might imagine that a belief in moral realism is a necessary condition of being a morally serious person, just as a belief in God or some equivalent supernatural being, supernatural force, or supernatural status quo is a necessary condition of being a religiously serious person. If so, anyone who wants to be morally serious should form certain meta-ethical commitments whether or not moral realism is philosophically true. However, to be a morally serious person it does not even seem to be necessary to believe in moral realism. One need only believe that there are situations that are (really) morally significant, that some issues are (truly) moral issues. Chapter 1 established that this condition obtains; morality is a genuine category of anthro-

⁴ This is not to contradict R. M. Hare's point that imperatives cannot be reduced to indicatives; they cannot be reduced to indicative statements that do not contain or presuppose normative terms. As he observes, value judgements correspond to imperatives of extremely broad and unrestricted application. *Language of Morals*, 5; cf. 178.

⁵ For detailed analyses of the status of moral predicates, and their bearing on the problem of moral truth, see Simon Blackburn, *Spreading the Word*, 181 ff.; John McDowell 'Values and Secondary Qualities'; and David Wiggins, 'A Sensible Subjectivism'. Confidence in the reality of moral properties and the objectivity of (some) judgements ranges from the highs of Nicholas Sturgeon, 'Moral Explanations', to the lows of J. L. Mackie, *Ethics: Inventing Right and Wrong*, 38–42.

pology and acknowledging its importance is all that is required for being a morally serious person.

It is sometimes suggested that a non-realist can have nothing to say against Hitler and is seriously impaired when it comes to choosing a critical morality. However, the presence or absence of a commitment to moral realism has no necessary bearing on agents' judgemental capacities or dispositions. A non-realist is no less likely than a realist to insist that one ought not to deploy lethal power against helpless persons and that anyone who does so is immoral. The non-realist, to be sure, when asked why he *claims* that Hitler's actions were inexcusable will not answer, 'Because it's true!' but this outburst by the realist adds nothing that is not already present in the non-realist's answer, namely, 'Because that's what I think!' Neither answer addresses the question why Hitler's actions *were* inexcusable: here realist and antirealist will give indistinguishable sets of answers.

Finally, one can care about moral policies and moral judgements and their effects on actions and belief without being a realist, just as one can be a dedicated aesthete who adheres to strict standards of taste and who judges others' appearances rigorously without holding a realist theory of aesthetic judgement. I may simply prefer the condition of the world in which others share my codes and agree with my judgements to the condition in which they don't.

A commitment to moral realism is therefore not necessary for caring about morality or being interested in a critical morality: Theory *M*. It is not sufficient either. The realist identifies *M* with the hypothesized set of true moral propositions but realism cannot explain why we ought to try to discover and to believe just those formulas of *M*. There is no general obligation to discover and believe every true proposition. Even if *P* is true, one should not (prima facie) try to find out that *P* if doing so will overtax one's mental capacities, and one should not believe that *P* if one has no justification for *P*. One should not try to find out or believe that *P* if doing so will spoil one's own or someone else's life. If trying to establish the entire set of true propositions of morality would overwhelm our brains, we should not try to do that. If bivalence holds of moral propositions, but we have no justification for believing any one, we should not endeavour to believe even one. Finally, if believing moral truths (or even certain true theorems of physics or mathematics) would spoil

our lives or cause us to spoil others' lives, we should forgo belief in them. Moral realism provides no assurance that system *M* is within our capacity to understand, cognitively accessible to us, and benign. It leaves open the possibility that the true system of morality could worsen our lives, were we to adopt it.

Moral realism is accordingly neither a necessary nor a sufficient condition of the existence of a critical morality *M* with normative force. That said, the hypothesis that there exists a determinate truth value for each and every moral judgement apparently does help to answer the following question: Why do we think that our use of special argumentative methods and procedures in moral discourse achieves good results—better results than would be obtained by unimaginatively following old habits and customs or flipping coins and reading tea leaves to decide what positions to adopt and endorse and what policies to approve and act upon?

However, the more compelling answer to this question is that the special procedures of argument and inquiry in the normative disciplines, including aesthetics and informal logic, ensure that the critical norms enunciated are accessible and benign. They also provide us with reasons for adhering to the norms in question, reducing their arbitrariness. This claim needs explanation.

Many, perhaps most, of the useful practical norms that we prescribe and submit to are both *authored* and *targeted*. The declarative statement, *You are not allowed to use your computer after 10 p.m.*, with its imperative and optative relatives, is an authored norm invented by me and directed to my son. The rule really exists—I created it!—but nobody other than my son is obliged to obey it. It is a fact that my son is obliged to turn off his computer by 10 p.m., but it is not the sort of fact a moral realist is likely to be interested in. By contrast, the declarative statement, *One is not permitted to torture animals for stimulation or diversion*, appears to correspond to an authorless norm and to have everyone as its target, and statements like this invariably attract the approving attention of realists.

The difference between authored and targeted, and seemingly unauthored and universally directed norms is not to be found in their logical form or in their ontological commitments. In both cases, the norm is enunciated with the aim of preventing certain happenings. The problem of bothersome, sleep-robbing, late-night computing is

addressed by the norm I invented. The norm implies an answer to the question ‘What should be done about late-night computing in this household?’ (Answer: It shouldn’t take place.) The problem of harm to animals is addressed by the other norm that we invented, though no one claims personal responsibility for it. The norm implies an answer to the question ‘What should be done about people anywhere who deliberately try to hurt animals?’ (Answer: They should stop/be stopped.) Moral judgements originate as authored norms. Most remain in this relatively subjective condition, even when there is wide consensus concerning them. They represent the elections of a set of morally interested agents.

Nevertheless, a moral judgement can, in principle at least, transcend its humble origins. A moral judgement can be said to be confirmed when it commands agreement by appropriately informed morally intentioned persons. To paraphrase Rawls, a moral theorist is not seeking the agreement of ‘rational’, i.e. self-interested, competitors to his own proposed course of action, but the approval of ‘reasonable’ assessors to his scheme for everyone.⁶ Confirmed statements can even be regarded (though I will avoid this locution) as ‘true’. Therefore, it is possible that there are some true moral judgements, even if bivalence is not a feature of moral judgements in general. Meanwhile, many moral judgements, including, *It is permissible to eat shrimp* and *It is impermissible to have an abortion in the fourth month*, are probably not susceptible of confirmation or disconfirmation and are almost certainly neither true nor false.

But what is the form of a moral theory and how are its propositions confirmed?

2.3. The Form of a Moral Theory

A moral theory is not a mere collection of rules or judgements concerning the distribution of advantage that some writer happens personally to favour. It is a collection of rules and judgements presented in such a way as to work persuasively on the audience for the theory. Whether in the form of Marcus Aurelius’ Stoic *Meditations*,

⁶ John Rawls, *Political Liberalism*, 52 ff. These assessors must have certain ‘macroethical’ commitments; see below, Ch. 4.

the treatises of Kant and Bentham, or contemporary accounts of social justice, a prescriptive moral theory typically incorporates three components.

First, there is a supraempirical account of human nature, and, typically, an account of the status of humans vis-à-vis animate and inanimate nature. The moralist makes a set of claims, or indirectly establishes a set of assumptions about the degree to which humans are subsumed under or stand outside the rule of physical or psychological laws and are like or unlike other animals. Terms such as motive, incentive, and sanction may signal the adoption of a naturalistic perspective, while terms such as will, choice, and freedom may signal the adoption of a metaphysical perspective. Second, the moralist presents a set of first-order formulas of obligation, together with a second-order rule or set of rules from which he claims they follow. Third, the moralist portrays the consequences of acting in accord with or against his preferred imperatives. He may issue threats or offer inducements. The non-compliant subject may be portrayed as out of harmony with nature, out of reflective equilibrium, out of touch with his feelings, or as superstitious, old-fashioned, irrational, or brutish. The theorist may speak in the request mode, or the wish mode, rather than the command mode or the declarative mode. He need not assert 'This is right!', 'Do this!' He may only be urging, more softly and a bit regretfully, 'Would that the world were such that we all did that!' Stoicism, utilitarianism, and Kantianism each correspond to narratively intelligible views or pictures of the world and of how its inhabitants ought to behave, just as theories of the extinction of the dinosaurs or theories about the interactions of subatomic particles correspond to coherent pictures of the world and how its constituent entities did or do behave.

A moral theory, with its tripartite structure, can have a purely intramental existence, but usually a theory is articulated and propounded to an audience with the aim of influencing their beliefs and attitudes, and, sometimes, their actions. The pragmatics of theoretical moral communication are complex. There are distinctions to be made between (a) the readers to whom the work is directed, the target audience; (b) the persons whose beliefs, attitudes, and actions the work purports to represent, the implied moral subjects; and (c) the creatures and things that would be affected by changes in or

by the stabilization of the beliefs, attitudes, and actions of moral subjects, the moral community. Theorists' beliefs about the constitution of these categories of addressee may be accurate or inaccurate.

(a), (b), and (c) may coincide. Aristotle's *Nicomachean Ethics* and the third book of Hume's *Treatise* appear to be addressed to an educated and literate subpopulation, instructing them what to believe about, and how to behave with respect to other members of the same subpopulation for the benefit of that subpopulation itself. However, such closure is rare, and the audience that is intended to receive the doctrine and to assess the reasoning behind it may be only a small sector of the implied moral community. Kant's *Foundations of the Metaphysics of Morals* includes the residents of the South Sea Islands as members of the moral community, in so far as they are regarded as objects of moral reproach. They are not, however, members of the audience for the work and would not have been considered by Kant as competent judges of his theory. Many of Bentham's writings have as their intended audience an educated and literate subpopulation, but they are concerned with instructing it what to believe about and how to behave with respect to a subclass of labourers and prisoners. Rawls's conception of the class of implied moral subjects and of the moral community has expanded from book to book, and perhaps the composition of his audience has done so as well.⁷

That the moral community is not a concept with a fixed reference is obscured in discussions that refer to 'us'—to what we think, to what our judgements are, and to what we value without discriminating between groups (a), (b), and (c). The situational examples offered by moral theorists do not ordinarily suggest that 'we' includes the few remaining members of hunting and gathering tribes, illiterate peasant women of the Caucasus, or political prisoners in solitary confinement. The moral community typically comprises a set of agents wider than the author himself and the members of his intended audience but narrower than the class of all humans past and present, or all rational beings.

A moral text is produced with the intention of revising or reinforcing beliefs and attitudes, or at least in imitation of texts that are produced with that intention, and, as an item of reasoned discourse, it

⁷ Compare Rawls's *A Theory of Justice* (1971) in this respect with *The Law of Peoples* (1999).

does so by advancing arguments, detailing analogies, and citing relevant facts, and by staging thought experiments and asking the audience to decide their outcome by intuition. The text may also scold, cajole, or threaten. An important *modus operandi* of the moral text that has been little analysed is its evocation of imaginary or fictional worlds. To understand confirmation effort in moral theorizing, it is essential to investigate this aspect of formal moral discourse.

Consider first what might be called ‘narrative-explanatory’ theories from the natural sciences, such as the theory of chemical bonds, Newtonian mechanics, or the climate-change theory of the extinction of the dinosaurs. Such theories describe the dispositions and tendencies of entities such as electrons, ions, hydroxyl molecules, planets, tiny spheres suspended from chains, meteors, shock waves, clouds, trees, and so on. Each theory is derived from observations of nature constituting its data set, and it generates predictions, not about what human beings are going to think or say, but about what objects will be observed, when, where, and in what configurations. Antecedently to its being confirmed, a narrative-explanatory theory depicts a *what-it-would-be-like-if* scenario for its audience. A theory about how the dinosaurs came to be extinct presents an account of what an observer who had dwelt among the dinosaurs would have witnessed. It might recount a sequence of events beginning with the impact of an asteroid and the shock waves propagated by the collision, proceeding through changes in weather patterns, temperature, and food supplies, and ending with the gradual or sudden disappearance of the animals. A theory of chemical bonds will convey a sense of what we would witness were atomic particles and their interactions visible to our eyes. To confirm the theory is to find out that things are as the theory represents them as being, even where visualization of the processes in question is impossible.

One might think that a moral theory has little in common with a narrative-explanatory theory of the physical world. Empirical statements divide possible worlds into those in which the statement is true and those in which it is false; moral judgements do not have this feature. There is something it would be like for the fundamental constants of nature to be other than they are, but there is nothing it would be like for it to be morally permissible (as opposed to being believed to be morally permissible) to torture animals for stimulation

or diversion. Hence, there is nothing that it *is* like for this to be morally forbidden. There is a set of phenomena that the facts about force, mass, and acceleration explain. Call these ‘the phenomena for Newtonian theory’. And there is a set of ‘moral phenomena’ in the form of observed behaviour, moral judgements articulated by individuals, and penalties and sanctions applied for moral infractions. Yet the moralist’s theory, unlike the scientist’s theory, is not explanatory in the sense of making us see why just these phenomena must occur, given the facts about morality.⁸

Kant’s ethical theory resembles in some respects a narrative-explanatory theory in the natural sciences. It represents to us a hidden set of entities and interactions. It explains how the good will hearkens to the moral law, ignoring the solicitations of the senses. Yet these entities and their activities are not detected and tracked by experience and observation. No matter how favourably positioned and well equipped we imagine ourselves as being, we cannot imagine ourselves observing the rights and responsibilities that Kantian theory ascribes to us.

These arguments seem telling against the notion that we can confirm moral judgements by sampling the environment and collecting hidden data from our present surroundings. Nevertheless, the analogy between physical and moral inquiry is worth pressing further. To give a prescriptive theory for actual human beings is to give a descriptive theory for ideal agents, a theory of their moral competence. But how can a theory of moral competence amongst ideal agents be confirmed when we do not have access to their judgements but only to our own (non-ideal) judgements and to the natural world?

The answer to this question is that we rationally and irrationally motivate the acceptance of the statements of moral theory by projecting narrative-explanatory accounts of how psychologically real but morally ideal agents behave and by asking for the approval of an audience that is presumed competent to assess the account. ‘Ideal agents’ are not to be understood in this context as agents who always or consistently perform heroic or supererogatory actions. Rather, they are to be understood as agents who act in ways such that their

⁸ This is the substance of Gilbert Harman’s essential point in *The Nature of Morality*, 7 ff.

world—or the aspect of it under consideration—is better than the available alternatives. An ideal world is one that is preferred to other, relevantly similar worlds.

2.4. Moral Theories and Paraworlds

A paraworld is a fictional world that is, in some respects, like our world—close enough that we would feel at home there. The inhabitants of a paraworld resemble persons in our world, but they act in a patterned way that is usually, though not always, significantly different from the way agents act in our world. In any event, agents in a given moral paraworld behave differently than agents in a rival moral paraworld do. Moral rules are better regarded as the ‘laws of nature’ of alternative worlds than, as Kant took them to be, a special kind of law embedded, like the laws of Newtonian mechanics, in our world. In evaluating a low-level universally targeted moral judgement such as ‘Vegetarianism is required’ or ‘One must never lie’, or a higher-level universally targeted judgement, such as ‘One ought never to carry out a plan that will worsen the position of the worst-off person in a situation’ or ‘One ought always to ensure that new policies are not Pareto-inferior to existing policies’, the evaluator has to consider what it would be like for it always to be the case that the norm was realized. The mobilization of the projective imagination is the ordinary response to a controversial moral claim.

A moral text that depicts a paraworld embodies a set of assumptions about what people are like, what they need and want, how similar to one another or different from one another they are, how they make decisions and interact with one another. The moralist may communicate to her audience her views regarding the differences between humans and animals, adults and children, men and women, or the rich and the poor, or she may give out that there is very little difference between them. A moral text may depict paraworlds of quiet deliberation, erotic or artistic adventure, quiet immersion in an intimate circle of family and friends, or ambitious striving in a race for the acquisition of resources. These worlds may be inhabited by humble ascetics, loving caregivers, flexible schemers, by expert

game players and strategists, or even by subrational improvisers who stop short of ‘one thought too many’.⁹

In his ‘Gauguin’ fable, for example, Bernard Williams projects a paraworld in which at least some agents do not put others’ emotional interests ahead of their own personal ambitions. The Tahitian idyll furnishes the backdrop of a paraworld in which ‘Gauguin’ is portrayed as sharing important human traits, talents, and frustrations with members of the target audience and at the same time as able to realize an ideal, the evasion of formulas of obligation propounded in rival moral traditions. It is a paraworld in which, despite its exoticism, the intended audience can recognize some familiar features of its own experience.

Paraworlds vary according to their creators’ implicit theories of human nature. A world in which agents are endowed with strong impulses to set fire to animals or to wage genocidal wars seems terrible to contemplate, the projection of a diseased or perverse imagination. A pessimistic depiction like Freud’s will be balanced in a moral text by the author’s suggestion that moral learning can keep destructive impulses under control. The more sanguine the theorist’s view of human nature, the less repression and coercion his ideal world will contain.

A paraworld has the following special characteristics:

The Reality Constraint

The actors in a moral paraworld are meant to resemble the moral subjects of our world. They are endowed with the motives, preferences, levels of rationality, and overall aims and purposes that are taken to characterize moral subjects in the actual world. They face situations and problems that are morally significant. The causal relations obtaining in our world are maintained in the paraworld.

The Idealism Characteristic

The actors in the paraworld behave in an ideal fashion, performing according to those formulas of obligation that their creator endorses, or refusing to perform according to certain non-exemplary formulas of obligation. Or, if the paraworld is presented for our condemnation, they observe the wrong rules, the rules comprised in a rival moral theory.

⁹ Like Bernard Williams’s wife-saving hero in ‘Persons, Character and Morality’, in *Moral Luck*, 18.

In Kant's paraworld, the agent is partially exempt from the laws of physical nature. His ordinary human body, plagued by selfish animalistic desires, is accompanied everywhere by an invisible noumenal self, motivated exclusively by reason. The creatures of Bentham's paraworld are ruled by two masters, pleasure and pain, and Bentham's moral agents recognize this and behave accordingly. They are not influenced by religion, sentiment, ascetic impulses, and tradition, unlike their benighted counterparts in our actual world.¹⁰ They aim chiefly to maximize aggregate utility. In some texts, a moral antitype is conjured up by evoking for the reader the ideal behaviour displayed in the paraworld of a rival. A utilitarian antitype in a paraworld invented by a deontologist might remove the organs from one healthy person to save five others. A deontological antitype in a world invented by a utilitarian might make others miserable by upholding the principle of sincerity to the foot of the letter. The paraworld may even portray as ideal agents subjects who behave no differently than it is believed that moral subjects, in particular, members of the intended audience, already do. The intention of the moralist is to convince the audience that behaviour that has been criticized as morally defective by rival theorists is not so. The claim that the actors in a moral paraworld behave in an ideal fashion does not imply that the actors all observe stringent moral rules, or that they all observe the same rules, or that they are calm or conforming. The ideal world of an anarchist might be one in which everybody behaves wildly and differently from everybody else. The ideal world of a libertarian might be one in which everybody behaves just as he or she pleases. It is their job as prescriptive moralists to persuade us that their worlds are realistic as well as good.

The prescriptive formulas comprised within a moral theory, then, constitute the author's instructions to the ideal moral agents of a paraworld concerning how they are to behave. Emotivism is accordingly false. 'Ought statements', whose surface grammar is that of a declarative sentence, are not equivalent to declarative statements about what the assertor likes, admires, or prefers. For statements about what people like, admire, or prefer, unlike moral judgements, have truth-makers in this world. The semantics of ought statements

¹⁰ Bentham, *Principles of Morals and Legislation*, ch. 1 'Of Principles adverse to that of Utility', 8–23.

are, by contrast, irreducibly modal. They have representational content, but this content is given by the exemplification in the idealized paraworlds of the 'is statements' to which they correspond.

If *N* has the thought that it is morally permissible to perform a particular kind of action *ACT* in a particular kind of situation *c*, she has the thought that *ACT* is a self-interested action in *c*, but that in the preferred moral world some persons do *ACT* in *c*¹¹. More generally, where *I* is the all things considered preferred, or for short 'ideal', world:

ACT is morally obligatory in *c* = Everyone in *I* does *ACT* in *c*
ACT is morally forbidden in *c* = No one in *I* does *ACT* in *c*
ACT is morally permissible in *c* = Some persons in *I* do *ACT*
in *c*

Further:

ACT is supererogatory in *c* = Relatively few persons in
I do (what is as exigent
as) *ACT* in *c*
ACT is prima facie obligatory in *c* = Usually persons in *I* do
ACT in *c*

These latter definitions might seem tendentious. A supererogatory action is one that is morally praiseworthy but not obligatory; how does the notion of occasional but infrequent performance in an ideal world capture this dual feature? The answer to this question is that the performance of supererogatory actions makes worlds better than they would be in the absence of such actions. We do not prefer a world in which no one assumes the role of Mother Teresa, for example, to a world in which at least one person does. Yet, in so far as we regard her actions as non-obligatory, we express a preference for worlds in which her role is not adopted by everyone who could adopt it. Similarly, if *N* judges that truth telling is a prima facie, but not an absolute and unqualified obligation, then she regards it as the case that persons in the preferred paraworld usually do not lie but sometimes do. That they do not lie is a valid generalization, analogous to 'Birds fly': it admits of exceptions. To say that honesty is a moral

¹¹ In standard deontic logic, an obligatory action is one performed in *all* worlds, a permissible action in *some* worlds, and a forbidden action in *no* worlds. The present scheme is offered as psychologically closer to the way in which we think about morality, representing to ourselves just one ideal world, even if the agents in it have different maxims.

ideal or a moral virtue is to say that honest practices are well represented in preferred worlds and that honest dispositions are frequent in them. To deny that non-violent conduct is even a *prima facie* obligation is to indicate that non-violent dispositions and behaviour are represented no more frequently in ideal worlds than violent ones.

It is important to be able to represent *prima facie* duties, since it is generally conceded that formulas of obligation generate inferences non-monotonically. According to most informants, the conclusion of the Kantian inference:

- P₁ Everyone ought never to lie
 C₁ I ought never to lie
 C₂ I ought not to lie at this moment

can be, in rational and considered judgement, rejected, if we add premisses to the effect that the lie is told to beguile a child, or that failing to tell it will result in grievous, irreparable losses, multiple deaths, etc.¹² The *prima facie* acceptable principle that it is always permissible to perform an action that will preserve more human life than one that will preserve less should not generate the conclusion that we may divest one well person of his internal organs to save five transplant candidates from death.

Theorists try to replace weak generalizations with formulas of exception, e.g., 'It is always permissible to perform an action that will preserve more human life than one that will preserve less, *except when* doing so requires an agent to kill an innocent person.' They do not, of course, any more than other fiction writers, attempt to imagine everything that occurs in a moral paraworld but only what happens that is of relevance to the persuasive task. Nevertheless, it is difficult to describe precisely the behaviour of agents in environmentally and institutionally complex paraworlds and to specify formulas of obligation that are both simple and exact.

¹² Even Kant showed himself surprisingly flexible in dealing with exceptions to his rule that one should never tell a lie or commit suicide: 'To be truthful in all declarations, . . . is a sacred and absolutely commanding decree of reason, limited by no expediency', Kant says in 'On a supposed right to lie from altruistic motives', in *Works*, viii. 427. The *Lectures on Ethics* contain nevertheless discussions of cases in which the telling of an untruth or the performance of misleading actions is permissible. See the *Lectures on Ethics*, tr. and ed. J. W. Ellington, 228. Elsewhere he suggests that 'Want of candour . . . is still very different from that lack of sincerity that consists in dishonesty in the actual expression of our thoughts.' 'Letter to Maria Herbert', in *Kant: Philosophical Correspondence*, ed. Arnulf Zweig, 188. Elsewhere, by differentiating between sacrificing my life, shortening it by intemperance, and suicide proper, Kant also qualifies the severity of his other major prohibition.

The content of Theory *M* could never be represented in the description of a single world, for how moral offenders ought to be or must not be treated is itself a moral question. If *N* is concerned with moral norms for the treatment of prisoners, *N* needs to invoke a paraworld in which at least some persons have transgressed legal norms, and, in realistic paraworlds, some of these persons will have violated moral norms. Theory *M* could accordingly only be represented by a hierarchically ordered set of paraworlds in which violators of lower level norms are written in as elements of the reality constraint in considering the morality of punishment, retaliation, and revenge. These complications remind us that the formulation of Theory *M* is, like the articulation of a physical ‘theory of everything’, a remote ideal without calling into question the confirmation procedures theorists appear to use in trying to certify or discredit individual formulas of obligation.

The audience’s robust moral intuitions limit the class of paraworlds whose generative rules are eligible for acceptance and adoption. The notion that the principles of a good theory ought to reflect moral judgements about which there is wide agreement follows from the need to justify a theory to a critical audience. A permission rule that allows anyone to perpetrate harm to anyone else whenever it would benefit him at all to do so generates a world that is morally inaccessible to most of us. Not only do we doubt that, if our behaviour were regulated by such a rule, things would go well; we do not think that we are the sorts of creatures who would find it fitting to our natures to observe such a rule. The ‘we’ who confidently reject the egoistic paraworld is larger than the ‘we’ who confidently reject a paraworld in which setting fire to military personnel sometimes happens.

A moral theory, then, can be regarded as confirmed if it is accepted by an audience that is competent to assess the degree to which it satisfies the reality constraint and the idealism characteristic. Whether any particular formula of obligation or any collection of such formulas in a system has ever been confirmed depends on whether there exists or has ever existed a competent audience and whether they have ever come to agreement. This question is not easy to decide. The competent audience might be conceived as the set of all reasonable persons, whatever their state of information. They, however,

have yet to agree on anything and this conception of a competent audience places moral theory on an entirely different footing from any empirical theory, since the class of all reasonable persons is not believed to be competent to settle, or even to try to settle controversies in the natural and social sciences. Only the opinion of certified experts is taken to count. At the same time, no merely institutional criteria seem adequate to the definition of a competent set of judges for moral theory. The association between morality and the dictates of conscience on one hand, and our knowledge that moral authorities of the past propagated harmful misconceptions on the other, make us reluctant to recognize a category of 'moral experts'. Idealization is involved in the very notion of a competent assessor of moral theory.

Sceptics deny either that a unique and competent set of judges can be identified, or that competent judges would agree about anything significant. Nevertheless, it is clear what the moral theorist is aiming at when he presents his work for evaluation, what confirmation effort comes to. Moreover, against the sceptic's protests, we can state with some confidence that, although we cannot provide a general definition of a competent assessor, or identify such persons in the population, the proposition that Hitler's actions were morally permissible has been disconfirmed. Persons who regard advantage-reducing rules as important—that is, who are committed to the reality of morality in the anthropological sense—and who are well informed about Hitler's actions overwhelmingly prefer a world in which no one engaged in the behaviour that the historical Hitler did to a world like ours. That numerous people had or still have other preferences does not alter its status as a disconfirmed proposition.

Prescriptive moralists tend to reinforce their audience's robust prescriptive intuitions (as I have just done) and to concentrate on transforming their softer intuitions in a way that gives rise to what might be called standard methodology or the method of cases. Suppose a prescriptive moralist wishes to budge readers from the soft intuition that affirmative-action programmes are unfair or from the soft intuition that they have no duty to assist distant strangers. She might accomplish this by portraying a benign paraworld, in which moral agents act according to a race or stranger-favouring rule without ill effects, and she might appeal to our liking for theoretical unity by showing that a simple set of instructions is adequate to generate that world. She can try to represent a condition in which preferential

treatment coexists with rewards for merit, showing that there is no inconsistency between the two principles, or in which assistance to strangers and assistance to family members are undertaken from the same motive of care.

Alternatively, the theorist may be concerned to preserve certain softer intuitions from challenge in a rival theory. Often this can be accomplished by pointing to certain features of human nature that make revisionary proposals difficult to implement. A meritocrat, for example, might challenge an egalitarian by describing a paraworld in which persons very different in their levels of talent and effort, such as he supposes persons really to be in our world, are rewarded indiscriminately, or according to other criteria such as age or sex, and asking us to appreciate the disharmony of the resulting world and the frustrations experienced by the hard-working and talented.

Even robust intuitions can be modified by an encounter with a moral text. Like a novel that is hard to start but that becomes increasingly compelling, a paraworld that is initially inaccessible to us can become accessible as we become persuaded that the author's efforts to incorporate the reality constraint have been successful, or as we are increasingly moved by his rendering of the idealism characteristic. The conservation of prior robust intuitions does not constitute an absolute constraint on the eligibility for endorsement of a moral theory, any more than it constitutes an absolute constraint on the eligibility for acceptance of a physical theory.

The claim that prescriptive moral discourse functions in this manner might be doubted. A rival tradition holds that credible prescriptive writing is the communication of moral perceptions and moral discoveries, issuing from or made by theorists who have a finely developed moral sense and a feel for logical coherence. On this account, they are able to perceive or intuit the wrongness and rightness of certain actions and situations and to communicate to an audience those veridical perceptions and intuitions, replacing the audience's pre-existing illusions and confusions. The moral theorist helps the audience to see the moral worth of an action or situation in the same way that a good film critic helps an audience to see what makes a film aesthetically flawed or a fine production as the case may be. On this view, the enunciation of normative positions is consequential on the seeing and intuiting of moral properties in actions and situations occurring in our world, not on imaginary goings-on in ideal paraworlds. For, if the explanation

of the meaning and function of ethical discourse demands explication in terms of the behaviour of ideal agents in alternative worlds, why shouldn't this be true of all normative discourse—whether in logic, aesthetics, or medicine?

According to some meta-ethicists, we can see that *It is wrong to lock little children in broom closets when they have misbehaved*, as we can see that *Hitler was an evil person*. Moral goodness and evil are relational properties that human perceivers are equipped—some better than others—to perceive. Alternative worlds are no more involved than they are when we see that a joke is funny, or even that a lime is green.

Despite the popularity of this theory, it is not a serious competitor. The theory of moral direct perception, considered as a phenomenological account of what it feels like to make a moral judgement *in situ*, is entirely compatible with the account offered here. To have the experience of perceiving or intuiting a moral property in an action or a situation is just to have a certain thought about a paraworld in the same way that to see or intuit a potential mate in a current date is to have a thought about a possible world or a future state of the world, even if the man in the street hasn't yet found his way to this analysis. A critic's normative claim that a film is (really) funny is a claim about what idealized people with good senses of humour *would* all experience, were they to watch the film. The claim that a sofa is really oyster, not beige, is a claim about what colour name linguistically and perceptually competent observers say it is in their worlds.

All normative utterances project a world in which persons like us comport themselves differently from the way they do in our world. In an aesthetically ideal world, they create and surround themselves with different objects, and in a logically ideal world they reason differently and draw different inferences than our agents do. To declare that *ad hominem* arguments incorporate fallacies is just to represent ideal epistemic agents as consistently avoiding recourse to *ad hominem* arguments.

2.5. How Remote are Paraworlds from the Real World?

The hypothesis that defending and challenging moral propositions, in contexts in which the establishment of a critical morality is the task at

hand, depends on the imaginative projection of paraworlds explains why moral philosophers so often appeal to literature as a source of examples and illustrations. Novels furnish ready-made paraworlds. Frequently, a novelist is advancing or criticizing a prescriptive thesis and draws his moral paragons and antitypes accordingly. Like narrators of realistic novels, moral philosophers constitute themselves as reliable observers of moral phenomena. However, the author can only theorize the actual as she or he perceives and experiences it, and unnoticed biases or even fantastic beliefs can deform her or his portrayal. The reader of a prescriptive text has to decide whether the moralist's conception of human nature and the demands placed on it are sufficiently accurate to make his recommendations credible and whether the rules of conduct implicitly or explicitly recommended are suitable for creatures such as we are.

The norms observed by the agents of a literary paraworld might seem adapted to them, given the sort of world they are in and the sorts of creatures they are, but it can seem to the reader that it would not be good for her or for us to live under their moral regime. Critics often challenge a moral theory by arguing that our world is fortunately unlike a given philosopher's paraworld. Julia Annas, for example, argues that the fictional world of *Effie Brist* is a Kantian world and an exceedingly unpleasant place to live. Martha Nussbaum suggests that 'Henry James's *The Sacred Fount* is a fascinating account of what the world looks like to a man who . . . [allows] theoretical intellect to determine his relation to all concrete phenomena, refusing himself any other human relation to them, and yet at the same time priding himself on the fineness of his perception'. A more Aristotelian world, she argues, is preferable to this one.¹³

It is often unclear whether a theory has failed properly to characterize the moral community, or whether its idealism characteristic incorporates a strongly revisionary element. Two examples can be cited:

Alasdair MacIntyre describes a 'practice' as 'any coherent and complex form of socially established human activity through which goods internal to that form of activity are realized in the course of trying to achieve those standards of excellence, which are appropriate to and

¹³ Julia Annas, 'Personal Love and Kantian Ethics in *Effie Brist*', 15; Martha Nussbaum, 'An Aristotelian Conception of Rationality', in *Love's Knowledge*, 81.

partially definitive of, that form of activity'.¹⁴ MacIntyre lists playing football, chess, farming, architecture, physics, chemistry, biology, painting, and the making and sustaining of family life, as practices. Now, few women have historically been involved in football, chess, architecture, physics, chemistry, and biology at a level involving the pursuit of excellence according to externally established standards, even if their current involvement is much greater than it was. Indeed, except for weaving and sewing it is difficult to think of an activity traditionally associated with women and traditionally performed by them in sizeable numbers that has ever been tested by and made subject to objective standards of excellence. MacIntyre cites the making and sustaining of family life as a practice, but women rarely take cleaning, cooking, and home nursing and education to specialist levels of excellence.¹⁵

On one reading, then, the moral community consists chiefly of men who, ideally, are all engaged in one or more practices. Their world also contains women, as well as trees, houses, animals, and other features of the landscape. One may feel such a world does not satisfy the idealism characteristic. Interpreted in another way, the moral community consists of men and women and, ideally, the women are fully involved in practices, either in football, chess, chemistry, etc., or in some expanded list of activities that count as practices. Under the latter interpretation, MacIntyre's prescriptive ethics is strongly revisionary, since it effectively asserts that, in a morally acceptable world, either women have to be involved in football, etc. to the same extent as men, or more women's activities have to be tested by and made subject to objective standards of excellence. In so far as MacIntyre does appear to be critical of women's wasting their time and their talents on undisciplined activities that are not subject to external assessment, there is good reason to ascribe to him the revisionary interpretation, but it cannot be said that his text addresses the point directly.¹⁶

¹⁴ Alasdair MacIntyre, *After Virtue*, 175.

¹⁵ E. F. Schumacher argues that there can be no expertise in family relations, since they present a set of 'divergent' rather than 'convergent' problems. *Small is Beautiful*, 79–80. It follows that the maintenance of family life can't be a practice in MacIntyre's sense.

¹⁶ MacIntyre is, nevertheless, critical of modes of life that, however elegant, refined, complex, and meaningful, leave or left people without significant work to do. He describes the lives of aristocratic women of the Meiji era as pathetically meaningless. See *After Virtue*, 223.

A second example, from Colin Macleod, presents the reader with a similar interpretive problem:

[W]e can roughly distinguish three kinds of choices which may legitimately ground differences in the holdings of individuals. First there are choices concerning the mixture of work and leisure in one's life. A person may be able to earn more income by working harder or by putting in longer hours. Second, there are what might be called 'production choices.' These are decisions about what sorts of goods and services to produce or provide. The more we make available what is in demand by others, the more resources we may acquire through market transactions . . . Third, there are choices concerning the sorts of risks one is prepared to take in leading one's life. While one person may choose to run high risks with a small chance of a huge payoff, another may prefer to minimize risk in order to secure a more modest but more likely payoff.¹⁷

In its context, the passage describes a paraworld in which holdings cannot be acquired except by work involving effortful application, and in which any subject can make such applications, though many will choose not to. Such acquisitions, and, presumably only (or chiefly) such acquisitions, are legitimate.

Had this passage appeared (with suitable linguistic alterations) in a seventeenth-century political treatise, readers would have been surprised by the novel plan of social organization being proposed. For most of human history, holdings were not a function of the holder's ability and willingness to produce and exchange socially valuable goods and services. They were taken to be legitimate if acquired by conquest, inheritance, or marriage, or as gifts. The modern reader may not share the old conception of legitimacy but he can nevertheless ask how remote the world evoked in the passage is from the real world.

What are the background assumptions in the passage quoted as to how the world works and how much verisimilitude do they possess? The generalization that reward levels depend on an agent's application and the risk level he or she is willing to assume is valid in many contexts. A philosopher can decide to revise his manuscript, or, alternatively, to entertain himself with little trips to the hardware store; to try to publish a paper in a bottom-drawer journal with a

¹⁷ Colin Macleod, *Liberalism, Justice and Markets: A Critique of Liberal Equality*, 50.

high acceptance rate, or to aim for a top-drawer journal with a low acceptance rate. These decisions will influence his outcomes. Most persons, however, lack full control over their output and their risk levels.

A mathematician cannot decide to double her output of mathematics papers or to write riskier ones that might bring her a higher pay-off. She will write the number of papers she can, and they will be timid or bold, depending on the quality of her mind. Farmers in rural subsistence economies, parents with young children, unemployed accident victims living on social assistance who cannot lift, type, or use a telephone comfortably—none of these agents can decide to produce more exchangeable goods, or to sacrifice some leisure. Other agents are uninterested in the prospects of a big pay-off. Or the question of the legitimation of their holdings is moot for them, since they don't have any. The holdings of many young mothers do not depend on their ongoing production-and-risk choices, but on a single past decision: whom they decided to marry.

On one reading of the passage, in a morally good world, people who for whatever reason do not choose to produce very much will not have much by way of legitimate holdings; on another reading, in a morally good world, everyone has substantial control over his or her production and risk choices. The portrayal of a moral paraworld different from our world, it might be insisted, is intrinsically critical because of the difference it tries to signal between actual and ideal conduct, or between inappropriately idealized conduct and truer-to-our-natures conduct. The proffered shoe only fits some moralists, however, in some parts of their texts, and not others. The thesis that Aristotle must have meant that women and slaves ought to realize themselves in politics or contemplation though they did not do so in his time is absurd, and the suggestion that Kant was advocating the full participation of women and servants in the intellectual and political life of his society is no more tenable. In their schemes, women and slaves or servants are not good or focal examples of moral subjects, and the discourse is not about them.

In summary, three conditions must be satisfied by the paraworld in which a moral judgement is descriptively realized if the judgement is to be considered as a confirmed element of *M*, the set of moral principles to which our actions ought to conform. First, the moral

community must be depicted in a manner that is biologically, psychologically, sociologically, and anthropologically realistic. Our dispositions and tendency, the realities of our various situations, whether they are posited as flexible or inflexible, must be portrayed in such a way as to stand up to criticism. Second, the behaviour of the agents in the paraworld that is commended to the audience must seem to be possible for members of the moral community who are constituted as they are. A theory that places overly exigent requirements on us cannot be confirmed. Third, the paraworld corresponding to *M* must be judged appreciably better than any of the alternatives from amongst the set of eligible paraworlds. The competent audience is not, however, required to agree on or to stipulate the criteria by which 'betterness' is to be judged, only on betterness itself.

This claim might seem outrageous. Surely, a theory of morals ought to tell us what makes one world morally better than another? This objection betrays, however, a confusion of levels. The task of specifying the criteria for paraworld betterness is analogous to the confirmation theorist's task. It is distinct from the scientist's task or the task of the ordinary philosopher seeking to explain scientific inquiry. It is far easier to get the agreement of competent assessors that $E = mc^2$ than to get agreement on the criteria that competent assessors use. A reputable branch of the theory of morals, analogous to confirmation theory in the philosophy of science, might well address the higher-order normative question of paraworld betterness—what makes one paraworld *really* better than another? My concern here is merely to give a non-normative philosophical account of how the notion of paraworld betterness is employed by rival moral theorists.

The agreement of a competent audience on these three features is necessary and sufficient, according to our claim, for normativity, for the existence of obligations binding on everyone, regardless of their personal views. Moral judgements, on this view, represent preferences, as non-cognitivists maintain, or more precisely, elections. At the same time, they are not mere subjective preferences, or emotion-based elections, or elections made regardless of what the judge believes others' elections to be. They have representational, not only conative, content. When we make a moral judgement, we represent certain patterns of action and forbearance as instantiated in a morally good world, without coming to know that they are.

Do *I* have to do what a confirmed statement in moral theory says people should do? Can meta-ethics show that agents are obliged to act in accord with confirmed moral theories? This might be doubted. No one, it seems, can demonstrate to me that I am obliged to follow the rules of logic in having a conversation, display culturally expected table manners, or dress in workplace-appropriate clothing. I am free to ignore all such norms, provided I can tolerate the resulting sanctions. I am equally free to ignore all formulas of moral obligation presented to me, one might argue, no matter what critical procedures they have been subjected to, provided I can accept the consequences.

Having the ability to ignore obligations is not, however, the same as not having them. In so far as a moral theory is a theory about what agents ought to do, agents ought to do as the theory says—provided the theory is confirmed and there are no good reasons for disregarding it. One ought generally to believe propositions confirmed by scientific inquiry, provided there are no important countervailing considerations. Further, *to the extent that* a theory has been confirmed, we are obliged—weakly or strongly—to do what it says. If we fail to do so, our fault is that much smaller or greater, depending on how well-confirmed the theory is.

Suppose, however, that we cannot confirm that one must never take another human life. Is murder then morally acceptable? For it might seem that if no one is impersonally obliged to refrain from murder, murder must be morally acceptable. The permissibility of murder does not, however, follow from confirmation failure. The fact that *we* cannot confirm that one must never take another human life does not imply that paraworlds in which *some* people kill or murder are preferable to those in which no one does.

2.6. Relativism

The notion that moral theories project paraworlds that enable us to evaluate their authors' virtual moral systems does not prejudge the case for universalism over relativism. There is no contradiction in the suggestion that, in a morally ideal world, distinct sets of agents organize their lives differently and behave differently. We can

distinguish between ‘world-level relativism’ and ‘theory-level relativism’. World-level relativism is realized if advantage-reducing practices within a world are different for different groups. It is a permission rule for groups. World-level relativism is simply a characteristic of some paraworlds. It is exemplified in the following conditions.

World-level relativism

In I , at least one group always/sometimes does ACT in c and at least one other group never does ACT in c . Or, at least one group never does ACT in c and at least one other group always/sometimes does ACT in c .

A trivial degree of world-level relativism results from virtually any circumstance-specification. The claim that persons may lie, but only to save their own lives is trivially relativistic. Relativistic content can also be disguised by global formulations. Suppose N believes that vegetarianism is the correct moral policy for persons in Southern countries with a developed food industry who have access to a wide variety of vegetables and starches but is not required for the fishers and seal-hunters of the North. She can advance action rules such as, ‘Anyone may eat meat in circumstances in which they are hungry in the North’ and ‘No one may eat meat in circumstances in which they are hungry in the South’. Though all persons in her ideal world confirm to what are nominally the same rules, they are relativistic in content.

Theory-level relativism, by contrast, is a philosophical thesis, not a characteristic of worlds. It is true if and only if two opposing moral claims are both confirmable to the same, impressive extent.

Theory-level relativism

In World 1, ACT in c is at least sometimes performed and in World 2 ACT in c is never performed; or in World 1 ACT in c is sometimes or never performed, and in World 2, ACT in c is always performed, and Worlds 1 and 2 are both versions of I .

One may reject theory-level relativism while accepting world-level relativism or (less intuitively) one may eschew world-level relativism insisting that all agents regardless of what group they belong to must behave alike in c , but that two qualitatively different worlds inhabited by uniform agents are equally ideal.

World-level relativism—moral diversity within a world—might be favoured on account of the varying circumstances alluded to in the

dietary example above. It might also be favoured on the assumption that humans have a drive to formulate distinctive styles of personal conduct. If this drive can only be repressed by coercion, a paraworld that incorporates the idealism characteristic will not instantiate a uniform code. Universal conformity to a certain pattern of behaviour—reflecting, for example, universal restrictions on meat-eating or numbers of wives—would, on this view, imply morally objectionable social dominance, through which one group succeeds in controlling another.

Theory-level relativism is not a dangerous or threatening doctrine. It is implausible, though not impossible. It is difficult to confirm to an impressive extent even *one* set of formulas of obligation, and the possibility that we could confirm *two* incompatible moral systems to an impressive extent seems remote. World-level relativism is a virtually inescapable characteristic of paraworlds and is no more intrinsically dangerous. At the same time, within a given prescriptive theory, what at first looks like a tolerant and generous provision for moral diversity can disguise morally objectionable social dominance. The relativistic aspects of a theory can be, in other words, grounds for rejecting it, and the ‘ideological’ appeal to diversity is criticized below in Sections 4.4 and 4.5.

The sheer difficulty of confirming moral propositions, meanwhile, leaves room for individuals to hold and live by different systems without violating objective norms, to practise what might be called ‘practical relativism’. For, although we can often be certain that they do not do certain things, we usually do not know exactly how moral agents in the best world behave.

Practical problems concerning what ought to be done—what it is right, acceptable, forbidden, advisable, rash, or permitted to do sometimes (but only sometimes)—have determinable solutions, making practical relativism an inevitability. To appreciate this underdetermined aspect of all practical decision-making, consider the targeted imperative:

(3) Please (you, let me) shut the window!

which corresponds to the superficially untargeted declarative:

(3′) We need to shut the window.

‘Please shut the window!’ is not susceptible of confirmation, but its *in situ* equivalent, ‘We need to shut the window’ does appear to be. If it

is 10°C outside, if A_1 and A_2 are flimsily dressed, if the room is underheated, and if A_2 has the power to close the open window, the judgement expressed when $3'$ is uttered by A_1 —it might be argued—is simply 'true'. Shutting the window is objectively practically desirable. The judgement that we need to or ought to shut it resembles judgements like, 'Your (buck-toothed) son needs/ought to have braces' 'Your (myopic) daughter needs/ought to have eye-glasses'. That you ought to get her a pair is seemingly an unauthored rather than an authored norm. You can even be said to have discovered *the fact that your daughter needs/ought to have glasses*. And though targeted to a specific individual, the normative statement can be said to imply that, other things being equal, anyone with a similar defect needs and ought to get glasses.

Statements of practical desirability are, however, only determinable under certain circumstances and determination-favouring circumstances often do not obtain when claims of moral desirability are advanced. A statement of practical desirability is determinable when the considerations supplied or taken for granted are easily recognized as reasonable and when no countervailing considerations are presented or discernible. Buck teeth and myopia over -1.50 diopters are aesthetic and practical deficits that are commonly and easily corrected, and, in most circumstances, there is no reason whatsoever not to correct them.

At the same time, a statement like $3'$ is not theoretically determinable in every context in which it could be uttered. Indeed, $3'$ may prove surprisingly contentious. Does the window need to be shut if A_2 has just climbed out of a sauna, if it is only possible to shut the window by climbing up onto a tall, rickety ladder, and if an under-ventilated gas fire is burning in the room? There may be no fact of the matter as to whether the window ought to be or needs to be shut. Agreement may signal nothing more than A_2 's reluctant decision to accept considerations proffered by A_1 , or the abatement of A_2 's motivation further to contest the issue. And just as there may be no fact of the matter about what constitutes a good diet or a sound financial plan, there may be no fact of the matter about how we ought to behave. Some diets are unhealthy and some financial plans are crazy but there are many reasonable, though mutually exclusive, contenders for the role of the best human diet and the soundest

financial plan. Confirmation effort is most effectively dedicated, in health, finances, and morals, to eradicating the worst and most common policies and practices, not to determining the best. The apparatus of idealized judges and ideal worlds is entirely conducive to this aim.

Moral disputes with others arise because of conflicting perceptions, both of what it is for things to go better or worse in a world and of what is possible, broadly speaking, for moral subjects. Moral agreement may not be reached when there is disagreement over how unpleasant some state of affairs is, when some people sincerely wish to have different experiences from others, and when it is unclear how difficult it is to implement some policy and what the long-term consequences will be. Disagreement about whether it is permissible to institute a taxation system that is steeply progressive, or obligatory to bring life to an end at the request of dying patients may be interminable under analogous conditions. This is not to deny that we are constantly making optimality judgements in the practical realm and adjusting our beliefs on the basis of experience.¹⁸ To the extent that a person is rational, he avoids driving whilst seriously inebriated, endeavours to save for his old age, and limits his use of household pesticides. Most people have the capacity to recognize the reasons for prudential actions, and, even if they do not exercise the capacity or their motivation to perform the actions is weak or non-existent, they are reasons for the agent.¹⁹ Though 'moral reasons' have some analogous features, moral judgements belong to the wider category of optimality statements and these are rarely determinable.

Optimality statements of the form '*p* is the best *q* for *r*' are not usually asserted on the basis of preferences but of experientially derived knowledge of what *r* is like, how *r* works or what processes occur within *r*, how *p* will function as a *q* for *r*, given the composition and structure of *p*, where *p* fits into the larger scheme of things, and what alternatives to *p* are available. By asking after what moral rules are best for us, we inquirers make plain to ourselves how essential it is to try to learn about our real as opposed to our perceived require-

¹⁸ Cf. Peter Railton, who describes his position regarding moral learning in 'Moral Realism', fn. 7 as 'stark, raving realism'. Railton's paper shows how experience can have a salutary effect on theorizing and moral self-positioning, but it is unclear why realism is warranted by this.

¹⁹ Williams, in 'Internal and External Reasons', mounts a sceptical attack, countered by Christine Korsgaard, in 'Skepticism about Practical Reason'.

ments, how our current modes of life address them, and what alternative modes of life might be available to us that regulate the damage humans can do to one another, to other creatures, and to themselves, differently. For most readers, and certainly for the present writer, the moral prohibition against bombing peasants has great lucidity. No situation is even imaginable in which the context would render bombing peasants a preferable policy to not bombing them. Yet, wherever interesting moral problems—problems to do with fair distributions, with justified warfare, with obligations to family members and strangers—face us, we will have difficulty deciding what is to be done. We will not be inclined to appeal to ‘the moral fact that *ACT* in *c* is forbidden’.

The position that morality presents a problem of optimization has some affinities with consequentialism. The scheme outlined here is not, however, convergent with it.

‘Consequentialism’ can be understood in two ways. On one interpretation, it is a controversial world-level moral theory—more accurately, a family of related theories. Let’s simplify by taking ‘strong consequentialism’ to be the world-level theory that it is always obligatory to maximize welfare. Strong consequentialism predicts that we consistently prefer paraworlds in which, *whenever* an agent can save two people by killing one, she does so. ‘Weak consequentialism’, the view that it is always permissible to maximize the general welfare, predicts that we consistently prefer paraworlds in which, *sometimes*, when an agent can save two people by killing one, she does so. Since it is difficult to establish that we do in fact consistently prefer strong or weak consequentialist worlds to worlds in which agents *never* kill one to save two when given the opportunity to do so—or in which they do so only under some more complex set of conditions—both strong and weak consequentialism are poorly confirmed. They represent authored norms.

On another interpretation, ‘consequentialism’ is the name of a family of theories of confirmation, not the name of a world-level moral theory. On this interpretation, a moral theory is the best if and only if the world it represents contains more welfare than other worlds. A confirmation-consequentialist could thus favour world-level virtue theory, believing that worlds composed of virtuous agents, and not welfare-maximizers, are definitely better because

they contain more welfare. Confirmation consequentialism is not, however, the best theory of confirmation. It is not a foregone conclusion that, in judging the betterness of worlds, we ought to attend only to information about overall welfare. We should also perhaps attend to information about relations between agents, or their autonomy, or the balance achieved between the aggregate of desirable states and experiences and what deprivations others can be made or allowed to endure.

Can we conclude that the greater the quantity of empirical information we can bring to the construction of a moral theory, the more adequate the theory? This hypothesis is suggested by the reality constraint. Competent judges, one might think, need to know their way around human affairs, and not only what usually happens, or has happened in the past, but also, like good lawyers, what can happen when one least expects it. If they are ignorant, their paraworlds will lack verisimilitude and their prescriptions will lack value.

To be sure, the uneducated peasant, as Tolstoy maintained, may have more genuine moral feeling than the corrupt sophisticate who has read a thousand books. However, the thought experiment that compares the virtuous peasant with the corrupt sophisticate does not address the question whether, other things being equal, moral theorizing is better when informed by more knowledge of the world. Take two persons *X* and *Y*, both of whom are endowed with an average degree of imagination and empathy. *X* is poorly informed about how the world works; she has little knowledge of economic and legal systems and their histories, or the variety of customs and institutions that have existed all over the world. She knows little about crowd behaviour, fear, aggression, and desire, how animals see the world, or the history of medical and religious views on euthanasia. *Y* is passably well informed on all these matters.

Suppose one learns that *X* and *Y* prefer different moral systems. According to the Tolstoyan, one has no basis for thinking that knowledgeable *Y* is more likely to have picked the better system than illusion-ridden or ignorant *X*; in fact, the presumption must be the opposite. If one has to choose between them sight unseen, one should choose naive *X*'s system. The Tolstoyan recommendation seems, however, unduly risky. The philosopher can easily 'imagine

a case' in which living under the system picked by ignorant X is preferable to that picked by knowledgeable Y . But if I were forced to stake all my future experiences in the social world on it, I would not decide to flip a coin.

Even if an omniscient being would necessarily pick a different—and better—system than an only moderately well-informed being would, we cannot require competent judges of moral theories to have the extraordinary qualifications of ideal observers. Nor is specialist knowledge in a given area always a stellar qualification for making ethical judgements. Doctors may be poorly equipped to construct a system of ethics dealing with the end of life, although they know more about the physiology of death than other people. International economists may be poorly equipped to construct a system of global justice, although they know more about wealth and poverty than other people do. A specialist focus virtually ensures that doctors and economists know less than other people do about other relevant subjects. All we can say is that, other factors being equal, the ethical judgement of the person with more physiological knowledge, when it comes to matters of life and death, or economic knowledge, when it comes to matters of global responsibility, is likely to be better.

To summarize, the theoretical question what a critical morality M would look like is equivalent to the question, what set of advantage-reducing rules would it be best for humans (or for some subgroup thereof) to adopt? Since we are interested in what is best for us as we are constituted, rather than what is true, we need not fear, as the moral realist must, that the permission and obligation rules we ought to live under are insufficiently benign. At the same time, interesting optimality statements of the form ' p is the best q for r ' are usually practically indeterminable. What is the best way to teach an introductory logic course? What is the best place for an unmarried woman to look for a husband? The only reasonable reply to such questions is, It depends. What sort of experience are you looking for? The analogous question sets limits on the content of viable moral theories. The next two chapters explore some limits to the pretensions of moral theories.

Limits on Theory I: Costs to Agents

Morality poses a set of problems about the regulation of self-interest in which we try to choose the best rules of conduct or policy using as good information as we can muster. As individuals deliberating over whether to follow a particular rule, or theorists trying to generate a set of acceptable moral formulas from a few principles, we recognize that there are trade-offs between the limitation of damage to others and the personal and group sacrifices of other desired goods and states, notably prudential and aesthetic goods and states, that we must bear if we adopt more stringent moral rules. We have to decide what quantity of non-moral goods we want to pursue in the form of personal ambitions and selfish or socially narrow goals, recognizing, in a way that is sobering, that doing so forces us to incur certain moral costs. A confirmed set of formulas of moral obligation would be one that is agreed by competent judges to have struck the best balance that can be struck and it would be generally binding. However, we usually have to settle for less, namely, for authored norms that have some aspiration to higher status and that are accordingly somewhat binding on all of us.

This chapter explores the problem of exigency. As recent critics have shown, the moral obligations urged on us by deontologists and utilitarians alike often seem excessively stringent, suited to persons more heroic, ascetic, and far-sighted than we consider ourselves to be. What features of persons and their environments favour lower levels of demandingness than those set by the two famous modern moral theories? How are the costs to an agent of acting in conformity with them judged to reduce their prima facie obligations, and what is the significance of costs to agents for prescriptive theorists?

3.1. Exigency in Moral Theory

Moral rules often forbid agents to take something they badly want to have, or do something they badly want to do. No matter how much you want a first edition of Hobbes's *Leviathan*, it is not morally permissible to steal one from your friend, your library, or your bookseller, even if you can do so undetected. No matter how much you like money and dislike your wife, you are not morally permitted to hide assets from your wife in a divorce case, even if her lawyer will never find out. Morally good people, we agree, do not do these things—or do them rarely, and only under truly extraordinary circumstances.

The satisfaction of having done the right thing often compensates for the sacrifice of not having or not doing. However, it would be Panglossian to suppose that the hedonic equation always balances. Hobbes's Theorem, as it was designated in Chapter 1, states that for at least some values, the addition of moral regulation to a world increases its hedonic content, the amount of pleasure, comfort, and happiness it contains, and is generally true. Hobbes's Theorem, though, does not apply in all local subsets of the world, or for any arbitrarily large quantity of moral regulation. One cannot always assure an individual or a group that his or their happiness or flourishing will be increased by the addition of increments of virtue to his or their conduct.

A thoroughgoing hypermoralist—and this is a creature of fantasy—has adopted a set of moral rules that are all subformulas of policy Q. He never promotes his own interests if someone else's interests can be furthered or maintained by his refusal to exercise a natural advantage, such as superior beauty or intelligence, or a social advantage, such as wealth, freedom, or influence. Even if they fall short of thoroughgoing hypermoralism, the beliefs and behaviour of martyrs, hermits, and ascetics, adherents of certain subformulas of Q, appear to most observers to represent either a fine but unattainable ideal or a frightening deviation from normality. The psychological platform it would take to support a motivation to adhere to Q is not easy to describe, for why should an agent who does not appear to value his possession of some good G—as is evident in his willingness

to renounce *G* for another's benefit—regard it as good that another should enjoy the possession of *G*?

Many ordinary pursuits, meanwhile, invite and tempt people to advance their own interests at the expense of others, to set moral considerations aside. While the acquisition of wealth and freedom is not a zero-sum game, since they can be increased in aggregate, one way to acquire more wealth is to dispossess others, and one way to acquire more freedom is to persuade or coerce others into giving up some of theirs. Influence is ordinarily acquired by displacing or silencing other influence-seekers. Wealth, freedom, and influence are self-multiplying in that their possessors are able to employ them to acquire more of those same goods than others have.

Most goods have at least some tendency to permit the accumulation of advantage. The beauty that enables one to become a fashion model is likely to confer additional opportunities and resources for beautification, such as a free wardrobe or charm lessons. The intelligence that brings admission to an institution of higher learning is likely to confer additional opportunities and resources for the perfection of knowledge and its effective presentation. Conversely, the initial possession of fewer natural and social advantages than others have tends to draw out the advantage-taking tendencies of the better-endowed in the absence of moral ideation. It is easier to steal from the poor than from the wealthy, to induce the already unfree to undertake additional labour, and to silence altogether those who do not have much say in things to start.

Prescriptive moral theories set constraints on the pursuit of self-interest and the accumulation of advantage. Credible prescriptive theories vary nevertheless in the number and intensity of demands they place on their subjects. They may be relatively exigent, inclining to the hypermoral renunciation of advantage, or relatively relaxed, loading agents with few or weak prohibitions. It is frequently observed that the famous modern moral theories—Kantianism and utilitarianism—tend in the direction of hypermorality.

Kant's demand level is high. His critical philosophy denied the possibility of getting to know truths about God, the soul, and the origins of the world, but impressed on readers the inevitability of certain beliefs and experiences and the ineluctability of certain duties. We *must* perceive in space and time, and we *must* apprehend causes

and effects. We *must* believe in God, hope for the world to come, and submit proposed courses of action to the universalizability test. According to his anthropology, humans have tendencies both to passivity and to aggressiveness, in so far as they are ruled by an inclination to indolence and by self-interest. Yet because they are endowed with a faculty of reason that enables them to conceive themselves as partially exempt from the laws of nature, they can be roused to an admiration of pure moral agency and respect for the moral law.

Kant was concerned not only to dissociate moral motivation from inclination and the pursuit of happiness, but virtuous practice and its acknowledgement from social display and social reward. ‘Through all the ills and torments of life, the path of morality is determined,’ Kant intones. ‘No matter what torments I have to suffer, I can live morally.’¹ Moral living can even mean not living. While the *Foundations of the Metaphysics of Morals* portray suicide as a kind of self-contradictory action, Kant maintained in his early *Lectures on Ethics* that it is better to choose death than to consent to life as a galley slave.²

Both Kantianism and utilitarianism posit an overall obligation to consider the interests of others that is independent of natural sentiment and independent of the rewards altruism may bring. The source of that obligation was held by Kant to be our rationality and his point was revived by Thomas Nagel, who insists that we are susceptible to the claims of others because, as rational creatures, we do not occupy a personal point of view exclusively. Where sympathetic creatures can adopt the perspective of another person and take her interests into account, rational creatures can adopt an impartial perspective and take into account anyone’s or everyone’s interests. An agent operating from a purely selfish platform is no easier to understand than a hypermoral agent. How can an agent who values *G*—as evidenced by his willingness to deprive another of *G* in order to have more for himself—fail to regard it as good that the other should enjoy the possession of *G*?

Rational creatures are thus responsive to what Nagel terms ‘agent-neutral reasons’. As he expresses it, ‘From the objective standpoint, the fundamental thing leading to the recognition of agent-neutral

¹ Kant, *Lectures on Ethics*, 156.

² *Ibid.* 155.

reasons is a sense that no one is more important than anyone else . . . [T]here is not a significant reason for something to happen corresponding to every reason for someone wanting to do something.³ At the same time, the claim that we can and sometimes do take agent-neutral reasons into account does not entail either that we ought always to do so, or that we can do so easily, as Nagel goes on to argue. The impartial standpoint that recognizes agent-neutral reasons is more sophisticated, learned, and comprehensive than the personal point of view. It is more philosophical. Yet it exists, so to speak, on all fours with our responsiveness to other reasons. Kant thought of the moral motive as a constantly operating noumenal force competing with empirical drives and instincts; or rather, he insisted that we are impersonally required to picture it as such.

Arguably, Kant's scheme was confused in its representation of the moral motive both as a real force and as an entity of reason; as non-contingently present in every human being and as threatened by our animality. Modern Kantians are inclined to see these tensions as productive. Meanwhile, critics of Kantianism, amongst them, internalists influenced by Hume, are dismayed by Kant's apparent scorn for the idea that a substantial investment in non-moral goods could be worthwhile. 'The most thorough and readily available medicine', Kant says, 'for soothing any pain is the thought, which can well be expected of a reasonable man, that life as such, considered in terms of our enjoyment of it, which depends on fortuitous circumstances, has no intrinsic value at all, and that it has value only as regards the use to which we put it, the ends to which we direct it.'⁴ The metaphysical notion of the good will as a protective device that the subject carries about with him everywhere like a talisman seems . . . superstitious.⁵ Ordinary agents cannot help causing some wear and tear, on others as well as on themselves, as they try to live their lives. The thought that

³ Thomas Nagel, *The View from Nowhere*, 171–2. Nagel has always defended the importance of 'impersonal moralities with universal pretensions', amongst which he includes utilitarianism and some forms of Kantianism. Ibid. 199. See also *The Possibility of Altruism*, *passim*. Altruism, on his view, is not a mere possibility, but it is not a consistently powerful and actual motive in human affairs either.

⁴ Kant, *Anthropology*, 107.

⁵ 'The Sumatrans', according to the 18th-cent. traveller William Marsden, 'are firmly persuaded that particular persons are, what they term "betuah" (sacred, impassive, invulnerable, not liable to accident); and this quality they sometimes extend to things inanimate; as ships and boats. Such an opinion, which we should suppose every man might have an opportunity of bringing to the test of truth, affords a humiliating proof of the weakness and credulity of human nature.' Marsden, *History of Sumatra*, 293. Kant refers to the history several times.

our enjoyment of life is intrinsically without value probably does not have the medicinal properties Kant claims for it.

Utilitarianism is the second great exigent moral theory. Bentham presents it as based on a simple yet comprehensive account of human motivation, and he displays his main theses—the irrelevance of motives, the supreme importance of outcomes—in a dry and factual manner. His thesis that ‘the principle of utility is capable of being consistently pursued’ is doubtful. In unreconstructed utilitarianism, I am actually obliged to suffer; for my actions must increase overall utility if they are to be morally correct. I am obliged to submit to and suffer the depredations of others and to perpetrate direct harm to them if, for some reason, doing or suffering evil will increase overall utility. This consequence sits uneasily with the claim that we are entirely governed by pain and pleasure. Bentham, to be sure, assumes that, in order to be eligible for utility-promoting punishment, a person must actually have committed a crime.⁶ He does not seem to consider cases in which deliberately inducing some quantity of undeserved suffering relieves a greater quantity of suffering. Nor does he regard everyday actions by individuals that do not increase the general welfare as immoral or proscribe them. Yet classical utilitarianism is an inherently interfering moral system. It evolved hand in hand with a belief in the need for scientific solutions to all manner of social non-performance. It called for expensive public works projects that would intrude into the lives of formerly free-living individuals with the construction of an array of jails, workhouses, and other allegedly beneficent institutions.⁷ Bernard Williams wryly refers to utilitarianism as ‘Government House Morality’ and his criticisms have unmistakable parallels with the postcolonial critique of totalitarian regimes of the left as well as the right. The collective good is understood in postmodern times to be an elusive notion, and the sacrifice of individuals in its name is no longer regarded as politically acceptable.

⁶ Bentham, *Principles of Morals and Legislation*, ch. 13.

⁷ On utilitarian experimentation in India, see Eric Stokes, *The British Utilitarians and India*. ‘The idea of some sudden sweeping transformation of Indian society, of an entirely new system of law to be constructed in the space of a few years, of a new judicial and administrative machinery under which India would be propelled at a bound from feudal darkness into the modern world—this sort of attitude would only flourish in an age brought up to believe in sudden conversion.’ *Ibid.* 242.

Contemporary normative theories have dispensed with Bentham's bricks-and-mortar approach to the exact specification of institutions, as well as with the strictness and purity of the principle of utility and with supporting metaphysics of the Kantian variety. It is no longer considered necessary to establish free will or a human exemption from the laws of nature in order to present a theory of obligations, and rights and permissions may be introduced to improve the palatability of utilitarianism. Their internal coherence and level of argumentative precision make the new versions technically superior to their predecessors. Even in their updated versions, however, the famous modern moral theories appear to a sector of their readers to suffer from the same underlying problems as their predecessors. The relationship between natural sentiment and moral obligation is unsatisfactorily developed in Kantianism. It is positively incoherent in unreconstructed utilitarianism, which implies that creatures who naturally hate pain are obliged to undergo undeserved suffering for the greater good. Life, whether in neo-Kantian systems of autonomy and universal pronouncing or neo-Benthamite theories of resource allocation, tends to look constrained and regimented. It is easy to sympathize with Williams's complaint that, with Kant and Bentham, 'One is left, at any level of importance, only with purely moral motivations and no limit to their application', and that 'there is, at the end of that, no life of one's own, except perhaps for some small area, hygienically allotted, of meaningless privacy'.⁸

Williams is not reporting on what Kant and the utilitarians have between them accomplished. His worry is not that, because of the promulgation of Kantian and utilitarian doctrine and the positive reception that they have met with, everyone's life has been diminished. He is committed to the claim that our lives could never be completely occupied by morality, except through the imposition of some unimaginable tyranny. He is addressing the representational content of these theories. We can imagine living as Kantians or utilitarians and, according to Williams, this would be bad for us.

Wherever exigent moral theories have been less than pernicious in their effects, critics claim, they have been useless. The sceptic insists that theories are at best epiphenomena of positive social change, not

⁸ Williams, 'Moral Luck', 38.

its motor. They may be genuine responses to the perception of moral inadequacy, but they may at the same time subserve it by expressing a complementary utopian vision to which ideational allegiance can be given in the absence of practical undertaking.⁹

The problem of exigency is expressed as a dissatisfaction with both the form and the content of the famous modern moral theories. The elevation of unconfirmed authored norms to authorless imperatives rooted in mind-independent reality is rightly regarded by their critics as mystificatory. At the same time, their prescriptions are seen as too interfering and repressive where individual lives are concerned, and as provoking moral scepticism by setting standards of benevolence and self-sacrifice that normal human beings cannot ordinarily meet. Attention, meanwhile, to the difference between the first-person and the third-person standpoint has provided the philosophical underpinning for lower-demand normative theories.

3.2. The First-Person Standpoint

Certain fixed and irremovable features of our cognitive and emotional make-up, it is sometimes argued, nullify the formulas of obligation that the exigent moralists urge on us. It matters that *I* am the one asked by the utilitarian to relinquish a large sum of money to feed hungry refugees, or to kill one person to save twenty. It matters that *I* am the one allegedly required by the Kantian to forgo the face- and situation-saving lie, or to keep the promise I devoutly wish I had never made. Certain things are unreasonable to ask of me, even if the outcome they happen to produce is judged to be better than the relevant alternative. How can it be morally incumbent on *me* to murder one person to prevent twenty others from being murdered, or even to prevent twenty other murderers from carrying out their plans to murder one person each? And if I am not always required to bring about the best state no matter what moral sacrifices it might entail for me, why am I always required to bring about the best state no matter what non-moral sacrifices it might entail for me? As Williams observes, ‘There is no limit to what a given person might

⁹ For a sceptical view, see Dwight Furrow, *Against Theory*, 116–17. For a more favourable estimate of the role of moral ideation, see David Brion Davis, *The Problem of Slavery in Western Culture*, esp. 292 ff.

be doing to improve the world, except the limits of time and strength.¹⁰ He accuses utilitarians of aiming to ‘increase a sense of indeterminate guilt in their readers’, a tactic that, he points out, is ‘likely to be counterproductive and to lead to a defensive and resentful contraction of concern’.¹¹

Subjective perspectives, it has been argued—what it is like to be *x*—are not fusible with objective perspectives.¹² As surprising as this might seem in light of his views on the inexorability of the moral law, these claims derive historically from Kant’s idea of a gulf between the moving objects of physical nature and human agents, between things and persons, the phenomenal and the noumenal world.

The first-person standpoint has been crafted by several centuries of literary practice in techniques of introspection and autobiographical narrative. It would be rash to conclude that all persons everywhere experience themselves in the same way and assign the same importance to their own projects. The few remaining hunters and gatherers may have fully human lives without experiencing the pull of pecuniary motives or having any interest in what the sociologist Albert Hirschmann designates as obituary-improving activities, such as the collection of citations and prizes. Nevertheless, the environment, as phenomenologists used to emphasize, presents itself for everyone as a field for action, an array of possibilities and obstacles, or even as an array of obstacles presenting possibilities. My appetites and desires contour the world that appears to me, determining what is figure and what is ground. If I try to assume the perspective of a detached observer of the human race, possibility and freedom seem to vanish. When I watch a colony of ants moving about on the forest floor, some seem to be scurrying randomly hither and thither, whilst others are engaged in definite tasks, lifting and dragging objects many times their size. I can see that their success and failure depend on certain physical parameters—an ant’s strength, the size and shape of the objects it is trying to budge, how much help it is getting, and its luck. An ant can get stuck, or take on a task that proves to be too

¹⁰ Williams, *Ethics and the Limits of Philosophy*, 77.

¹¹ *Ibid.* 212 n. 7. The same narrowing of concern was predicted by E. O. Wilson, *On Human Nature*, 195.

¹² ‘We will not know how scrambled eggs taste to a cockroach even if we develop a detailed objective phenomenology of the cockroach sense of taste. When it comes to values, goals, and forms of life, the gulf may be even more profound.’ Nagel, *View from Nowhere*, 25.

much for it. A detached observer might see human activity in this way, partly as pointless scurrying and partly as involving endeavours whose eventual success or ultimate futility is mostly determined by factors outside the control of the busy agents. However, this is not how we experience life from the inside, and it is difficult to see why the objective standpoint should be regarded as more authoritative. My desires make the world appear to me as it does; my project is something I cannot consistently relativize to the projects of the other, for they are they and I am I: the Other appears as an obstacle or an avenue for me, depending.

As Kant sighed at one point, 'Man is not so delicately made that he can be moved by objective grounds',¹³ and as Hume crowed, '[T]here is no such passion in human minds, as the love of mankind, merely as such, independent of personal qualities, of services, or of relation to ourself'.¹⁴ The cost to me of actions that benefit others is often high from my first-person standpoint. The actions that I deem to be good from an objective point of view often require me to inhibit my spontaneous and pleasurable impulses (and spontaneous behaviour just is pleasurable), or to exert myself in ways to which I am not accustomed (and unpractised actions are by and large unpleasurable). Not only is it difficult to act towards them exactly as I wish they would act towards me, or to act in ways that benefit others at my expense, it is difficult to adopt the objective perspective at all, for it is psychologically disquieting.

On simple utilitarian principles, spending money in expensive restaurants and on high-priced entertainment and luxury articles is forbidden. And this is not a peculiar deduction of utilitarianism. A Kantian too would will that it would become a law of nature that, when A_2 could be saved from starvation for a year by A_1 's forgoing a dinner, A_1 would forgo the dinner. Yet the gulf between theory and acceptance, and between acceptance and action, can seem enormous. For Nagel, our ability sometimes to assume the objective posture results in a permanent state of tension between the good life and the happy life that usually cannot be erased by deciding to limit one's aspirations to one or the other.¹⁵ I can recognize, adopting the objective point of view, that certain outcomes are morally good

¹³ Kant, *Lectures on Ethics*, 46.

¹⁴ Hume, *Treatise*, III. 11. i. 481.

¹⁵ Nagel, *View from Nowhere*, 189–207.

ones, and that they are within the capacity of people like me to advance. Still, my own concerns, my comforts and discomforts, weigh more heavily with me than they do with a randomly selected stranger, and I weigh my own comforts and discomforts more heavily than I weigh those of a randomly selected stranger.

Partiality towards our own causes, like our spontaneous reactions of gratitude and resentment, appears to be resistant not only to extinction but even to modification. It might seem that in a limited range of cases we can extend our concerns in the direction of impartial benevolence, but always for a specific reason and not because the action performed accords with a formula of obligation derivable in a moral theory. Indeed, the influential notion that objective normative assessments are incommensurable with first-person normative assessments entered metaphysics with P. F. Strawson's celebrated discussion of reactive attitudes.

Strawson was concerned about the encroachment of scientific ways of thinking into normative frameworks. Regarding the spadework in separating the human realm from the realm of inert nature as already accomplished by Kant, and as issuing in an ontological distinction between 'persons' and 'things', Strawson insisted that we could never abandon our commonsense view of persons as for the most part responsible agents, no matter how irrefutable philosophical arguments for determinism might appear. Mere arguments, even if we could not answer them by showing how free agency was possible, could make no dent in our normal and culturally universal view of persons as the source of happenings in the world compelling moral evaluation.¹⁶ We could, Strawson conceded, occasionally see persons as 'things', as momentarily exempt from attributions of responsibility or even as exempt in the long term and as candidates for rehabilitation or treatment. Our reactive attitudes might be suppressed by detailed knowledge about some particular case. We could never, though, suppress the automatic habit of reacting to persons with resentment and appreciation for their offences and kindnesses. There are limits, Strawson argued, to the changes in our basic commitments and practices we can make for purely theoretical reasons.¹⁷ Analogously, it might seem that while I might forgo a restaurant meal and send

¹⁶ P. F. Strawson, 'Freedom and Resentment', 83.

¹⁷ *Ibid.* 84, 95–6.

money to a charity because a television appeal or an account I read in a magazine moved me, I will never do so because I have come to believe that utilitarianism is true or irrefutable. At best, I might offer a utilitarian principle *post hoc* by way of explaining or exalting my motives.

The claim that certain dispositions are invulnerable to revision from certain sources corresponds to psychological and social reality at several levels. As living creatures, we are disposed to resist interference with our projects, with our pursuit of desired objects, and with challenges to the veridicality of our own perceptions. We are reluctant to embrace knowledge that would deprive us of the conviction that we perceive the world accurately and understand why things happen in it, that it is just, and that our conduct is innocent. As affluent moderns, we are disposed to resist any forces that would, really or symbolically, take away any of the things we enjoy. Strawson was able to convert what the Stoic moralist considered a lamentable weakness, my inability not to react with irritation when a clumsy person steps on my foot, into a kind of virtue. His text was implicitly prescriptive; it suggested to the reader that his ordinary reactions of resentment—perhaps even his tendency to vindictiveness—were elements of a form of life that did not intersect with the practice of metaphysical inquiry but that had its own importance and dignity. It held out a permission—an exemption—to carry on resenting, praising, and blaming. Analogously, what at first appears to be a psychological defect—my inability to treat my own preferences and desires as on an equal footing with everybody else’s—can be converted into a positive attribute. This disability or defect is a sign that I am a real human being with a unique and interesting self, not a robotic saint or a mechanical utility-computer.

Various proposals have been made for reducing the discrepancy between ordinary inclinations and classical utilitarian mandates. Samuel Scheffler proposes that a moral agent be permitted to include in her utility calculations the benefit to herself of performing a given action, and loosens the requirement that she make all her calculations from a third-person standpoint.¹⁸ For example, it is more valuable to

¹⁸ '[A] moral point of view gives sufficient weight to . . . [the independence of the personal point of view] only if it reflects it, by freeing people from the demand that their actions and motives always be optimal from the impersonal perspective.' Scheffler, *The Rejection of Consequentialism* (2nd edn.), 20. Cf. Nagel, *View from Nowhere*, 171 ff.

X to save her daughter from drowning than it is to her to save a stranger. The utility to her of saving her daughter, calculated from her perspective, would likely outweigh the utility to her of saving two strangers. Therefore, a modification might relax the prima facie utilitarian requirement on *X* to maximize happiness by saving one or two strangers rather than saving only her daughter. To be sure, there is a limit to the use of this strategy. If agents' estimations of utility to themselves are allowed by the prescriptive moralist to weaken prima facie obligations as these stand under unreconstructed utilitarianism, to just the degree they like, these alleged moral subjects no longer seem to operate within a normative system at all. In the limiting case, the formulas of obligation to which they respond are identical with the actions that will give them the most personal satisfaction. Any revision to the standard theories must retain a sufficient level of demandingness to avoid this *reductio ad absurdum*.

This goal can be achieved by recognizing only small agent-favouring departures from utility-maximization as legitimate. The needed prescriptivity can also, however, be introduced by raising the normative bar elsewhere in the theory. Standards of non-moral achievement—aesthetic, affiliative, or economic—that are quite demanding may be imposed on agents as their moral requirements are relaxed. Criticisms of Kantianism and utilitarianism often allude to what in earlier times were considered non-moral claims, such as the claims of children, spouses, friends, or one's vocation, for time and attention, treating them as focuses of devoted obligation.

Williams, for example, attempts to show that a person's aesthetic project can have an importance that renders him impervious to certain kinds of criticism. 'Gauguin' arrives at the idea of going to Tahiti to start a new life as a painter of tropical beauty and beauties. He worries about the decision before making it, but he does not, in Williams's story, query his plan to determine whether it satisfies the test proposed by Kant—that he could will it as a universal rule that everyone in his situation should do likewise. He does not make a serious effort to determine whether the benefits he can produce will outweigh the unhappiness his actions will cause. He pursues his project and goes to Tahiti, where things turn out well for him. He becomes a celebrated artist whose work gives pleasure and inspiration to many, so in the end he does produce a good deal of benefit for others as well as himself. Things do not turn out badly for those

affected by his decision. Williams defies the reader to say that ‘Gauguin’ did something wrong, not just in going to Tahiti but in failing to submit his proposed course of action to theoretical review.¹⁹ It is irrelevant that such a review might have turned out in ‘Gauguin’s’ favour, that it might be easy to defend his decision on simple utilitarian or even Kantian grounds. For Williams, any assessment of ‘Gauguin’s’ situation in terms of moral theory would have been an imposition. What he refers to as ‘the morality system’, not just some specific duty, is the focus of his scepticism. We cannot, he thinks, assume that there is ‘some currency of satisfactions, in terms of which it is possible to compare quite neutrally the value of one set of preferences together with their fulfillments, as against a quite different set of preferences... [W]e... cannot in principle gain a standpoint from which the alternative fillings of our life-rectangle could be compared without prejudice... The perspective of deliberative choice on one’s life is from here.’²⁰

Exactly what the presentation of this example is intended to show or does show has been the subject of much discussion. It hints at several specific moral issues, chiefly, financial responsibility for dependent spouses and children and sexual exclusivity, both of which *prima facie* duties involving self-sacrifice for the benefit of others ‘Gauguin’ repudiates. On the assumption that the example concerns these putative obligations, does it show that success at some non-moral undertaking compensates for a moral delict? Or that it is tyrannical to require agents to put all their actions to some impersonal test? Or that the future is so unpredictable that the thought experiments and calculations required by moral theorists as a preliminary to action cannot be performed? Does it question the right of anyone besides ‘Gauguin’ himself to judge his life? Any of these theses might be extracted, but the claim that Williams appears most interested in defending is the claim that if a person were to behave as his fictional Gauguin does, and if the outcome were to be as it was in the case of the real Gauguin, there would be no basis for saying that she acted in a morally irresponsible manner.²¹ In accord with the hypothesis that

¹⁹ Williams, ‘Moral Luck’, 20–39.

²⁰ *Ibid.* 35.

²¹ The historical Gauguin’s departure was far from impulsive or unpremeditated. Though the plan he unveiled in a letter to his wife to ‘flee to the woods on a South Sea Island, and there live in ecstasy, in peace and in art... with a new family, far from the European struggle for money’ is expressed rather

exemptions and permissions relative to background theories of morality tend to be compensated for by increased demandingness elsewhere in the system, Williams places a burden of artistic achievement on his hero, and saddles him with risk; if his project does not succeed, he may be hounded by regret. Because, the implication is, to live a good life one must normally achieve certain goals with respect to self-expression, the forging and maintenance of personal relationships, and the contribution of something valued by others in the world, Gauguin is not exempt from all normative requirements.

The position that while philosophical argument cannot justify selfish pursuits, it can show us that, when properly compensated by non-moral achievement, they do not stand in need of justification has been explored by other writers. Susan Wolf insists that we need to recognize ‘the normal person’s direct and specific desires for objects, activities and events that conflict with the attainment of moral perfection’.²² She lists a number of activities that consume time and money, participation in which precludes the exercise of charity, that are worthwhile and appropriate. According to Wolf, no philosophical theory can tell us how important moral goodness is compared with other forms of goodness, or how much money, time, and effort we ought to devote to other-directed moral pursuits as against selfish pleasurable pursuits; consequently ‘the posture we take in response to the recognition that our lives are not as morally good as possible need not be defensive’.²³ Wolf does not claim that it is morally permissible for anyone to dine on champagne and caviar when there is famine in Ethiopia. Nevertheless, in the discursive context—in which it is a *prima facie* assumption that actions ought to be regulated according to utilitarian criteria, with money spent where it will produce the most marginal utility—the illocutionary force of the argument is the offer of a permission or an exemption from a putative obligation.

shamelessly, the historical Gauguin was self-consciously aware of a conflict between personal needs and social demands. In a book composed for his daughter, who was distressed by his departure, he wrote, ‘You sacrifice yourself for your child, who in turns becomes an adult and will sacrifice himself. And so on. There will be nothing but sacrificing people. And this will go on for a long time.’ David Sweetman, *Paul Gauguin: A Complete Life*, 341.

²² Susan Wolf, ‘Moral Saints’, 424.

²³ *Ibid.* 435–6.

Nagel, finally, displays a certain ambivalence towards his own discovery that selfishness is irrational. ‘Suppression of the full force of the impersonal standpoint’, he asserts, ‘is denial of our full humanity.’²⁴ Yet he places only limited confidence in the motivating power of agent-neutral considerations. He emphasizes how difficult it is to maintain the impersonal standpoint and consistently to regard our own interests as no more deserving of satisfaction than anyone else’s. For the later Nagel, neither the personal nor the impersonal perspective can be considered privileged. We are somewhat susceptible to arguments from the objective standpoint. We are sometimes moved. To deny that the impersonal standpoint has any authority would be to embrace irrationality. But if someone maintains that we ought always to be moved or that the third-person perspective has a built-in authority that the first-person perspective does not, he owes us an account of why this should be so. Such an account cannot be given in non-moral terms. The third-person perspective is essential in moral theory not because it is specially rational, but because to assert an obligation to do what is objectively best, rather than what one can do, given one’s particular powers and situation, is often to renounce advantage.

Many criticisms of deontological and utilitarian frameworks have in common an implicit appeal to the costs that are incurred by agents when they conform to exigent advantage-reducing policies. An increasingly used tool in the prescriptive moralist’s kit is the argument from heavy costs, and it is to a discussion of that argument that I now turn.

3.3. The Argument from Heavy Costs

An important principle of descriptive moral psychology is that informants tend to judge that the physical and psychological costs to an agent of carrying out a *prima facie* obligation reduce the strength of the obligation. According to the principle, a morally good action that is extremely hard to carry out will tend to be perceived as supererogatory and optional for agents. More precisely:

²⁴ Nagel, *Equality and Partiality*, 20.

Heavy Costs Principle

(1) Subjects usually regard the difficulty of fulfilling a formula of obligation prescribing an action *ACT* in context *c* as tending to weaken its authority.

It needs to be emphasized from the outset that the heavy costs principle is a statement belonging to moral psychology, not prescriptive moral theory. It is a claim about how we judge, or perceive the world, or think about our obligations, not about what we are justified in doing. That said, the principle that the costs to the agent of assuming an obligation are perceived to diminish the strength of the obligation evidently applies quite broadly to all manner of social and political obligations. It is a social rule that one ought to reciprocate invitations. The rule is, however, responsive to the assumed capacities of people in specific circumstances to conform to it. One who lives in a grand house with a chef and servants and never reciprocates dinner invitations will be considered miserly and be blamed for it, whereas one who cannot cook and lives in an apartment with only a hot plate will be seen as only weakly obliged to return invitations. A member of the underground resistance who blurts out names under torture, violating the rule to keep the party's secrets, will be judged less harshly than one who sells them for money, even if she needs money quite a bit. 'Ought' thus appears to be a scalar concept: My perceived obligation to perform a given action can be strong or weak, depending on my circumstances.

'Can' is also a scalar concept, as Björn Eriksson points out.²⁵ I can do some things easily, others only with difficulty, and this applies to the psychological as well as the physical realm. It is easy for me to clean my living room, given a whole day in which to do it. With some effort, I can clean my whole house in one day, but this is not easy for me. It would be impossible, from an organizational and physical point of view, for me to clean the 500-room office building in which I work in one day. Given my psychological dispositions, it would be difficult for me to work in a public relations firm, but it would be utterly impossible for me to work on a catwalk building an eighty-storey skyscraper.

²⁵ Björn Eriksson, *Heavy Duty*, 70 ff., 163.

Moral obligations too are perceived as scalar and appear to be diminished when an agent is unable to perform some action or could perform it only with great difficulty. If an agent cannot swim at all, his failure to save a drowning child is seen as tragic but not culpable. If he can swim, but does not save the child because she is very far out on a frigid, storm-tossed ocean, and it would be all but impossible to save her, his not saving her is still regarded as tragic and not culpable. Culpability is not usually assigned by observers until an unperformed good action gets within range of the agent's actual capacities. If the agent can swim fairly well, and if the child is not too far away, but he feels that someone more vigorous ought to take on this rescue opportunity and does nothing, observers will judge him as morally deficient. Being a little tired is not thought of as a defeating condition of a *prima facie* human rescue obligation. There is a range of degrees of enablement that are related not just to physical capacities, but also to the strength of emotions, and the strength of ingrained habits, that affect my ability to perform any task. In deciding whether to blame or excuse, we consider these capacities.

Two factors appear to influence observers' assessments of the level of obligation in particular contexts. The first, just as utilitarians might hope, is the benefit that an action in conformity with the formula of obligation will bring about, or the harm that it will prevent. The second, however, is the cost or burdensomeness to the agent of conformity with the formula. These factors offset one another. There are many benefits that an agent could bring about with some effort that she is considered only weakly obliged to bring about, and many harms she could prevent with great effort that she will be considered only weakly obliged to prevent. His rich uncle, Eriksson judges, is more obliged than he is to endow a hospital that will save the lives of 500 persons in a poor country because it would be easier for the uncle to do. Eriksson could sell his house and move to a block of flats in order to give away the money, but psychologically this is hard for him even to contemplate.²⁶ One is likely to judge that someone busy raising a family of five small children is less obliged to spend an evening attending a letter-writing session to free political prisoners than someone who lives alone. Leaving aside Eriksson's

²⁶ *Ibid.* 161.

prescriptive determinations, it is clear that our spontaneous judgments at least tend in the direction he indicates. Agents are normally considered strongly obliged to bring about a large benefit at low cost to themselves, weakly obliged to prevent small harms at large costs to themselves, and moderately obliged to bring about middle-sized benefits or prevent middle-sized harms if it is somewhat burdensome for them to do so.

Perceived burdens are subject to multiplier effects in accord with the following psychological principles.

(1) *Relative Deprivation Principle*

Burdens that the agent knows are shared by others in the group with which he compares himself are diminished; burdens experienced in isolation are magnified.

If everyone else in one's reference group is obliged by circumstances to eat fish during the month of February, it is less burdensome not to have meat. It is grating, by contrast, not to be able to take a vacation when one's co-workers can, to wear shabbier clothes, and to serve plainer fare at your table than others do.

(2) *Before the Revolution Principle*

Burdens that the agent did not formerly experience are magnified; burdens that the agent thinks she will soon be relieved of are diminished.

It is humiliating to be defrocked, to suffer a cut in salary, to move to inferior accommodations, or suddenly to find oneself in the midst of a legal battle when all was going well. It is less burdensome to be born into a harem than to be captured and put into one. Conversely, if one believes that things will get better soon, that new opportunities beckon, that assistance is on the way, burdens are easier to bear.

(3) *Externally Imposed Hardship Principle*

Burdens that are imposed by someone else are magnified; burdens that are the by-product of the agent's own projects or the result of the agent's deeper instincts are diminished.

It is not hard to accept deficits such as being a junior in a professional hierarchy that are the result of your own choices and fit into your own plans. The trouble involved in taking care of a baby when one

expected or hoped to have a family is usually cheerfully accepted. Babies would by contrast be regarded as an offensive imposition if they were deposited on the doorsteps of some households selected by lottery, legally requiring to be raised to maturity.

Information about burdensomeness to the individual, as well as information about overall utility and disutility, affects our propensity to blame.²⁷ The rich man who engages in petty deception by hiding his extra bottle of liquor from the customs officer, or the academic who uses departmental postage to mail her credit card payment may be thought reprehensible, even if the offence is small, for it would have been easy to conform to the regulation.

The finding that burdens are judged to offset obligations at the same time as the utility of outcomes is judged to justify the imposition of burdens is explanatory in two respects:

- (a) It predicts that high-demand formulas of obligation will often be judged inapplicable to a particular case or generally inappropriate.
- (b) It predicts that motivation, self-reproach, and blame will admit of degrees, according to how hard it is thought to be to perform an action and how morally desirable the action is.

Extremely high-demand theoretical statements, such as the claim that it is never under any circumstances permissible either to contribute to or to fail to prevent the occurrence of pain in another human being, except in the service of that person's physical health, are very difficult to confirm, though individuals may be deeply committed to authored norms such as pacifism and global benevolence. This is simply because it is hard for individuals and groups to pursue their aims and get on with their lives without neglecting some human suffering and often without contributing to it. Manufacturing and selling automobiles, for example, causes numerous injuries and fatalities and is not undertaken for altruistic reasons. One may reasonably hold the view that automobile travel is more dangerous than it ought to be and that the greed of manufacturers who will not cut profits to save lives is objectionable. But competent judges will not rule out as morally illegitimate any human activity whose practice has ever been found

²⁷ Bjorn Eriksson, *Heavy Duty*, 182 ff.

to have made a causal contribution to the suffering of a human being. We can also appreciate that, even if a particular formula of obligation can be confirmed by competent judges, agents may not experience the motivation to conform to it. Note that none of the items below entails that *I do ACT in c*.

I believe doing *ACT* in *c* is morally right

I believe the morality of doing *ACT* in *c* to be well confirmed

I believe I ought morally to do *ACT* in *c*

I believe everyone ought morally to do *ACT* in *c*

I believe that I ought morally to do *ACT* in *c* if I believe doing *ACT* in *c* to be morally right and doing *ACT* in *c* is morally right

I believe I will be seriously morally at fault if I fail to do *ACT* in *c*

I believe I should not commit serious moral faults and that I will commit a serious moral fault if I fail to do *ACT* in *c*

Etcetera

Normally, the belief that it is right and good to do *ACT* in *c* is paired with a motivating disposition to try to do *ACT* in *c*, for to experience moral motivation is to be aware of the coexistence of just such paired beliefs and actions, and manifestly we are animals who experience moral motivations. In so far as we do experience moral motivations, we cannot seriously entertain the possibility that a set of confirmed formulas of obligation is objectively binding on us, though we have no inclination, or only a weak and consistently ineffective inclination to act in accord with the formulas. Yet belief and motivation can on occasion come uncoupled without my losing my entire moral capacity, only my moral motivation in this case. There is no impossibility in my believing that considerations speak overwhelmingly in favour of doing *ACT* in *c* and yet not wanting to do it.

The argument from heavy costs is employed prescriptively whenever it is claimed that a proposed formula would be so difficult for agents in general to comply with and is so repugnant to their natural sentiments that it cannot possibly constitute a genuine obligation. The suggestion that one ought to give all one's income to charity except what is required for biological survival presumably falls prey to the argument from heavy costs. An argument from heavy costs may

also be employed to excuse a particular agent from compliance with a rule in a single case, even though the obligation is regarded as remaining in force. Some particular failure of veracity or fidelity or loyalty might be excused on the grounds that the special nature of his case made the corresponding duty hopelessly difficult or impossible for a particular agent to fulfil.

Prescriptive employment of the argument from heavy costs is necessary in moral argument but risky. It would be an elementary mistake to infer from the premiss that a given *prima facie* obligation is judged by someone to be weakened by the costs of performing the action to the conclusion that the obligation is nullified. It would also be a mistake to infer from the premiss that people generally found it hard to live up to some rule or bring about some outcome that it was not morally required. At the same time, it would be an elementary mistake to infer from the existence of a formal distinction between 'is' and 'ought' that information regarding costs is always irrelevant to the determination of obligations.

A rigorist might insist that moral psychology has no bearing on the question of our real obligations and that whether a defensive backlash is generated by the popularization of high-demand theories is not the concern of the moral philosopher. What we are naturally inclined to do and what we are obliged to do are two separate questions. If it has been established that we really ought to wear our old clothes until they fall apart and give up expensive entertainment, sending the money saved to charity, as Peter Singer argued we should in a celebrated article of 1972,²⁸ then we have to do that regardless of the psychological difficulty.

This position is, however, confused. The rigorist is right to distinguish between appearance and reality. He is right to insist that *if* it can be established that we ought to wear our old clothes, etc., then we really have to do that, even if our motivation is weak. He nevertheless underestimates the difficulty of confirming moral obligations. The exigency of an authored norm is its proponent's personal concern. By contrast, when a policy, for example the policy of sending all one's income except what is required for subsistence to a charitable organization, is presented as objectively binding, i.e. as confirmable, it has

²⁸ Peter Singer, 'Famine, Affluence and Morality'.

to gain the backing of competent judges who assess it in light of the reality constraint and the idealism characteristic. The rigorist is wrong to maintain that exigency cannot be a genuine reason, as opposed to a motive, for rejecting a particular formula of obligation, for one is not always being unreasonable in citing the heavy costs of meeting a putative obligation as reasons for not doing so. Accordingly, costs to agents can really disconfirm certain *ACT* propositions.

But can the argument from heavy costs furnish a reason for rejecting, not simply this or that formula of obligation, but all moral theorizing across the board? Clearly, Kantianism and utilitarianism are vulnerable to it. High-demand moral theories not only impose burdens, they multiply them. For not everyone is threatened with the demands of these modern moral theories, but chiefly a small Euro-American fraction. Their burdens are amplified according to the relative deprivation principle, for most inhabitants of our planet concern themselves very little with the problem of benevolence to strangers, living instead in blissful ignorance and indifference. Further, high-demand moral theories are a new imposition whose subjective costs are increased according to the before-the-revolution principle. Historically, they arose as a response to industrialization, in the wake of the emergence of contract relations between persons who did not know each other well, popular government, and the spread of objective knowledge about others. The appeal of ancient ethics, with its focus on relationships like friendship, into which the problems of mass society do not intrude, is easy to explain. And, finally, high-demand morality is not a by-product of deep instincts and dispositions, and its requirements weigh on us more heavily according to the externally imposed hardship principle. Kant's insistence that morality is a set of laws that rational creatures spontaneously prescribe to themselves is a noteworthy attempt to lighten our perceived burden by depicting high-demand morality as having an internal source. Yet he fails to convince us that his rules are confirmed and thus objectively binding.

The argument from heavy costs is effective against Kantianism and utilitarianism. If the reflection, ratiocination, and self-denial required to act as these theories prescribe strain our ordinary capacities, their prescriptions are inappropriate. For we cannot maintain that the theories are excellent and yet be unwilling to shoulder the burdens

they entail. In summary, prescriptive moral theorists must take into account the relationship between the formulas of obligation they advance and our ordinary or average capacities. A confirmed theory *M* cannot consist of formulas of obligation that are so exigent that only heroes, ascetics, omniscient beings, or persons devoid of worldly ambition, can do what is morally required of them. Nor can it be so obscure and complicated that ordinary human beings cannot understand what the theory says they ought to do. Facts about the first-person standpoint accordingly place limits on the substantive and formal demandingness of morality.

Nevertheless, the argument against exigent systems from the first-person standpoint is not as powerful as it appears at first glance. It slides from the unquestionable premiss that her first-person experiences and concerns are ineffably different for the subject who has them, to the weaker but still indisputable descriptive conclusion that first-person concerns and experiences are highly motivational, to prescriptive conclusions that are questionable. Its costliness to agents in general counts against any formula of obligation, but never decisively. The costliness of a policy *to me* is not relevant to its assessment, except in so far as costs for me would be costs for anybody else like me.

The heavy costs considerations advanced above do not threaten the enterprise of proposing, defending, and criticizing particular moral claims, or the generalizations that permit us to deduce particular formulas. In commending certain ways of life as appropriate for persons with our psychological constitutions and condemning others as inappropriate, critics of Kantianism and utilitarianism are contributing to the same theoretical enterprise as their predecessors. Chapter 4 introduces a more radical strain of antitheory, frequently conjoined with the argument from heavy costs, that does, by contrast, pose a challenge to the very idea of a moral theory. This challenge is argued to be unsuccessful.

A few predictions concerning the relations between meta-ethics and substantive moral theory can be ventured in the meantime. Costs to agents are frequently cited against the claim that a more egalitarian distribution of social goods is morally required. Now, there is no logical inconsistency involved in holding egalitarian or other high-demand moral views while subscribing at the same time to the heavy

costs principle. However, a commitment on the part of well-off people to greater social equality is obviously more expensive to them than a commitment on the part of the badly-off to greater social equality. It is accordingly difficult to instantiate the role of a well-off egalitarian who sets great store by the argument from heavy costs, and much easier to instantiate that of a well-off oligarch. Edmund Burke was perhaps the first great conservative to cite the heavy costs principle when, in his *Reflections on the Revolution in France*, he lamented the loss of ‘all the pleasing illusions which made power gentle and obedience liberal’ that had been ‘dissolved by this new conquering empire of light and reason’.²⁹ If our prediction is borne out, the heavy costs principle will tend to figure prominently in conservative thought.

Prescriptive egalitarians who, like the present author, accept the heavy costs principle, need a good deal more by way of theoretical apparatus to explain how egalitarianism can be a well-confirmed position. The challenge is addressed in Chapters 5, 6, and 7.

²⁹ ‘All the decent drapery of life is to be rudely torn off. . . . On this scheme of things a king is but a man, a queen is but a woman, a woman is but an animal, and an animal not of the highest order.’ Edmund Burke, *Reflections on the Revolution in France*, 67.

Limits on Theory II: Immanent Standpoints

Hume argued that virtuous behaviour is a set of performances through which a society maintains goodwill and easy communication amongst its members, fostering trust and reducing envy. Moral actions and responses arouse sentiments of approbation in observers who appreciate their usefulness. Morality comprises a system of encouragements and deterrents, discursive and practical, emerging from a central core of basic but non-theoretical agreement. It could be said to be immanent in human relations; it does not need to be imposed, in Hume's view, by specialized experts.

In his writings on virtue, Hume does not enunciate formulas of obligation, and neither arduous deductive systems nor ascetic exercises of self-denial pertain, in his view, to the establishment or maintenance of morals. 'Grace . . . ease . . . genteelness . . . must be considered as a part of ethics, left by nature to baffle all the pride of philosophy.'¹ He recites a list of character traits, including the dispositions to justice, fidelity, honour, allegiance, benevolence, humanity, clemency, temperance, sobriety, and chastity, explains their utility in enabling men to live together in confined spaces, and urges readers to cultivate them. The virtues are artificial, in that they require practice and polishing, but they are independent of ratiocination, and, once ingrained, they can be exercised without reflection or deliberation.

Because Hume's approach to ethics is so often cited as an alternative to Kantian and utilitarian prescriptivism, it is important to distinguish between the theoretical strain and the antitheoretical strain in Hume's moral discourse. One might suppose that a virtue theory like

¹ Hume, *Enquiry Concerning the Principles of Morals*, sect. 8. 14.

Hume's is only nominally a *theory*, that it is fundamentally opposed to an imperative-based theory like utilitarianism or Kantianism, formulated in terms of policies adopted and followed by agents. This impression is, however, misleading; virtue theory is an ordinary moral theory, subject to confirmation in the usual way.

In a Kantian paraworld, agents behave in a manner that we external observers can describe as falling under a universal rule. To stipulate that ideal agents not only behave in this manner but also consciously plan their actions to accord with universal rules is to advance a more complex theory, one even more difficult to confirm than simple Kantianism. In the competing paraworld of a virtue theorist, virtuous agents are ubiquitous. They too, however, act in what from the observer's perspective is a regular and reliable fashion. Faithful agents serve their masters loyally; chaste agents are selective in their *amours*; clement agents do not punish wrongdoers harshly. To possess a virtue is to tend to the production of some stereotypical set of actions. Is the behaviour of ideal virtuous agents, as conceived by virtue theorists, more improvisational, less predictable, than the behaviour of Kantian agents? There is no reason for it to be so. Though the virtue theorist will insist that her agents do not regulate their conduct by reference to rules, she is likely to conceive them as regulating their conduct by reference to character. Thus, Hume's agents are easily imagined as investing a good deal of time in moral introspection and in discussing and criticizing one another's characters. In that case, they worry, fuss, and obsess about being moral as much as cartoon Kantians and utilitarians do, only their concern is not with universalizing their maxims or advancing the general welfare but, according to Hume's somewhat cynical view, with maintaining their reputations.

Virtue theory and its relatives, then, are simply ordinary theoretical moralities; their assertions are confirmed or disconfirmed, to the extent that this is possible, in the same way as other theoretical claims. They do not pose a meta-ethical challenge to the very idea of a moral theory. Yet there is a radical antitheory strain as well in Hume which does pose a challenge to the very idea of a moral theory and which has been influential.

4.1. Immanentism

A radical critic of moral theory might insist that there is no point in moral theorizing by reference to alternative worlds and no possibility of expert confirmation of formulas of obligation. Williams represents this position when he says that ‘morality is not an invention of philosophers. It is the outlook, or, incoherently, part of the outlook, of almost all of us.’²

According to one conception of moral philosophy, which I shall designate ‘immanentism’, the moral philosopher’s role is to make visible and understandable the moral aspects of particular ways of life. He does not discover moral obligations, and moral progress does not consist in the invention of better moral theories that capture our obligations more accurately. For we cannot possibly confirm new and surprising moral claims by ratiocination and controlled experiment as we can discover new and surprising scientific truths. We must already have constructed as much morality as we need, for we have been living together for a long time. These ideas are memorably expressed in a well-known passage of Hume’s, through the image of two rowers who fall into an easy rhythm without a word having passed between them.³

Nor is the rule concerning the stability of possession the less derived from human conventions, that it arises gradually, and acquires force by a slow progression, and by our repeated experience of the inconveniences of transgressing it. On the contrary, this experience assures us still more, that the sense of interest has become common to all our fellows, and gives us a confidence of the future regularity of their conduct.⁴

A Humean might remind us that punishment and praise, criticism and appreciation, were directed at us, and at our instructors before us from our earliest years. The habits of sincerity and reciprocity have integrated themselves into our normal routines, so that to find oneself not behaving in a certain way can cause acute discomfort and anxiety.

² Williams, *Ethics and the Limits of Philosophy*, 174.

³ Hume, *Treatise*, III. II. ii. 490.

⁴ *Ibid.*

The rules we have internalized govern in the first instance our relations with familiars, but we also have rules for dealing with strangers, people who move into our orbits for the first time. Social feedback continues as we encounter more complex and delicate situations, and the actions and decision of agents alter normative standards in turn. Adherence to social norms involving sincerity, reciprocity, and respect for property rights is practical in terms of social energetics; transparency of communication and the performance of contracts reduces the amount of time needed to monitor social transactions and enforce agreements and frees time and attention for other tasks. These norms and their associated modes of inculcation and enforcement represent an efficiency that members of a social group can find their way to without external direction.

MacIntyre expresses this sense of pretheoretical rootedness when he insists that ‘I inherit from the past of my family, my city, my tribe, my nation, a variety of debts, inheritances, rightful expectations, and obligations. These constitute the given of my life, my moral starting point. This is in part what gives my life its own moral particularity.’⁵ We do not, on this view, choose our moral commitments by surveying a vast array of possible moral codes and assessing them in terms of betterness relations. Rather, we inherit a moral code and we make, in light of experience, small, piecemeal modifications to the shapes of our lives and those of our societies.

An immanentist may acknowledge that many of our actions do in fact conform to what utilitarianism or Kantianism prescribes, even while insisting that the moral theory articulated by philosophers is not a significant engine of personal or social moral development. No one seeks to maximize the satisfaction of all her non-moral preferences, even the ones that can be concurrently satisfied. Most people are fairly truthful and benevolent and will go out of their ways to minimize pain and suffering in their immediate environments. The reason for this conformity to advantage-reducing principles, the immanentist insists, is not to be found in exposure to academic moral discourse, but in our upbringing in communities in which moral practices were a subset of cultural practices.

⁵ Alasdair MacIntyre, *After Virtue*, 220.

In the view of the immanentist, our behaviour exhibits partial but only partial conformity with the requirements posited by the celebrated moral theories. Where we do extend our efforts and concerns outside the usual realm, or sacrifice our projects or submit to frustration of our desires, exhibiting unusual altruism or moral restraint, it is not because we have decided to apply 'theoretical morality in this case'. According to what might be called the non-theoretical reasons principle, departures from our usual preferences and habits or heroic self-suppression can only come about for internal reasons and not because of what an expert has shown it is right, theoretically, to do. This man gives away a fortune; that slave is manumitted; this female is allowed to hold an important political office. These occurrences are never explainable as consequences of a theoretical commitment to social equality; they depend on the perceptions and motives of individual actors.

Berating people for failing to behave as exigent moral theories say they ought to, or for failing to consult such theories before they act, is misguided, according to the immanentist, since these theories have no special authority. To show that agents who disregard the imperatives of impersonal moral theory are culpable or are acting in bad faith, one would have to establish that they ought to care, or have an external reason to care about governing their actions in accord with impersonal theories. Kant tried to show, Williams remarks, 'how the moral law can unconditionally apply to all people, even if they try to live outside it'.⁶ Yet Kant's constructions are unconvincing. Though we resent and blame offensive actions, we do not resent and blame any violation of Kantianism, utilitarianism, etc.

The immanentist seems able to account for the phenomena of morality, as most people experience them, without supposing that agents stand in any necessary relationship to moral theory. Framing and evaluating moral rules and distributive schemes appropriate to whole worlds, he insists, is an airy theoretical exercise, a game played with its own rules, that does not touch our moral experience. His position is similar to that of the Pyrrhonist sceptic who maintains that experience commits us to no particular theory of material objects; we can thus speak of 'moral Pyrrhonism'. It is supported by the

⁶ Williams, *Ethics and the Limits of Philosophy*, 191.

observation that impersonal moral theories have existed for only a fraction of the history of the human race, while actions and reactions we would not hesitate to describe as moral have existed for much longer. All past and present cultures have positive morality, but only a small fraction, have, in addition, contested theories of morals that purport to state, not only the rules themselves, but also the rules for making up or rejecting the rules.

Immanentism is a meta-ethical view that ought to be without prescriptive content. The knowledge that *N* believes moral theorizing to be useless and mostly uninteresting should carry no information as to where *N* stands on questions such as whether it is acceptable to eat chickens raised in cages or treat African-American applicants especially favourably in law school admissions. A theory-shy immanentist can be deeply concerned with social injustice—exploitative labour relations, sexual discrimination, and global economic inequality—just as a theorist persuaded by the heavy costs principle can be an egalitarian. Yet we can hazard the following meta-theoretical prediction: we will not observe many instances of immanentism or moral Pyrrhonism paired with revisionary ideas concerning social justice. This is not to say that the rejection of the modern moral theories is motivated by the desire of members of an advantaged class to retain their advantages. This would be an unwarranted assumption and tantamount to an unjustified *ad hominem* accusation. It is up to the sociologist of knowledge, not the philosopher, to determine what motivations or interests lie behind the promulgation of any doctrine.

Before considering the relations between antitheory sentiments and substantive positions further, it will be useful to explore in detail the preference of the conservative for a world conceived as evolving through the accumulation of minute, anonymous adjustments. For we can represent not only institutions and practices in a world, but morally significant changes in institutions and practices in a world. Revisionary theorists envision worlds that are quickly and dramatically improved, through the agency of prophets and the mass conversion of followers, or through enlightened lawgivers and the cooperation of a compliant populace.

Conservatives have a strong preference, not only for the kinds of worlds in which social distinctions are marked, but also for worlds

that take shape in a certain slow, undirected way. Plato is not a conservative theorist on this definition, for his preference for a hierarchically ordered social system without occupational mobility is countered by his understanding of how such a world comes to be, i.e. through the intervention of a *force majeure*. Marx's position is notoriously complex, for his textual presentation of a world slowly evolving towards equality through the unfolding of inner processes of development is widely appreciated as a rhetorical device for encouraging quick and dramatic improvement by prophetic means.

4.2. Fast and Slow Paraworlds

Consider 'Gauguin's' not merely forgivable, according to Williams, but creditable lack of interest in assessing his actions according to the specifications of an impersonal moral theory. How can his actions be read with respect to his broader social milieu? In Williams's paraworld, 'Gauguin' launches himself forward into a wholly new mode of existence, but family life in Copenhagen remains more or less the same. As pretheoretical core morality is supposed to provide a stable background for moral contests slowly and somewhat arbitrarily decided, so a social world symbolically held fixed provides the background for personal adventure.

Reflection on the sorry plight of the real Gauguin-in-Copenhagen might suggest to a reader that marriage is an institution requiring moral review. Perhaps lifelong obligations cannot justly be contracted in the heat of passion, and perhaps relationships requiring cooperative labour on a daily basis interfere with the aim of developing talents to the full. Perhaps marriage is a relic of some ancient institution and no longer has the same purpose, point, or justification in the modern world, and Gauguin was a harbinger.

This, however, is not the pathway Williams is urging us down in 'Moral Luck'. Elsewhere, he professes rather his admiration for Hegel's notion of *Sittlichkeit* over Kant's 'abstract' morality, praising Hegel's notion of 'a concretely determined ethical existence that [is] expressed in the local folkways, a form of life that [makes] particular sense to the people living in it'.⁷ Though Hegel was off the mark, he

⁷ Williams, *Ethics and the Limits of Philosophy*, 104.

thinks, in imagining a secret teleology of history guiding moral evolution, 'the Hegelian problem is the right problem at least to this extent; it asks how a concretely experienced form of life can be extended, rather than considering how a universal program is to be applied'.⁸ Williams argues that it is inadvisable to commit whole societies to potentially dangerous and irreversible programmes of reform; the leaps ought not to be too great, too frequent, or too sudden. We should not try to lock in any substantive set of values for future generations, he says, but only the value of free inquiry.⁹ Though individuals are exempt from any overarching requirement to remain rooted in a familiar well-tested mode of existence, and to modify it by degrees, larger groups, the implication is, are not. Taking on individual risk can be a good part of human life, and risk-prone individuals should not be held back to the norms of *Sittlichkeit*. Whole societies, by contrast, ought not to lurch into an unknown and unpredictable future. They, it seems, cannot be fenced in by their dominant discourses and practices; they cannot need to break out.

From the Hegelian perspective, actual states of affairs are right so long as they endure. Whatever is now, and especially whatever has been of long standing in human affairs, is exempt from justifications by comparison to what merely could be. The forms of life that we have adopted are, on this view, like Hume's coordinated rowing, the most natural and effective for us, and cannot be called to account. While modern immanentists distance themselves from some extreme forms of effective coordinated social practice to be found under the headings of militarism, imperialism, chattel or wage slavery, charismatic leadership, and sexual subordination, they for the most part do not perceive these patterns of social dominance operating within modern institutions. Where they do, they either deny their moral wrongness, or deny that much can be done to obviate it, unless history on the grand scale happens to be spontaneously tending in that direction. There is a kind of logic in the position that fast individuals need a slow society, even if slow societies can operate without fast individuals. Hegel was fascinated by the remarkable personality. The figural relationship of the Napoleonic individual to the ground of his culture leaves all its subgroups—its merchants, peasants, prisoners,

⁸ Williams, *Ethics and the Limits of Philosophy*, 104.

⁹ *Ibid.* 173.

aristocrats, housewives, blurred into the whole, unconscious of their agency or identity. They are part of the stable background; the farmers and fishers continue to farm and fish, while the generals change the boundaries of nations. In Williams's scheme, the wives continue to bring up the children, while the artists change the boundaries of aesthetic experience.

But if a form of life is not generally satisfactory, surely deprived agents will act to alter conditions? If we see two rowers rowing smoothly, are we not entitled to assume that they are both getting where they want to go? Ordinarily we can. Yet we cannot claim to see smooth rowing all around us. The appearance of harmonious coexistence or even harmonious cooperation between two agents or two groups who are in limited or imperfect interaction with one other may disguise accommodation to impermissible advantage-taking. When A_1 presses an advantage against A_2 , A_2 may respond in three ways, by ignoring what is happening, by acquiescing, in the recognition that his choices are constrained, or by openly protesting or resisting. Acquiescence, in turn, can take many forms. It can involve motivated ignorance, a decision to leave the terms of the A_1 - A_2 interaction unexamined, or the suppression of the impulse to grieve one's condition or to struggle towards the exit.

A third kind of acquiescence does not depend on a wilful refusal to ponder the A_1 - A_2 interaction, or the suppression of rebellious motives. It may call upon the active engagement of A_2 's intellect and imagination in constructing a rationale that appears to legitimize A_1 's advantage-pressing. Such a rationale may even take the form of a first-person narrative in which acquiescent A_2 represents himself as pursuing his own self-interest and as making deliberate self-advantaging choices along the way, choices that lead to the present state of interaction with benefit-appropriating A_1 . These narratives and the ceremonies to which they are attached contribute to cultural density in a way that can appear to furnish excusing conditions for prima facie injustice in the same way that glorious weddings may appear to their participants as well as to observers to erase the more repellent features of many arranged marriages.

The claim that social evolution is driven from inside, from collective wants and desires, not—it is implied—the theoretical insights of alienated geniuses, might be taken as an empirical claim

about how, despite a few frightening blips—the Terror, Stalinism—history has unfolded. It may also constitute a statement of reasoned authorial preference for a world that does not change very fast or very fundamentally. This can be seen from a consideration of the disqualification thesis.

4.3. The Disqualification Thesis

Some kinds of prescriptive theorizing are, in the eyes of critics of moral theory, morally inappropriate. No moral theorist, a Pyrrhonist critic might insist, is entitled to dictate, even through the projection of an ideal world in her own imagination, for the entire species. We are not in a position to pronounce on what others may or may not do or have. Even when no morally objectionable attitudes are involved in making judgements about other persons and groups, we are rarely qualified to do so.

Let's term the prescriptive belief that one ought not make certain moral judgements or assert certain moral claims because one is either not entitled or not qualified to do so the 'disqualification thesis'. Sometimes entitlement and qualification considerations are blended, when, for example, it is held that an intelligible moral judgement presupposes an argumentative context in which there is live dispute. While one might suppose that proponents of the disqualification thesis are concerned only with our entitlement and our ability to make moral judgements about persons geographically distant, or at least psychologically remote from us, this turns out not to be the case. Some proponents of the thesis insist that there are certain judgements we cannot make about our own culture. We can accordingly distinguish between the disqualification thesis (external) and the disqualification thesis (internal).

Before going on to criticize some applications of the disqualification thesis, it is important to distinguish between judging and intervening. Making a judgement and expressing it openly can be and often is an act of intervention. In deciding whether to intervene, by making judgements or in other ways, we have to decide how bad some action or practice really is, how difficult it will be to change it, what else of value might be lost, and how much force or persuasion

would have to be used. These considerations enter into any decision to interfere with another person's or group's settled habits and customs; they are as applicable to one's spouse as to distant strangers. The appropriateness of the judgement I make in such a context—and my qualifications or entitlement to make it—can nevertheless be distinguished from the effects of my asserting it to various audiences, including the audience the judgement is about.

It is one thing to insist that one should never interfere with the established customs of strangers, whether through the use of police or military force, or by telling them off or writing editorials against them. It is another to insist that one cannot or should not form any opinions of the morality of the conduct of strangers. In what follows, I shall ignore the question of intervention to focus on the simpler question of the appropriateness of moral judgements of strangers, including persons who are exotic or remote from us, historically or psychologically.

The first thesis to be evaluated is that, under the following conditions, an observer is disqualified from offering a condemnatory moral judgement with respect to a situation *S*, or with respect to the actions and responses of one or more agents *A* and patients *P*, the participants in *S*, even when *S* involves a set of interactions between *A* and *P* such that, were interactions of that sort to occur between a set of agents and patients in our local culture *W*, they would be judged morally defective. Gilbert Harman presents an instance of the general thesis when he states that we cannot judge of Hitler that it was wrong of him to have ordered the extermination of the Jews on the grounds that 'Hitler, like the cannibals, is outside our morality'.¹⁰

Conditions for Disqualification (External)

- (1) *S* is remote, historically or geographically, and/or *A* and *P* in *S* are psychologically remote, from *W*, the world inhabited by the prospective judges and its actors.
- (2) The *As* do not share the prospective judges' values and standards.

An additional adverse judgement-precluding condition is sometimes added (though not by Harman), namely:

- (3) The interactions between *As* and *Ps* were or are part of a way of life that is stable. They are woven into the fabric of

¹⁰ Harman, *Nature of Morality*, 109.

the culture and find resonance in other practices. They organize the life of the culture as projects organize the lives of individuals.

Harman suggests that an agent who is judged negatively has to be regarded as accepting a certain rule at the same time as he or she is perceived to have violated it. Hitler fails to meet this condition. For, in failing to endorse a prohibition on genocide, he situated himself outside our morality. He was surrounded by like-minded persons, all of them remote from our morality. He is accordingly disqualified as a candidate for judgement and we are disqualified as judges.

Are we disqualified from judging Hitler? Some evidence for the truth of this controversial thesis might be furnished by speakers' usual reactions to questions like 'Do you think that what Hitler did in Germany in the 1930s and 1940s was morally wrong?' or 'Do you believe that genocidal programs like Hitler's are always morally wrong?' If these questions are posed in a breezy, unassuming way, informants may be puzzled and hesitate with their answers. Informants might also judge that sentences like 'The Samurai ought not to have tested their new swords by chopping off the heads of passers-by' or 'It was unethical of the Aztecs to have cut the hearts out of their captives' have no conceivable use or are even meaningless.¹¹

These considerations nevertheless fail to show that speakers are not qualified or entitled to make adverse judgements concerning the morality of Hitler, Genghis Khan, the Aztecs, the Samurai, and other bloodthirsty individuals and groups. The hesitation of speakers under the conditions just described has various explanations besides their recognition of their own disqualification. For example, we often understand terms referring to defunct individuals and groups descriptively. Though the description theory of proper names that treats a name as designating the person who best fits a description associated with the name has been officially discredited, it accounts for what it is easy or hard for us to think. All that many people know about the Aztecs is that they were the medieval Central Americans who for a long time practised human sacrifice, and all that many people know about the Samurai is that they were a medieval Japanese warrior caste of great ferocity. Similarly, many people know little

¹¹ See Williams, *Ethics and the Limits of Philosophy*, 160–1.

about Hitler beyond the fact that he was the man who formulated and carried out a genocidal programme in the 1930s and 1940s in Germany. They may find it hard to give much content to the thought that whoever did those things ought not to have.

Yet on reflection, it is clear that the Aztecs, those very people, could have decided to give up the practice of human sacrifice by morally persuading themselves or by being persuaded by their victims to do so, and that Hitler, that very person, might never have formulated his plans or might have been dissuaded from carrying them out. So the alleged incoherence must be pragmatic, or of some other nature, rather than semantic.

Evidently, the irritating character of the question ‘Do you think what Hitler did was morally wrong?’ and the perceived aberrance of the related judgements depends on the position, expectations, and intentions of the prospective judges, not to the internal economy of the agent being judged, for a German citizen in 1939 could have condemned Hitler and his henchmen on the grounds that certain prohibitions against harm were neither being acknowledged nor respected. Now, when *N* asserts sincerely that it was wrong of Hitler to have ordered the extermination of the Jews, *N* represents himself (given the interconvertibility of declaratives and imperatives) as endorsing the moral rule ‘Don’t order the extermination of the Jews’, or some more general rule, such as ‘Don’t engage in genocidal practices’, and he implies that Hitler violated the rule in ordering the extermination of the Jews. If *N* asserts that it was wrong of the Aztecs to tear the hearts out of their captives in a religious ritual, he represents himself as endorsing a prohibitory rule forbidding such treatment of captives, or indeed of anyone, that he implies their actions violated. Standardly, moral judgements are asserted by a speaker when she perceives there to be a present need—admonitory, expressive, or behaviour-modifying—for the enunciation of the rule. Such conditions will often correspond to what Paul Grice¹² describes as the doubt-or-denial conditions for informative utterance, though the context can make the enunciation of the judgement pertinent for other reasons.

There are, however, a number of arguments that might still be advanced to support the disqualification thesis (external). Perhaps it is

¹² H. P. Grice, ‘The Causal Theory of Perception’.

arrogant to judge remote others in distant cultures, even if it is technically possible to make interpretable judgemental statements regarding them and even if one has no intention of intervening in their lives.

One such argument is the argument from entrenchment: Suppose a practice or an institution has endured in a remote culture for many years. Supporting symbolic structures, such as religious rituals and ceremonies, have grown up around it. It then earns the right, it might be argued, to be treated, if not with respect and deference, then at least with moral indifference. We need not approve it morally, but we ought not to condemn it either; indeed, we ought to adopt no moral attitude towards it whatsoever. On this view, the Roman Empire has a different moral status from Hitler's planned Thousand Year Reich. The latter was incompatible with human values and could not have been realized. However, the former institution existed for a long time. Describing the Roman Empire as a morally unjustified institution seems otiose. It just . . . was . . . though eventually needs, preferences, and tolerances changed and the Roman Empire collapsed. Slavery—whether ancient slavery or the New World slavery of the eighteenth century—might be regarded in a similar light.

Second, there is the argument from opacity: It is a condition of judgement that we understand the practices we are judging. The application of any value judgement—for example, a judgement concerning the excellence of a wine, or the ineptness of a detective story, or the beauty of a painting—presupposes familiarity with and an understanding of the kind of object that is being judged. Often this familiarity is absent when we try to judge the practices of a distant culture or the actions of persons whose psychological make-up is atypical. Perhaps we have good insight into what was going on in Germany in the 1930s, and can therefore make appropriate judgements. But, according to the argument from opacity, we cannot make confident judgements against a remote culture.

The anthropologist Evans-Pritchard claimed in this connection that while we modern Englishmen might think that a woman who crawls in the presence of her husband is expressing abjection, we don't know enough about other people's customs to say that African women who crawl in the presence of their husbands are expressing

what our women would be expressing by crawling.¹³ After all, it might be argued, if we observed dimorphic Martians adopting certain characteristic poses with respect to one another, say, lifting or lowering their antennae when they met, we would have no basis for supposing that one morphological group was showing a morally objectionable subservience to the other, and we may be in the same position in observing *S*. So we should be cautious in supposing that the behaviour of the *Ps* in *S* unambiguously indicates their status as victims; we should even be cautious about ascribing the roles of *A* and *P* to the actors in *S*. For all we know, crawling by African women is an expression of high status and *noblesse oblige*.

Third, there is the argument from pointlessness: We experience less impetus to condemn or celebrate long ago or faraway occurrences that cannot be influenced by our thoughts and expressions than to condemn or celebrate occurrences that can be. (We might term this the ‘That-was-in-another-country-besides-the-wench-is-dead’ effect.) Moral judgements are the sorts of things that can in principle make a difference, and if a judgement cannot possibly make a difference, it is pointless to propound it.

In response to the argument from entrenchment, it needs to be pointed out that while institutions are ‘tested’ by time, time does not actually approve or validate structures or organisms. The knowledge that some entity has survived over some interval does not even enable us to determine whether it is strong or the challenges to it have been weak. A bridge that would have collapsed might hold up thanks to a spell of dry weather; an alga species that would have died out might benefit from nutrient-rich run-off spewed out by a new factory; a boring author may turn up year after year on the undergraduate syllabus because curricular changes are hard-fought. And an institution can be stable without being morally unobjectionable. Its longevity may be due to the very features that make it morally objectionable, namely, that the strong have been successful in pressing an advantage over the weak. Conversely, hypermoral institutions, like experimental communities, may be unstable. We can differentiate between the overall moral goodness of an institution and its likelihood of persisting under given conditions. Though a good

¹³ E. E. Evans-Pritchard, ‘Position of Women’, 40–1.

institution that is durable is preferable to a good institution that is ephemeral, we can judge these features separately.

The opacity argument is equally unconvincing. It is true that the notion that it is a practical necessity to sacrifice captives to ensure the recurrence of the winter rain, or that it is necessary for the integrity of a warrior band to waste the villages of rivals does not come easily to us. We may not see the point of being in a warrior band. We cannot know what it was like to have been an Aztec priest or a Samurai warrior. Nevertheless, we are not as badly off as we are in trying to imagine how scrambled eggs taste to a cockroach, to borrow an example of Nagel's. We can think our way into the position of the people who believed these things, and there is no reason to be incredulous that they did. Moreover, our difficulty in imagining what it was like to have been an actor in remote situation *S* is a distraction. The professions of Aztec priest and Samurai warrior doubtless had many secret intricacies that we never will understand, but more relevant to the question of legitimate judgement is our being able to understand what it is like to be a sacrificial victim or a peasant in a burned-down village, one of the patients. There do not seem to be as many intricate mysteries here.

But perhaps there is much that we do not understand about the role of the alleged victim? In response to Evans-Pritchard's version of the argument, it should be pointed out that from the observation that a conventional gesture can be misinterpreted or overinterpreted by an inexperienced person, it does not follow that we are always prone to misunderstanding and misjudgement. Perhaps we underestimate the dignity of the peasants in the villages burned by the Samurai who might have felt themselves to be involved in nationally important or even cosmically significant events, or the sacrificial victims of the Aztecs, who may have thought an honour was being done them and looked forward to evisceration with pride. Their interpretations need not, however, affect our judgement that what the Samurai warriors and the Aztec priests did showed no moral concern for the victims, in so far as they pressed their advantage over these weaker parties to reinforce their own social positions. It is not necessary that the peasants and sacrificial victims should have accepted our analysis for it to be a correct account of what was happening. Even if we cannot interpret the behaviour of e.g. Martians or crawling women, it is

possible that, if we could interpret it, we would judge it morally unacceptable. For the disqualification thesis to be generally credible, we would have to believe that we can rarely or never correctly interpret what the *As* are doing and what is happening to the *Ps* and that our efforts to do so are futile.

Finally, in response to the pointlessness argument, it can be observed that it may do some good to condemn the dead or the remote, even if they are beyond remonstrating with or cannot take part in moral dialogue. Condemning the dead discourages new present-day followers from reviving doubts and denials of a prohibition by promulgating some immoral programme and condemning the remote may have a similar warning function. It may alert us to the presence of similar practices in unsuspected quarters. Williams appears to overstate the case in insisting that legitimate criticism requires the possibility of a real confrontation between the present-day philosopher and the original agent or his current representatives. All that is necessary is that such a confrontation can take place in the mind of the subject entertaining the proposition.

All three arguments for the disqualification thesis (external) may be turned to domestic use. A representative of the disqualification thesis (internal), Hilary Putnam, finds that some practices that are inside our morality are immune from critical judgement. Certain kinds of radical scepticism about our institutions are incoherent or, as he puts it, 'silly'.¹⁴ It must be impossible for our local culture, the implication is, to get things fundamentally wrong in matters such as the organization of work or the constitution of family life, even if they appear to be so to an observer who is psychologically alienated or reluctant to participate in its central institutions. A critic who is unmoved by the allure of what Thorstein Veblen called pecuniary society, or who, like Simone de Beauvoir, fails to appreciate the charms of the nuclear family, or who, like Freud, concludes that civilization is a source of anguish that is barely supportable without the use of Schedule I narcotics, might be considered by a Putnamian as too distant or too alienated from our institutions to be in a position to offer meaningful

¹⁴ According to Hilary Putnam, "“Is our own way of life right or wrong?” is a silly question, although it isn't silly to ask if this or that particular feature of our way of life is right or wrong, and “Is our view of the world right or wrong?” is a silly question, although it isn't silly to ask if this or that particular belief is right or wrong. As Dewey and Peirce taught us, real questions require a context and a point.' 'Objectivity and the Science-Ethics Distinction', 154-5.

criticism of the economy, or the family, or civilization. Radical criticism, on this view, always presupposes an alienated standpoint that is automatically disqualifying.

Putnam's thesis might be formulated as follows: When the following conditions are satisfied, one is disqualified from offering a condemnatory moral judgement of the situation *S* or the actions and responses of *A* and *P*, the participants in *S*, even though from some detached or alienated perspective they might appear morally defective.

Conditions for Disqualification (Internal)

- (1') *S* in our world *W*, is very different from any situation in *W'* envisioned as ideal by the prospective judges. The *As* and *Ps* in *S* are psychologically very different from the actors in *W'*.
- (2') The prospective judges do not share the values and standards of the *As* and *Ps* in *W*.
- (3') The interactions between the *As* and *Ps* are part of a way of life that is stable. They are woven into the fabric of the culture and find resonance in other practices. They organize the life of the culture as projects organize the lives of individuals.

The *prima facie* plausibility of the disqualification thesis (internal) is considerable. Its defender will point out that it is *theoretically* possible that the way we live and what we consider normal and justifiable behaviour is sadly lacking in just the way radical critics tell us. It is *theoretically* possible as well that we have no right to our goods and our houses, that our partialist concern for friends and family is morally unjustified, and that our educational and economic institutions are fundamentally unjust. It is also theoretically possible that I am a brain in a vat or dreaming all the time. We can doubt this or that, he will insist, in particular doubt-inducing circumstances, but we cannot doubt the veracity of all our perceptions, or even our entire body of optical or chemical knowledge. Analogously, we can doubt that some particular law or practice is morally justifiable, but we cannot question the entire economic or domestic basis of our society.

However, the analogy between the theoretical possibility that I am a brain in a vat and the theoretical possibility that I live in a world that is normatively quite a mess, in the ways suggested by Veblen, Beauvoir, or Freud, is poorly motivated, for the suspicion of institutions

and practices need not be hyperbolic. It may be motivated by historical, anthropological, or sociological investigation, none of which can have a bearing on whether I am a brain in a vat. The disqualification thesis (internal) states that the dominant practices of a culture are above suspicion in that culture and qualifies accordingly as ideological. It is true that it is difficult for criticism from detached or alienated perspectives to take hold, and the critic himself may have little reason to wish to influence a society from which he stands radically apart. The stability of *S* should not, however, be confused with an element of the reality constraint.

The disqualification thesis, in its external and internal forms, is therefore not a theorem of the descriptive theory of morals. It may be regarded prescriptively as an injunction, not merely against interfering with the habits and customs of others, but also against engaging in the symbolically hostile action of passing negative judgement on others. The assumption of a posture of disqualification can have genuine moral significance and we should expect that, like most prohibitory rules, a rule against judging others can be formulated at various levels of demand. The recommended forfeiture of the right to judge might be partial. It might be held that condemnation by outsiders is permitted when and only when there exists the possibility that a condemnatory movement could arise within the culture itself. Or the proponent of disqualification might permit adverse judgement if reasoned persuasion from outside is possible 'in principle'. Each of these positions represents an eligible moral restraining rule. They are in competition with and need to be defended against weaker prohibitory rules that permit judgement in a broader range of cases.

4.4. Opacity and Disqualification

Michael Walzer's discussion of the morality of exchanging women illustrates many of the features of moral discourse identified in the preceding discussion, from the employment of paraworlds in moral argument and world-level relativism, to the deployment of the disqualification thesis (external and internal), supported by considerations of entrenchment, opacity, and pointlessness. It illustrates the striking connection between immanentist hostility to theory

and a disposition to social conservatism reminiscent of Evans-Pritchard, at least with respect to certain questions.

Walzer asks us to imagine a society, remote from our own, in which 'women (all women) seem to have been socially constructed as objects of exchange and where rules of exchange follow from the construction. . . . [W]omen are transferred among households, from one patriarchal jurisdiction to another, as if they were objects of exchange.' The women, as they are portrayed, are not interested in remaining in their natal villages, or in making their own decisions where to live, or even in parity with men in these respects. Yet they have not, he stipulates, been brainwashed, physically coerced, or made desperate, and being an exchange object is said to '[bring] some benefits to at least some women (even if the benefits are much greater for men)'. He describes the exchange of women as 'only one part of a larger pattern of relationship, fitted to a system of beliefs, symbolically represented, ritually enacted and confirmed, handed down from mothers to daughters over many generations'.¹⁵

Walzer asks what we should say about that society and how we should evaluate its institutions. He tries to show that, even if the exchange practice is inconsistent with our local ideas concerning enlightened treatment of women, as long as we cannot find any resistance to it amongst the subjects themselves, we cannot confidently describe it as unjust. 'Social construction', he says, 'makes for us a complex and rich world, many features of which will seem so obvious to us that we will not be prompted to ask whether they are, of all possible features of all possible worlds, objectively best. They will have a more immediate objectivity.'¹⁶

It is important to understand Walzer's genuinely pluralistic commitments. He is not prescribing an exchange system for all women, or claiming that men have an intrinsic right to exchange women. In the terms of our analysis, Walzer is merely advancing the claim that a morally acceptable world *might* contain *some* exchanged women—not randomly distributed through the population, but congregating in particular groups. To confirm the proposition that exchanging women is morally acceptable under certain circumstances, one need

¹⁵ Michael Walzer, 'Objectivity and Social Meaning', 174.

¹⁶ *Ibid.* 173–4.

merely gain agreement that the idea of a just world composed, like our world, of non-ideal agents does not exclude the practice.

The audience to whom this prescriptive thesis is addressed can be assumed to occupy the baseline position that the women of their community ought not to be exchanged. Further, they are inclined to think it would be an injustice anywhere, that a world containing any groups of exchanged women is distinctly worse than a similar world with none. The audience might try to explain the nature of the injustice by saying, 'It is wrong to treat women as a commodity' or 'It is wrong to treat any human being as a means rather than as an end.'

The mitigating circumstances that Walzer adduces in order to challenge the higher-demand position that the imaginary community can be judged by our more abstract standards and found wanting include the following:

- (1) The system has been in place for a long time and expectations and practices have developed around it.
- (2) Though the system benefits men more, it benefits women somewhat.
- (3) The women are not coerced and do not object to the system.

Circumstance 1 is relevant because people enjoy their customs and rituals and do not like to change them. Everyone's life planning in the society under consideration depends on the assumption that women of marriageable age are going to be exchanged. Tasks such as the preparation of a trousseau for the exchangeable women, or the selection of a bride by the father of a young man might be welcome interruptions in otherwise wearying routines. Circumstance 2 is relevant, because many practices are justified by the fact that the aggregate benefit furnished is large, even if it is not shared evenly. Consideration 3 is also relevant, for we might think of the women rather as we think of monks who voluntarily enter on a life of service to others and obedience to their superiors. We may not like serving others, and we may not believe that the objects of the monks' or the women's deference are as important as they do, but they might just as well think of our secular, acquisitive mode of life as crass. If the exchanged women are helping men, the situation may be morally in order. Helping is generally morally good, for A_1 in helping sacrifices some opportunity simply in order to improve A_2 's state.

Even taken in conjunction, however, these considerations are not particularly convincing. A dense web of enjoyable cultural practices can be expected to grow up around any marriage system; there is no reason to think that the exchange system has a special richness. And since we are considering only judging and not intervening, the costs of changing a set of practices do not need to be considered. Further, though we know that the system benefits the women somewhat, we do not know what benefits are available in differing proportions to men and women living under alternative marriage systems. Implicitly, we are invited to compare the system with one in which women do not benefit at all. Finally, helping, though *prima facie* morally creditable, is not always a good thing to do. Helping, when it is not reciprocated or otherwise acknowledged, may entail a too-costly sacrifice of the non-moral values of personal dignity and autonomy. When thinking about hypothetical women and giving them certain attributes in our imaginations we can make them seem like monks—ideal monks, not real monks, who may be deluded and exploited. It is impossible, however, to regard exchanged women as individually called to a life of service. Rather, they are born into a condition of future servitude. They are described as acquiescent, but, unlike monks, they cannot opt into or out of the arrangement. Unlike monks, they are not deferential and submissive towards something grand and impersonal; they defer to other human beings whom they perhaps invest with godlike qualities.¹⁷

Walzer's position wavers between the agnostic view that we ought not to condemn the practice of exchange in a distant culture that we may not understand—a thesis about our right or our ability to judge—and the positive view that there is nothing wrong with exchange in some conceivable cultures—a thesis about the moral acceptability of some policies and practices. The agnostic argument can be conceded immediately. Where we do not understand a culture and have no way to determine what is happening in it, we should not enter a negative judgement against it. If all we could know about a culture was that women were exchanged in it, our judgements of it would always be superficial and worthless. By contrast, the judge-

¹⁷ 'The engineer, so precise when laying out his diagrams, behaves at home like a minor god: a word, and behold, and his meal is served, his shirts starched, his children quieted'. Simone de Beauvoir, *Le Deuxième Sexe*, ii. 501-2.

ment that it is exceedingly unlikely that a culture could be organized in such a way that there would be nothing wrong with its exchange system for women is anything but superficial.

The women who are objects of exchange in the example are hypothetical beings, given by the author's description in an imaginary case, not by ostension. The author has not searched through historical or anthropological sources to find a culture where women are exchanged, described its practices in all their depth and richness, and argued that the practice meets the standards of moral acceptability. The exchange of women, we can say, turning to the real world, generally involves an affront to morality. The exchanged woman may play an important role as 'cultural diplomat'. The exchanged woman travels or sends her children back and forth between her new and old villages; indeed, unlike her husband, who remains in his parents' village, she lives in multiple worlds and must adapt to multiple systems of meaning.¹⁸ Nevertheless, becoming an object of exchange exposes women in our world to the dangers of overwork and abuse, for it is the exchanged woman's labour, usefulness as a sexual resource, and her childbearing capacities that are deemed valuable. Women who live with or near their parents and siblings, rather than with or near their husbands, siblings-in-law, and parents-in-law, tend to have a higher status in their households and to have more protection from assault.¹⁹

Does the reality constraint require the theorist to build the features of overwork, abuse, and excess risk into his model? Not necessarily. It might be argued that the usual by-products of known exchange systems are not intrinsic to the practice. But, in that case, to show that an exchange system can be morally acceptable, the theorist has to provide a credible alternative model. To judge that the situation in Walzer's paraworld was morally in order, we would need to know that the men did not enjoy the advantages at the women's expense that they normally do in exchange systems. Walzer could try to build these compensating and mitigating conditions into the description of his exchanged-women paraworld and then invite us to consider it as

¹⁸ Elise Boulding, *The Underside of History*, 46–7.

¹⁹ Women who perform unsatisfactorily in their new households may be sent back to their parents who may be forced to refund their bride prices. Fear of embarrassment, punishment, and disgrace may explain women's seemingly graceful acceptance of the system. *Ibid.* 45 ff.

just. So various protections and entitlements for the women might be brought in to offset the disadvantages of having to move to a strange village away from one's family and accept a stranger for a husband. Alternatively, he could simply assert that in so far as these further conditions might obtain, there might be a situation in which women were exchanged in a morally acceptable way.

The conclusion that a society of exchanged women *might* be just and that it *might* be wrong for an outsider to condemn it is, in the final analysis, unsupported by positive argument: no description of a possible society meeting these criteria has in fact been provided. It is clear at the same time that Walzer intends to comment on practices in our world, perhaps to criticize preferential systems of marriage, or to commend women's acquiescence in systems organized principally for the benefit of men. Under this 'Putnamian' interpretation, the discussion of exotic tribal women as they might be judged by modern liberal outsiders, functions as a vehicle for a discussion of our local women aimed at defending certain elements of a status quo against attempts by radical social critics to reset the level of demandingness.

Walzer does not argue in the article under consideration that the status of females in our world *W* is beyond such 'internal' criticism. Elsewhere, he poignantly calls attention to the history of institutions 'that seem designed, above all, to break the spirit of young women'.²⁰ Noting that 'freedom in love radically alters the standing of women, but . . . doesn't . . . end their oppression',²¹ he seeks a third way—community supervision of marriage arrangements—between what he perceives as the evils of overmanaged alliances on one hand and of radical freedom on the other. Nevertheless, his discussion invites us to pose, and to answer in a particular way, the following question, which turns out not to be about exotic marriage customs at all, but about something we observe all around us. If, in our world, women do what men want them to and are acquiescent, despite their getting what looks like the worse half of the bargain, why should we argue with them?

The question reflects a genuine puzzlement that most members of the audience for the work feel. We are more than a little anxious about the subject of acquiescent women: our intuitions about this are conflicted. Had the thought experiment described a situation on Mars

²⁰ Walzer, *Spheres of Justice*, 239.

²¹ *Ibid.*

involving a practice of asymmetrical exchanges between ‘pods’ of an acquiescent class of differently coloured caterpillar-like creatures whose mode of life was otherwise opaque to us, the disqualification thesis would not have been compelling to debate. An aura of anxiety would not have surrounded it. At the same time, the women in Walzer’s paraworld are said to be so unlike our women that we cannot really understand them; they might as well be caterpillars. In the end, we do not know what to say.

In summary, Walzer’s thought experiment is provocative. It attaches to an anxiety we experience in our world *W* and is a stimulus to moral reflection. However, it does not help to establish the disqualification thesis as a defensible theorem of the descriptive theory of morals, and it is too sweeping to constitute a reasonable all-things-considered contribution to prescriptive moral theory. What a prescriptive moral theory requires is a set of criteria for determining when a morally acceptable or unacceptable situation exists; the description of an opaque situation involving inscrutable women cannot point us towards any particular answer regarding the justice of conditions obtaining either in *W* or *W*’.

4.5. In Defence of Theory

The Pyrrhonist position in epistemology rests on the confidence that nothing will go seriously wrong if we distance ourselves from a commitment to the existence of material objects distinct from our perceptions. The corresponding position in moral theory—moral Pyrrhonism—is that nothing goes seriously wrong if we do not worry about the relationship between our current practices and unobserved ideal moral system *M*. The epistemological Pyrrhonist nevertheless fails to recognize that a better knowledge of nature might be acquired were we to attempt to go behind the appearances. For what we produce ourselves and spontaneously approve may be based in illusion or simply incomplete.

Hume was confident that no one would ‘tread as willingly on another’s gouty toes, whom he has no quarrel with, as on the hard flint and pavement’.²² As an avid reader of ancient history, Hume

²² Hume, *Enquiry Concerning the Principles of Morals*, sect. 5. 39.

should have grasped the importance of the qualifier ‘given that he had no quarrel with him’. ‘Everyone’ agrees that murder is a terrible crime and all moral codes proscribe it. But they define *it*—the kind of killing that is a terrible crime—differently. The readiness with which sophisticated humans will kill and despoil their close neighbours in warfare for trivial or confabulated reasons should dispel any sunny assumptions about the existence of a uniform and well-internalized moral code we have wordlessly fallen into. ‘Don’t steal’ is ambiguous, and its applications contestable. What about landlords, taxes, and derivatives-traders? North Americans might all be willing to sign their names under the overgeneralized formula ‘All persons should be regarded as having an equal claim to life, liberty, and the pursuit of happiness’, or to go to war under this slogan, but their day-to-day political conduct will belie this alleged commitment. In short, it is difficult to substantiate the claim that there is a pretheoretical core of moral regulations to which all normal humans subscribe, in the sense that they acknowledge their rightness and strive to conform to them, even if they are from time to time overcome by situational pressures and fail to do so.²³

The resources of Hume’s moral theory in what might be termed the realm of microethics are not negligible. The virtues are moderating influences on the exercise of advantage—which sometimes, though not always, amounts to viciousness—in the one who possesses them. The agent exercising the virtue of fidelity stands with a cause even when his interests are no longer served by it, and the sincere agent conveys the actual state of her mind even when it exposes her to certain losses. A gentle demeanour signals that one will not resort to violence or intimidation to attain desired ends. With some virtues, such as temperance and sobriety, there is no intrinsic advantage-reducing feature, though the intemperate often do place substantial caretaking burdens on others. Adherence to virtuous behaviour in Hume’s sense precludes cruelty, betrayal, secret manipulation, and other forms of mistreatment of others to which agents are occasionally tempted.

²³ Michael Ruse is confident that ‘No one would say that it is morally acceptable for grown men to have sexual intercourse with little girls.’ Ruse, *Taking Darwin Seriously*, 212. Of course this is an exaggeration. Ruse means, I think, to advance the view that there is a well-confirmed norm here, regardless of what thousands of people say and do.

The virtues are nevertheless elements of an archaic theory of upright conduct that bundles the observation of moral rules together with other requisites for being an admirable person. The resources of virtue theory for addressing problems of advantage and disadvantage in what might be termed the realm of macroethics are accordingly limited. Hume envisions a society whose members are in constant interaction with one another, confronting one another face to face. The moral community Hume appears to be describing consists of members of a single social class, even if those members differ to some extent in their possession of wealth, wit, beauty, and other social advantages. The virtuous persons of the *Treatise* and the *Enquiry* do not appear in threadbare clothes, they do not have brown or black skins and labour on plantations, nor do they exercise the virtues from or in hospitals and prisons. Further, the virtuous may be presumed to have an interest in avoiding interaction with the pitiful and unfortunate, in so far as such persons are known to arouse the most distasteful sentiments.²⁴

We can easily imagine a world of individually virtuous persons, each of whom reliably exhibits each Humean virtue and displays no vice with respect to each person whom he encounters. The world is nevertheless *prima facie* unjust, since goods are distributed unequally between groups that do not encounter each other—perhaps deliberately. Since not everyone is rowing in the same boat, non-theoretical agreements that fulfil everyone's preferences do not emerge. Game-theory simulations suggest that fair distributions emerge when agents with similar quantities of resources encounter one another repeatedly and strike bargains. It is reasonable to predict that global fairness will not be maintained when agents are antecedently segregated into haves and have-nots and do not encounter one another at random.

The deficiencies of virtue theory can be addressed. A virtue theorist could add 'global benevolence' or 'concern with equality of the sexes' to the list of essential virtues. This move will, however, seem arbitrary, for there is little reason to think that people with these virtues are more admired by their peers and have better reputations than their deficient counterparts, features essential to Hume's account

²⁴ Hume, *Enquiry Concerning the Principles of Morals*, sect. 6. 33.

of why the virtues are good. On the contrary, an avid interest in distributive justice and equality of the sexes can make a person distinctly unpopular amongst his peers. Alternatively, she could follow Hume himself in supplementing the theory of personal virtue with a theory of social justice; in morally ideal worlds, she will say, the inhabitants are both personally virtuous and socially just.

Whether its solution is to be found in a strengthening of Hume's notion of private benevolence or supplementation of virtue theory with a theory of justice, the existence of the problem is undeniable: Virtue theory contains a bias towards moral neglect simply on account of its relatively narrow focus. The world slices, one might say, of the discourse of virtue are small ones, and revisionary aspirations are not at home in a Humean framework.

Why did modern moral theory after Hume evolve an impersonal and universalistic component? This is obviously a complex historical question. A satisfactory answer to it would mention the breakdown of the image of a hierarchically structured cosmos, the sense of global connection despite cultural difference furthered by maritime trade in the late eighteenth century, the beginnings of scientific anthropology and ethnology, and the increasingly visibility of slave labour and of the emerging underclass in a rapidly industrializing Britain. The immanentist is apt to discount the fact that the famous modern moral theories grew up from inside human life, from the observations and experiences of their founders. They did not come from the Martians or from a star.

Just as natural science can be considered both as an extension of everyday human knowledge based on the operation of our normal sensory faculties and inference mechanisms, and as a corrective to ordinary perception utilizing artificial instruments and schema for perceiving and analysing data, theoretical morality both extends and corrects.²⁵ We can employ both scientific knowledge and folk wisdom in making practical decisions and we can employ theoretical perspectives as well as community conventions in making moral decisions. The only-because formulation of the non-theoretical reasons principle suggests that there is a sharp distinction between an internal, situational reason for doing something in a particular case

²⁵ As Hume recognized in claiming that it was meaningless to dispute whether justice was 'natural' or not. *Enquiry*, app. 111, 173.

and a non-situational, theoretical reason. This conceptual boundary is blurred in experience. Our factual knowledge about the world has a powerful bearing on how we see and react to particular events. So does acquaintance with fiction. Both distant and imaginary places and happenings furnish information or informational content from outside our own lives requiring assimilation, but the knowledge gained is inside us. In acquiring knowledge, our representations of the world and our responses to it are altered by what we are able to absorb. The mere memorization of a theoretical formula or the inscription of one into a personal notebook does not automatically alter anyone's behaviour; I can possess the formula without its making any difference. The sources of the alterations that constitute learning nevertheless extend beyond the personal experiences derived from getting about in one's familiar territory.

Accordingly, no a priori limits can be set to what perspectives one is able to internalize or act upon as a result of engagement with formal moral discourse. Individuals and cultures have been able to give up practices and attitudes of whose rightness they were at one time firmly convinced and upon which a whole range of social and personal expectations depended. It might be argued that moral progress occurred only when individuals made some outcome their concern, or their project, and pursued their goal with the same single-mindedness with which Gauguin pursued his desire to become a skilled and original painter. No empiricist should deny, however, that we are influenced by theoretical discourse, and that our passions and projects would not have the form they do in its absence.

The aristocratic militarism of the Samurai and the priestcraft of the Aztecs were anchored by expectations and obligations inherited from the past and passed down through families. Yet the conduct of these persons vis-à-vis peasants and captives was morally unacceptable. And the 'context and point' demanded by Putnam as a condition of asking about the validity of the ways of life that incorporate those practices can be easily supplied. The subcultures concerned possessed three important characteristics. First, they were founded in social dominance—the dominance of warlords vis-à-vis the populace, priests vis-à-vis captives. Second, they represent recurrent patterns in history; warlords and priests have often assumed privileges of life and death over peasants and captives. Third, they are candidates for moral

criticism, since they are manifestations of advantage-taking behaviour. We can judge the way of life of the people involved we are unable to judge the way of life of Martians, whose odd squeaks and gestures really are unintelligible. The beauty of Japanese armour and the complexity of Aztec art tell us that these practices were embedded into surrounding aesthetic and religious practices of great depth and interest. Nevertheless, these features do not render the agents or patients involved exotically inscrutable, or immune from criticism, or show that they are redeemed by the cultural manifold in which they are situated.

As Strawson, and before him Kant, pointed out, I can never see myself from inside merely as a product of the external play of social forces. The narrative I give about how I arrived here and why I do this or that reflects only the dimmest awareness of the way in which my culture sets the parameters for aspects of my agency and my experience. My ability to present to others a personal narrative of deliberative agency, which explains how I came to occupy my role, is determined by my being a member of a species that narrates itself to itself and to others. The cultural and individual narratives and explanations we offer tend to occupy epistemological space so completely that more objective accounts are seen as alien or superfluous. Embedded persons have both advantages and disadvantages when it comes to understanding why things are as they are. They can explain things that seem illogical or unfair to the outsider, but it is another question whether their explanations are, objectively speaking, any good.²⁶

To summarize, the embeddedness of a theorist—who is at the same time a moral agent—in a particular culture with its own history and traditions can limit her ability to understand and assess other cultures. She should be modest in advancing formulas of obligation pertaining to persons very unlike the ones she knows. At the same time, she is not a wholly disqualified observer of others' practices, especially if she takes the trouble to learn about them, and she should resist the ideological deployment of the disqualification thesis. Chapter 5 will show that moral theorizing is constrained by an anonymity requirement. The anonymity requirement provides the precise condition

²⁶ As Dwight Furrow comments, 'A kind of moral blindness seems to be built into the structure of narrative histories because the sort of moral obligations they generate are tied to partial, contingent perspectives', *Against Theory*, 65.

that distinguishes a viable moral theory—whether formulated as a virtue theory, or in deontological or consequentialist terms—from an ideology, the latter understood narrowly as the promulgation of a doctrine by a privileged group that is, in the classic formulation of Karl Mannheim, ‘unable to see certain facts which would undermine their sense of [rightful] domination’ and that consciously or unconsciously ‘obscures the real condition of society both to itself and to others and thereby stabilizes it’.²⁷

As was suggested earlier, confirmation is a matter of degree. To the extent that a rule or policy is confirmed—and this can only be done through controlled discursive practices—one ought to act in accord with it and is blameworthy for not doing so. But isn’t the claim *One ought to act in accord with well-confirmed moral norms* itself a theoretical statement of prescriptive morality? Why must I conform to it?

The answer to this provocative question is that *if* you regard conformity to well-confirmed moral norms as optional, you thereby hold a certain prescriptive thesis. You believe that, in the best world, people do not by and large act in accord with well-confirmed moral norms. As such, your thesis can be assessed for its theoretical cogency and criticized as falling short of ideal levels of demandingness. Your thesis is not irrational, but it is insufficiently moral, for the claim that conformity to well-confirmed norms would be optional clearly favours the better-off. They have reason to exercise the option rather than conforming to the norm and sacrificing an advantage they could otherwise retain. The conservative bias in antitheoretical moral discourse is not accidental.

²⁷ Karl Mannheim, *Ideology and Utopia*, 36.

The Anonymity Requirement and Counterweight Principles

Virtue theories like Aristotle's and Hume's do not privilege *me* over the other persons with whom I normally interact. The same obligations of justice, fidelity, honour, allegiance, benevolence, humanity, clemency, temperance, and sobriety are required of each of us to the same degree. They do, however, appear to award a privilege to *us*—to groups. The privilege takes the form of an implicit permission to ignore much social injustice that does not concern property relations, narrowly defined. A world of virtuous Aristotelian or Humean agents can contain slaves or exploited labourers, women without rights and privileges, and other deprived and neglected persons; Hume's personal understanding of the world was that it not only could, but did contain such persons, and that their existence did not diminish the quantity of virtue. This is not to say that Kant's personal understanding was very different. His applied moral theory awarded distinct social privileges to members of particular groups: to men v. women, employers v. labourers, and invaders v. indigenous persons. A revisionary radical in his pure moral theory, Kant retreated to a more comfortable status-based theory when he came to discuss law and politics.

If we care about the distinction between theory and ideology we must observe a strong constraint on moral theorizing, the anonymity requirement. Together with certain counterweight principles that diminish the force of the argument from heavy costs, the resulting package contains what is needed to defend egalitarian proposals against the challenges to them mounted on the basis of the first-person standpoint, costs to agents, and the weight of custom and convention. Though a consistent and thoroughgoing antitheorist will

reject the package—for if moral theory is useless and uninteresting, constraints on moral theorizing are even more so—he forfeits at the same time his claim to be advancing prescriptive claims that are more than authored norms.

5.1. The Anonymity Requirement

The perspective of an agent on the world is a situated one. It is not only personal or subjective but also marked by the agent's identity as a member of this or that group with this or that history.

The significance of group membership for moral theory resides in the fact that policies influencing our levels of well-being are normally formulated with reference to groups. Tax schemes specify how income earners at various levels are treated. Draft regulations separate women from men. University admissions policies divide high-achievers from low-. Individuals may or may not be subject to psychological chauvinism—dislike of or contempt for members of other races, ethnic groups, nationalities, sexes, income-tiers, intelligence-levels, etc., or the desire that one's own group be perceived as superior in some domain. Nevertheless, in so far as we are concerned with our own welfare, it matters to us how other groups fare. In some cases, one can expect to be made personally better off by policies favouring another group. Putting swimming pools and other recreational facilities in the inner city rather than in the suburbs can make life safer and more enjoyable for suburb dwellers who are less likely to be mugged when they come to the city. In other cases, group-favouring policies such as affirmative action can diminish an individual's chances for success.

Purely selfish agents would choose their social policies according to how much they expected, personally, to benefit from the policies. They would support politicians who promised tax benefits to persons at their income level, whether high or low, assuming they did not think their own security would be thereby threatened. They would favour affirmative action if they were members of under-represented groups and otherwise oppose it. They would fight against public transit systems if they owned cars and vote for them otherwise. If they were residents of Central African nations, they would favour

transfers from wealthy nations to poor ones, but not if they were residents of wealthy nations, unless they expected a greater indirect benefit to flow back to them.

Though most of us are somewhat selfish in our favouring of social policies, we are not entirely so. We are able to some extent to evaluate the impersonal goodness of social arrangements without giving special weight to our own interests. What we normally do to a greater or lesser extent—evaluate policies impersonally—a moral theorist is obliged to do consistently. A theorist acquires credibility by adopting an anonymity requirement as a constraint on theory construction and evaluation. The requirement blocks the formulation of selfish policies and those that inappropriately favour certain groups.

This claim that moral theorizing is constrained by such a requirement calls for further explanation and defence. Consider two conditions, Condition 1, in which persons *A*, *B*, *C*, *D*, and *E* each have \$20 and Condition 2, in which person *A* has \$80 and persons *B*, *C*, *D*, and *E* each have \$0. According to classical utilitarianism, Condition 1 is preferable to Condition 2. When we take the further step of assigning individuals to identities or ‘roles’, the assignment of me to the identity of person *A*, according to utilitarians, cannot influence the evaluation of the two conditions. If Condition 1 is judged superior on the grounds that it contains more total or average utility, it must be so judged wherever I conceive myself as ending up. But what is the force of ‘cannot’? I may clearly prefer a state of the world in which I have \$80 and four other people have \$0 to a state of the world in which each of us has \$20. Therefore—it might be argued—the utilitarian assumption that assignments are irrelevant to the choice of systems must be wrong.

Classical utilitarianism is deficient considered either as a descriptive theory about how agents make moral decisions or as a prescriptive theory about how they ought to. However, its deficiencies do not reside in its refusal to consider post-assignment preferences in theoretical contexts. In projecting an ideal paraworld, whether it is with the intention of changing the background beliefs of the audience, or opposing revisionary proposals, the theorist recognizes that the persons in that world can all be assumed to weigh their own pains and pleasures more heavily than those of other people and that some of the parties have an interest in the maintenance or advancement of the

status of their group. A theorist frames rules for various reference groups; people of a given sex, age group, temperament, income level, nationality, occupation, drive level, political persuasion, and so on. She does not evaluate the rules according to how well she or her reference class does in the worlds that realize them.

As a reputable theorist, I cannot write a set of rules that are endorsable only if I am myself in the envisioned paraworld. More precisely:

The Anonymity Requirement

When acting as a moral theorist, N cannot deem one rule preferable to another because of her identity. If W_1 , exemplifying one rule, is deemed better than W_2 , exemplifying some alternative, this judgement must be preserved over any possible assignment of N to various roles in W_1 and W_2 . Further, the judgement of the audience A charged with confirming the theory as promulgated must be robust over their envisioned assignment to various roles in W_1 and W_2 .

It follows from the anonymity requirement that I cannot prefer a set of rules R to a set of rules R' because R contains rules that work to my advantage, or to the advantage of my family, my class, or my country. I am prevented from passing off justifications of my existing privileges as soundly reasoned theory and from awarding myself special exemptions, even on the grounds that they tend to the general good, unless I would be able to appeal to the same grounds from any other subject position. Gyges, if we consider him under the rubric of a moral theorist, violates the requirement. He endorses a rule that reads, 'Only a certain person P may steal, deceive, and murder undetected'. He prefers that rule to one reading 'No one may steal, deceive, and murder undetected', but only if P is Gyges.

At the same time, the anonymity requirement does not entail the repugnant conclusions associated with the claim that everyone's interests must be counted equally by everyone. It does not demand that agents weigh costs and benefits to all agents equally, or that any particular value be maximized, regardless of the costs to someone. Worlds in which agents weigh everyone's interests equally do not necessarily satisfy the requirement and their impartial weighing is certainly not sufficient for its satisfaction. For a world *in which* agents

regard themselves consistently from a third-person standpoint—that is to say, considering their own interests as no more important than the interests of some arbitrarily designated person—will clearly be structured quite differently from a world in which agents are subject to some other equally impersonal constraint. It is clear that a paraworld in which agents do, or could come to, regard themselves consistently from a third-person standpoint, does not satisfy the reality characteristic. It is a further question whether a world in which agents were able to weigh costs and benefits perfectly impartially and did so would be preferable to a world in which they were able to but did not.

As a theorist, I can therefore judge that it is sometimes, though not always, impersonally better that agents weigh their own interests more heavily than they weigh the general welfare. This can be seen from a consideration of various versions of the ‘trolley problem’. Should I run over one person to prevent my trolley running over five? Should I murder my spouse if it will save twenty? Should I lie to my spouse to prevent five people lying to their spouses? Simple-minded utilitarians may answer yes to all these questions, but it would be difficult to confirm the claim that such things should be done. It would be too much to ask of me that I murder my spouse to save twenty orphans, if a maniac were to offer me the opportunity to do so. It may also be too much to ask of anyone that he shoot down a plane with 300 passengers to prevent it crashing into the White House.

On these questions, intuitions differ so profoundly that, unless theory-level relativism is true, the norms that are enunciated are only authored norms, with no claim to objectivity and so general bindingness. The anonymity requirement does not settle the question what the right thing to do in such dilemmas is. However, it implies that in establishing what the right thing to do is, my judgement must remain consistent whether I suppose myself to be the potential killer, the spouse, an orphan, a passenger, a White House employee, or anyone else who has an interest in the outcome. All things considered, I can prefer a world in which (in some circumstances) I am killed to a world in which some agent (in those circumstances) murders someone else, thereby saving me along with nineteen others. The role of having sacrificed my life to spare another agent from having to do a terrible thing—murder a spouse in order to save nineteen others and me—is not an unacceptable one.

The anonymity requirement is consistent with world-level relativism. It permits a theorist to write different rules for different subclasses on the grounds of varying circumstances (including different preferences) or avoidance of imposition. The obligation to act benevolently, for example, might well be regarded as different for the well-off and the badly-off. Artists might be allowed certain exemptions and permissions that others are not. However, if *N* writes a rule that exempts artists from certain responsibilities, her endorsement of it as morally preferable to the envisioned alternatives cannot depend on her assumption that she is an artist. If *N* judges that a world that might contain some exchanged women is at least as good as an otherwise similar world that contains none, his judgement must be robust over his assignment to the role of woman or man, exchanged or not exchanged.

As noted, the promulgation of a permission rule of the Schefflerian variety that states that ‘Everyone may assign greater weight to his own interests than to the interests of others’ need not violate the anonymity requirement. Yet, certain subformulas of the rule may constitute a violation under specific empirical conditions. Consider the formula of obligation ‘Everyone is permitted to devote all his resources to his family and friends and to give none to distant strangers’. One might defend this *ACT* on the grounds that a world in which both the wealthy and the impoverished devote all their resources to kin and kith is all-things-considered better than other paraworlds in which the wealthy do not do this. Suppose, however, that the poor have few or no resources to circulate amongst their family and friends. In that case, the effect of everyone’s conforming to the permission rule will be indistinguishable from the effect of everyone’s conforming to a nominally different rule that reads simply, ‘Wealthy persons may devote all their resources to friends and family’, or from one that reads ‘The wealthier a person is, the greater the resources he is permitted to devote to friends and family’.

The situation is similar with respect to a rule *R* that reads as follows, ‘Anyone who has a very great talent that can profitably be exercised is permitted to cease to provide for his children as long as someone else can be found to sustain them’. If the world is divided into two classes, *K* and $-K$ such that the $-K$ s lack talent, the $-K$ s’ talents are unprofitable or impossible to exercise, or the $-K$ s are unable to

find anyone else to sustain their children, the effect of realizing *R* will then be indistinguishable from the effect of realizing the following rule *R'*: 'Ks who have a very great talent that can profitably be exercised may cease to provide for their children as long as someone else can be found to sustain them'. If the Ks are more likely to possess talents that can profitably be exercised and are empirically more likely to be able to find someone else to sustain their children, the effect of the rule will be indistinguishable from a vaguely stated rule *R''* that states that 'More Ks than -Ks are permitted to cease to provide for their children on the grounds of a talent that can be profitably exercised, etc.' To fulfil the anonymity requirement, these equivalent rules should be judged acceptable regardless of the possible roles that could be occupied by the proponent and evaluators of the rule.

The anonymity requirement does not require me to be psychologically indifferent to my potential assignment to various roles in alternative worlds. I might prefer the role of an artist to all other roles in many possible worlds because I like to draw. In that case, I have a non-moral hedonic preference for worlds in which, other conditions remaining the same, I am an artist. Numerous roles in other worlds can be hedonically unacceptable to me without being morally unacceptable to me and so just morally unacceptable. I might prefer non-existence to solitary existence in a world in which I was the last living survivor of a collision between the earth and an asteroid. In such a world, there would no longer be any moral relations between animate creatures, hence, few if any opportunities to engage in morally unacceptable practices. My role might not be morally unacceptable but unacceptable nevertheless on hedonic grounds as being too lonely. By contrast, being the last survivor of a nuclear holocaust might be considered a morally unacceptable role, since my being in that state could have been prevented by human exertion and benevolence. We must distinguish, however, between the judgement that my role in the actual world is morally unacceptable, which is to be understood as the thought that a world with no one like me in it is morally preferable to all worlds with anyone like me in them, from the judgement that I ought to eliminate myself from the actual world. The costs of transition from one state of the actual world to another—in this case, the loss of a human life—may be too high to justify the transition. In general, proposals for changing existing conditions or

redistributing goods have to be evaluated for their morality separately from questions of the goodness of hypothetical conditions or protocols for the distribution of goods. Immanentists are properly sensitive to these costs; they need merely to be dissuaded from overestimating them.

There are three main objections to the anonymity requirement, that it is too strong to serve as a basic constraint on moral theorizing, that it is too weak, and that it is pointless. I take these up in turn.

(a) *Objection: The anonymity requirement is too strong*

Why is the requirement good for a theory to satisfy? a sceptic might wonder. Is the anonymity requirement so well confirmed that it could be regarded as an authorless norm with respect to rational rule-projection? Evidently, it is not a condition of rational rule-projection, for Gyges, for one, fails to observe it and, *pace* Nagel, it is not clear that Gyges is irrational as opposed to merely immoral in refusing to renounce the advantages that go along with being invisible. Doesn't the anonymity requirement, taken as a prescriptive meta-rule, beg the question against the critics of exigent moralities whose claim it is that agents need not observe the constraints of the third-person perspective in deciding on courses of conduct? The requirement is, the critic might slyly suggest, another authored norm, issued by theorists who have been too heavily influenced by the famous modern moral theories. It is on the same logical footing as the household rule that computers under the direction of subadults have to be turned off by 10 p.m.: just another bossy imposition. Agents have access to information about themselves, it might be argued, that is strongly relevant to what they ought to do: information about their own needs, desires, abilities, and levels of tolerance. No matter how well informed a moral theorist is, she does not know what it is like to be me or a member of one of my reference groups. *Her* estimation of what it is like to occupy *my* role is not worth much, theoretically.

In response, it can be conceded that each of us has only vague ideas about what it is like to be some other person or to belong to some other reference class. Nevertheless, arguments in moral theory presuppose that we have enough access to otherness to make

judgements. When Williams invites the audience to agree to the specific exemption from marital duties he requests for ‘Gauguin’, he presupposes that we can understand ‘Gauguin’s’ situation—perhaps from knowing something about Gauguin—well enough to venture an opinion. Credible oligarchs who favour exemptions for the rich must present themselves as knowing what it is like to be down and out.

The ‘Gauguin’ of Williams’s fable is not a moral theorist. He is not conceived as endorsing a rule that states that husbands should remain at home, except when they are artistically gifted, or that artistically gifted husbands should remain at home, except when ‘Gauguin’ is one of the reference class. Nor does he endorse a rule that states that persons, with the exception of himself, should submit their proposed courses of action to a Kantian or utilitarian test. His relationship to moral rules in the story is purely negative. He thinks over his problems and his conflicts, but he is portrayed as one of the many persons who adopt a plan of conduct with moral implications without consulting a theory. The disqualification thesis (external) denied, any observer is free to criticize him, approve of him, or refuse to pass judgement on him. Williams, however, in narrating the story of ‘Gauguin’, is in a different position. He is a theorist, citing the case of a-man-whose-relationship-to-moral-rules-is-purely-negative. He can only be understood as arguing prescriptively—and impersonally—for an exemption from dutiful behaviour for persons with large talents, capacious desires, and relative immunity from retaliation. Williams’s claim that there is no legitimate vantage point from which ‘Gauguin’ might properly be reproached is defended by appealing to the reader for his endorsement of the author’s judgement about what ought to happen or need not happen in a particular case.¹

Any agent knows more about himself than any theorist can know about him. All this means, however, is that agents are in a good position to provide information that can disconfirm theories about what people like themselves ought to do.

¹ Williams might be understood as rejecting the Anonymity Requirement. Effectively, he claims that it is acceptable for me to judge the situation in which I kill one person to save twenty to be worse than the situation in which X kills one person to save twenty. As a theorist, however, I have to decide whether a world in which it might happen that I killed one person to save twenty is preferable to worlds in which this could not happen.

(b) *Objection: The anonymity requirement is too weak*

Suppose the world is divided into two classes, the Reds and the Greens. Suppose a Red theorist propounds a set of rules according to which the Reds may live from the labour of the Greens, but not vice versa. Her judgement is robust over her assignment to the role of a Red or the role of a Green; if I *were* a Green, she thinks, though I most certainly am not and can scarcely imagine being one, it would not be a bad thing for the Reds to live from my labour. So the requirement does not exclude exploitative systems or even systems of slave holding.

This suspicion is correct. The theorist has done nothing to violate the canons of proper theory construction. The anonymity requirement contains a bias against exploitative and other unjust worlds since, by and large, theorists and evaluators do not rate highly worlds in which they are exploited or exploit others, but it does not automatically deselect exploitative or unjust worlds. An application of the heavy costs principle could, in theory, override the moral reluctance to be an exploiter. Exploitation can only be excluded by a strong moral preference for non-exploitative worlds, that is, by a substantive commitment, not the use of a formal criterion.

The anonymity requirement is evidently not satisfied by Aristotle's views on natural slavery, for Aristotle would not have considered a state of the world in which the proponent of those views, Aristotle, could have been a natural slave to be a possible state of the world, and therefore could not have considered his position as a natural slave acceptable. Technically, Aristotle does not have a 'theory' of natural slavery—he merely advances some authored and targeted norms in the context of a descriptive theory of different types of human nature. However, one can imagine a theory of natural slavery resembling Aristotle's that is tested by the requirement. Would a competent audience necessarily reject it? According to Rawls, we can show that slavery is unjust by appealing to 'the fact that it allows some persons to own others as their property and thus to control and own the product of their labour'.² This seems a rather elliptical proof of the unacceptability of slave-holding worlds; it provides little clue as

² Rawls, *Political Liberalism*, 122.

to why, all over the world, for much of history, slavery has been considered by reasonable people as a reasonable way to organize labour. How did we get to disconfirm this popular view? For I take it we have disconfirmed it.

The awfulness of an institution need not correspond to the brevity of a proof required to show its wrongness. A more leisured, but ultimately more satisfactory, approach to the question ‘Could slavery be just?’ requires us to take the theoretical proposal that in some morally acceptable worlds there might be some slaves seriously. This granted, there are three ways in which a theory of slavery might fail a confirmation protocol. First, the claim that some persons are born with the dispositions and preferences of slaves and do not mind being slaves might turn out to violate the reality constraint. Or, even if the theory passed that test, it might fail with respect to the anonymity requirement, in case the competent audience was unwilling to concede that, if they had been born with the dispositions and preferences of slaves, or if they did not mind being one, the role of slave would be acceptable to them. Finally, even if the theory were to pass both tests, it could fail to incorporate the idealism characteristic. It is conceivable that there are natural slaves, that the competent audience would judge that the relevant roles are acceptable, yet still prefer a world in which no slaves were held by anyone to a world in which there were some slaves for the moral reason that slavery involves egregious advantage-taking. But is not slavery simply a violation of human rights? Can’t this point be used to exclude a theory as *a priori* ineligible even before it is put to the test of the reality constraint and all higher tests?

Rights, however, are best viewed as shorthand notations for theory-writers. The right-not-to-be-a-slave eliminates potential theories from further consideration. A right is not a feature of living human beings that prevents us doing certain things to them; for the fact is, human rights can be and often are violated. The ascription of a right nevertheless tells us not to bother debating further the acceptability of permissions, e.g., to enslave our fellows. When we hypostasize rights, we tend to see them as features of the reality constraint, as a kind of aura protecting people from abuse. They are, however, more properly seen as determined by the idealism characteristic.

The tendency to hypostasize rights as possessions of people disguises the fact that one can behave unacceptably to people without

violating what are usually considered their rights. It is not only seriously wrong to chain and flog persons of low motivation to get them to work, it is also somewhat wrong to pursue and harass persons of low motivation to get them to work. The rights theorist cannot account for the latter wrongness in a manner consistent with the former, for being harassed by officials in a welfare office or being the subject of viciously punitive editorials cannot be considered a violation of human rights. Yet if it is thoroughly wrong to control and own their labour by owning persons, then it is somewhat wrong to exercise certain kinds of control over labour by exercising powers that resemble the powers of ownership. The violation of a right should therefore be seen as a kind of limiting case of moral awfulness. One cannot do worse to people than violate their rights.

(c) *Objection: The anonymity requirement is pointless*

Suppose I am a well-off person who judges that all members of my reference class, persons with an income of $> \$100,000$ per annum, ought to reorganize their consumption patterns and send $\$3,666.66$ annually, for reasons to be set out below, to UNICEF. I represent us as all doing so in an ideal world. I am now faced with the prospect of acting on the norm I have judged appropriate, i.e., of having to send $\$3,666.66$ to UNICEF when others in my reference class are not going to do so. *They* are going to spend the money on parties and vacations. While it would be easy for me to send in $\$3,666.66$ if they were going to do so as well, it will now be difficult for me and my obligation is at least *prima facie* reduced. I should therefore simply have asked how much *I* ought to send to UNICEF, giving no thought to others in my determination. Therefore, the anonymity requirement is pointless. It is an element of the reality constraint, after all, that I have only limited influence on others, and that getting out of step with kith and kin, or merely with my normal group of associates, imposes indirect costs, for example, the costs of being unable to participate in certain social activities or reciprocate the generosity of others, or being thought holier-than-thou.

The appropriate question for me to ask, however, is 'What ought well-off people who will be acting alone in their social groups give to UNICEF?' The amount may well be less than what well-off people

who will be acting in concert with others ought to give, but this should not be too surprising. Obligations can be weakened by context, though they can also be strengthened. On the Ericksonian principles discussed earlier, my fidelity obligations, if I live in a promiscuous society, are somewhat weaker than they are if I live in a chaste society, for a given amount of moral effort can prevent more harm in one case than in the other. If it is the norm in my subculture to give money to UNICEF, rather than taking vacations, I am somewhat more culpable if I take vacations instead, since less moral effort is required in one case than the other to give money to UNICEF. Of course, neither charitable obligations nor fidelity obligations are precise determinables and, to the extent that they are not determinable, I am morally free to exceed or fall short of expectations on either score.

The anonymity requirement functions in a way that is related to, but not equivalent to, Rawls's stipulation that the formulation of a theory of justice is undertaken behind a veil. It disposes of some serious problems with the veil notion. Rawls proposed that an acceptable theory of justice is the theory that would be agreed to as a result of bargaining amongst a set of individuals, not one of whom 'knows his place in society, his class position or social status . . . his fortune in the distribution of natural assets and abilities, his intelligence, strength, and the like. . . . [T]he parties do not know their conceptions of the good or even their special psychological propensities. The principles are chosen behind a veil of ignorance.'³

The reason the original position must abstract from and not be affected by the contingencies of the social world is that the conditions for a fair agreement . . . between free and equal persons must eliminate the bargaining advantages that inevitably arise within the background institutions of any society from cumulative social, historical, and natural tendencies.⁴

This explanation of how the veil assures a fair outcome is puzzling, because it is unclear how there could be any bargaining amongst individuals who were essentially identical, in their degree of ignorance and their absence of personal and cultural tastes and preferences. The notion of a contract or a bargain struck amongst some group of persons implies a process of give and take, concessions and com-

³ Rawls, *Theory of Justice* (rev. edn.), 11.

⁴ Rawls, *Political Liberalism*, 23.

promises, and the merely formal multiplicity of the discussants provides no basis for such a process. Rawls might as well have said that an acceptable theory of justice is the one that would be preferred by a single individual who did not know his place in society, his intelligence, strength, preferences, and so forth, on the assumption that it would be impossible that two individuals fitting this description should have a rational preference for different theories.⁵

However, the single-chooser formulation does not appear to capture Rawls's intended conception either. The notion that some form of 'bargaining' occurred amongst the contractors in the original position was meant to reflect the idea that different interests were at stake and that those in different roles were in a position to provide information to others concerning their needs and preferences. In short, Rawls's formulation in the passage just quoted is contradictory.

It might seem that to avoid the paradox, we could represent the deliberators as differently motivated, and perhaps even as differently empowered, but as equally reasonable. However, if the deliberators are conceived as equally sympathetic to every possible set of preferences, and as equally well informed about one another and the world, their multiplicity is still only formal. They still constitute in effect a single chooser. Contractualist theories confuse two separate issues, the pragmatic need for agents to tailor their desires, demands, and expectations to one another, to agree on what is going to happen in their world, and the epistemological need for normative theorists to coordinate their information and to agree on what agents are like and what they ought to do. For when Scanlon says that 'when we address our minds to a question of right and wrong, what we are trying to decide is, first and foremost, whether certain principles are ones that no one, if suitably motivated could reasonably reject',⁶ he leaves it unclear whether we decide on principles by engaging in a discussion of our individual desires 'on the ground' with our direct competitors for goods, honours, and other resources, or whether we decide on principles by discussing people in general's desires for goods, honours, and other resources with our direct competitors in moral theory. Scanlon refers to 'unity with our fellow creatures' as the desideratum specially

⁵ As Ted Honderich pointed out in a memorable review essay, 'The Use of the Basic Proposition of a Theory of Justice', 81.

⁶ Thomas Scanlon, *What We Owe to Each Other*, 189.

or even uniquely available in contractualism, but the unity of purpose in a well-run institution or the satisfaction citizens from all walks of life might feel in reaching a compromise on taxation policies that everyone can live with is not the same as the intellectual unity with our fellow thinkers that we seek as investigators of the moral sphere.

The claim

- (1) A_1 is morally obliged to do ACT in c with regard to A_2

is, on our analysis, equivalent to the claim

- (2) A_1 always does (advantage-reducing) ACT in c with regard to A_2 in preferred worlds satisfying the reality constraint.

To confirm 2 is to have confirmed 1; to try to confirm 1 is to try to confirm 2. Now, contractualists appear to take 1 in turn as implying

- (3) If A_1 were to propose ACT in c to A_2 , this would be reasonable; and A_2 's agreement to ACT in c would be reasonable; and if A_2 were to propose ACT in c to A_1 , this would be reasonable; and A_1 's agreement to ACT in c would be reasonable.

The 'agreement' might concern anything from the division of labour within a household, to the manner in which a county will deal with resident drug addicts, to the policies of a nation with respect to agricultural subsidies that influence the standard of living elsewhere.

Now, 3 does not of course entail either 2 or 1. It is reasonable for me to invite my daughter to accompany me to the movies on a Sunday afternoon, and her agreement would be reasonable, but going together to the movies is not a feature of overall morally preferred worlds or morally obligatory. Is 3 nevertheless a necessary condition of 1 and 2? The answer to this question depends on how 3 is interpreted. If it is equivalent to 4, the answer is clearly no:

- (4) If A_1 were to propose ACT in c to A_2 in our world, A_2 's agreement would be forthcoming; and if A_2 were to propose ACT in c to A_1 in our world, A_1 's agreement would be forthcoming.

The contractualist will not take the expectation of actual agreement between existing persons or parties as a necessary condition of obligation, for he is fully aware that not all parties in our world are reasonable and morally motivated. Claim 3 must be understood as referring to idealized agents who are reasonable, very well informed, and moderately benevolent. If A_1 and A_2 are now conceived as just

such impartial theorists who agree on what is to happen to them (as it happens), now that they have shared relevant background information, contractualism simply replicates the theory of confirmation offered earlier. If, by contrast, their agreement is significant because they both get the policies they want, contractualism is first-order moral theory. It carries the implication that

- (5) Worlds in which reasonable, well-informed, morally motivated people obtain the conditions and policies they want are morally preferable to those in which other kinds of people obtain the conditions and policies they want.

The contractualist idea can be filled in as follows: In the best worlds, agents do not merely happen to act well; they act well as a result of engaging in certain discussions and deliberations that terminate in agreement all around. *No reasonable, well-informed, etc. agent in a contractualist world is ever forced to go along with a moral policy he does not like*, and we might suppose as a corollary that, in contractualist worlds containing the usual proportions of unreasonable and semi-reasonable agents, *the more reasonable a given individual is, the less he finds himself forced to engage in moral performances involving the relinquishment of advantage that he does not wish to relinquish or requesting the relinquishment of advantages that others refuse to give him*.

Though 5 above does not follow logically from the definition of an obligation, its potential for independent confirmation is high; it is difficult to see how one might prefer a world composed of persons like us in which less reasonable individuals were subject to less forcing than more reasonable individuals, and in which reasonable individuals rarely got what they asked for from other reasonable individuals. The contractualist is nevertheless required to substantiate better the intuition that frustrating or disappointing unreasonable people is morally better than frustrating or disappointing reasonable ones. He may be able to firm up this intuition. In any case, it is important to distinguish between an interpretation of contractualism according to which it corresponds to a plausible meta-ethical theory of confirmation and a very different interpretation according to which it corresponds to a mildly arbitrary moral theory.

The response of the contractualist might be that our claim to have explicated the concept of moral obligation is just as unsatisfactory. For

if the reality constraint precludes us representing persons as fully reasonable, how can we convincingly represent such unreasonable beings as conforming to certain moral formulas of obligation that satisfy the idealism characteristic? The response to this probing and appropriate query is simply that *if* a formula of obligation has the true status of an unauthored universal norm, *then* it is ideally acknowledged by everyone, regardless of their degree of reasonableness and their level of moral motivation, and that this interesting feature of obligation (which does not require the existence of any obligations) cannot be expressed in contractualist terms. The topography of contractualism, one might say, is too flat.

5.2. The Partiality Exemption

Consider the claim that a moral system that includes the rule that we ought to transfer some non-trivial portion of our assets to needy strangers is too difficult for us to endorse because natural selection has made us so that we are disposed to care mainly for our kin. We can care a little for our kith too, it might be conceded, but nature has not bestowed on us a disposition to universal benevolence. In so far as a theory represents the bestowal of goods on strangers as obligatory, it imposes an intolerable burden.

This claim can be interpreted narrowly, as pertaining to a prima facie charity obligation to give specific sums to Third World stranger-assisting organizations like UNICEF or Médecins sans Frontières. A wider interpretation is available as well. Nature, though perhaps not natural selection, has made us in such a way that we are disposed to advance the interests of certain groups with which we identify. Our spontaneous inclination to relinquish advantages to members of other groups is weak and our obligation to do so is correspondingly weak.

To evaluate this claim, consider a case in which partiality seems justified. The relations between the Leibniz Society and the Kierkegaard Society are not very close. Each group has certain interests—in fostering the study of Leibniz or Kierkegaard, in expanding its membership or restricting it, in making the work of its members respected and visible, in furthering the researches of younger scholars, and so on. The two groups are indifferent to one another,

though they do not have contempt for one another or think of the other's projects as less than worthwhile. It does not occur to members of the Leibniz Society that they ought to help members of the Kierkegaard Society by, for example, stuffing envelopes for them, or by vacating their meeting room early so that the Kierkegaard Society can have a longer session. Why should they? Group membership normally entails certain group interests and a reluctance to sacrifice advantages.

The anonymity requirement does not directly preclude a policy whereby groups are permitted to ignore or reject the claims and requests of other groups whenever they feel no natural sympathy for them. A policy permitting neglect considerably reduces the complexity of individuals' lives. If the Kierkegaard Society asked us as members of the Leibniz Society to vacate our room a half-hour earlier, we could refuse simply on the grounds that we are not interested in furthering their aims and projects. We are not obliged to debate their request at length or to give reasons for our refusal. Is there a corresponding permission to ignore distant others? We could help people in Central Africa, but do we have to?

A permission to ignore of this type would free individuals from the need to consider the justice and reasonableness of multiple claims for recognition and assistance by Third World strangers and persons with disabling medical conditions. We Northerners could just decide to ignore Central Africans on the grounds that, while there is nothing wrong with them and nothing unworthy about their projects, we are not especially interested in their getting on with their lives. We are not obliged, we might maintain, to debate their requests for help at length or to give reasons for our refusal to help.

There are nevertheless moral reasons not to write too many policies that permit groups to ignore and reject the claims and requests of other groups on the grounds of natural sympathy and partiality. While anyone can benefit on occasion from ignoring another person's grievances, ignoring and rejecting are the prerogatives of the advantaged party in any situation and the policy favours them. The Leibniz Society and the Kierkegaard Society confront each other as equals and for relatively small stakes. A policy of neglect and indifference will be acceptable to me no matter which group I imagine myself assigned to. This is not true of Northerners vis-à-vis the Central Africans.

To what extent are we justified in ignoring complaints and grievances we judge to be bothersome and how far must we expend time and effort in hearing them out despite their bothersome nature? Rarely, if ever, is the quantity of advantage we should relinquish to another group whose personnel may seem strange to us and whose interests are not our interests a practical determinable. There is a gradient of possible answers to these questions. A possible world in which powerful groups consistently ignore the claims of less powerful groups in whom they are not interested in order to save time and stress is not, however, minimally morally acceptable and is even inaccessible to most of us.

The descriptive theory of morals tells us that people in different societies, and different people within a single society, can be expected to have various accounts of their obligations to close relatives and to distant ones, to acquaintances and strangers. The prescriptive moralist is faced with the task of selecting amongst the multiplicity of possible formulas of obligation that concern the needs of persons of various degrees of relatedness and closeness. The claim that partiality is justified can be interpreted as the claim that, in an ideal world otherwise resembling ours, the needs and wants of kin and kith are routinely satisfied before the needs of strangers.⁷ If the claim is sustainable, such a world is preferable to an otherwise similar world in which people usually or always satisfy the needs of strangers before turning to their family and friends, as well as one in which they satisfy the needs of strangers before addressing the wants of kith and kin. What reason is there to believe that the first distribution protocol is ideal?

Let there be two worlds, each divided into two equally endowed populations of haves and have-nots. In W_1 , the resources of the haves circulate entirely within families, who purchase for their children goods and experiences such as computers, bicycles, and ballet lessons. In W_2 , some resources are transferred to the have-nots, alleviating their most urgent needs. It is hard to appreciate that W_1 contains more human flourishing. From the perspective of a situated 'have', there is more human flourishing apparent in the shining eyes and

⁷ For a spirited defence of partiality, see John Cottingham's 'Partiality, Favoritism and Morality'. Cottingham argues that, except under special circumstances in which there is a duty of impartiality, 'It is morally correct to favour one's own'. *Ibid.* 358.

happy cries of his child on getting the \$700 bicycle instead of the \$200 bicycle and in his not having to endure any disappointed looks and reproaches, than in his getting a ‘Dear Contributor’ letter from UNICEF thanking him for sending \$500. Yet his experience does little to confirm the theoretical judgement that worlds in which resources are kept circulating locally are better. Human flourishing appears to require some expropriation of resources from well-off families. The question is not whether, but ‘how much?’

It is burdensome to meet the needs of strangers when kith and kin are vociferously representing their wants. The utilitarian thesis that one should adopt the policy that maximizes aggregate well-being impersonally calculated is exigent. An agent could continue to move deprived persons from a condition of neutrality—the state in which existence is not distinctly preferable to non-existence—up to a positive condition of well-being by dispensing his assets bit by bit until he was on the point of falling below neutrality himself. Such a policy is too costly. The agent’s desire to consume, to learn, and to express herself will be frustrated if she adopts it. Not only will she have to give up a few luxuries, such as shade-grown organic coffee, imported ham, and Perry Ellis sportswear, she will have to cancel magazine subscriptions and stop buying books and seeing movies. Normal relationships with kith and kin will be distorted by the refusal to maintain an accepted level of expenditure on them. As Mary Douglas has pointed out, material goods ‘are needed for making visible and stable the categories of culture’.⁸ They support what she terms ‘rituals of consumption’, including dining out, entertaining, decorating, and dressing, that offer opportunities for the exchange of information and the cementation of personal and familial ties.⁹ To make matters worse, the hypermoral agent will be forced into relationships with distant strangers with whom she has nothing in common beyond membership in a common species with the same basic needs. To determine where her newly freed-up resources can do the most good, she will have to give a lot of thought to matters to which she doesn’t normally give much thought, such as sub-Saharan politics. These considerations seem to militate against the adoption of exigent rules.

⁸ Mary Douglas and Baron Isherwood, *The World of Goods*, 59.

⁹ *Ibid.* 81.

The question how far we are obliged to depart from the norms of the culture and from settled consumption habits has no objective answer. 'Around here', we outfit our children with good bicycles, and they feel aggrieved if we do not. Nevertheless, local norms still leave a great deal of scope for self-positioning. What kind of bicycle do I buy? How often do I replace children's bicycles? New or used? How much of a concession should be made to the child's desire for reputation and status? The role of the prescriptive moralist is to advance a position on wealth retention and wealth distribution that is a moral position—i.e., that requires some renunciation of advantage—and to make the paraworld in which the corresponding obligations are instantiated appear to be realizable. In doing so, she must take into account the costs levied on agents by the obligation and the multipliers and reducers that influence the estimation of their weightiness.

Though the question how much income a given agent ought to divert to strangers is not a practical determinable, 'Fermi methods'¹⁰ of estimation can nevertheless prove useful. Even when one's access to information is limited, one can make headway with a few rough and ready figures. According to Partha Dasgupta, 'The financial requirements for a broadly-based human resource development strategy designed to meet basic needs would total approximately 5.5% of GNP.'¹¹ Suppose that global GNP is thirty trillion US dollars. Then approximately 1.8 trillion dollars is needed to meet basic needs universally. There are various ways to calculate what ought to be sent each year.

I can assume, for example, that . . .

- (1) Each person in a world of six billion will contribute equally to the subsistence of the rest. Someone like me should send in \$300.
- (2) Only the better-off half will contribute to meeting the needs of the worst-off half, and the better-off half do not need any more resources than they already possess and can do with less. Someone like me should send in \$600.

¹⁰ The physicist Enrico Fermi showed how it is possible to make useful determinations of such seeming unknowables as the number of pianos in Chicago. The figures arrived at are inaccurate, but not by orders of magnitude.

¹¹ Partha Dasgupta, 'National Performance Gaps', quoted by David Braybrooke, 'The Concept of Needs', 62.

- (3) The top 5 per cent of world income earners should meet all the needs of the worst-off. Because three hundred million persons will be providing the total, someone like me should send in \$6,000.

I could send in the excess over what I require to meet my own basic needs, approximately \$75,000, and if I am a saintly person I will do so. However, it would be hard to confirm the theoretical claim that I am impersonally obliged to do this. Conversely, it is difficult to see why options 1 and 2 should fall victim to the argument from heavy costs since their impact on my life is trifling. Both are consistent with Liam Murphy's plausible suggestion that I cannot be impersonally obliged to impose greater deprivations on myself than those that would be imposed on all of us, were all of us (in the reference group deemed most relevant) to do what we ought to.¹² A world in which some people do impose extra deprivations on themselves is nevertheless morally preferable to a world in which no one does, and of course any sum that I fix on is morally creditable, though, other things being equal, the larger sums are morally more creditable than the smaller, lying higher on the exigency gradient.

5.3. From Theory to Practice

The following general prescriptive principles seem to be tenable in light of the argument from heavy costs. The first limits the scope of the idealism characteristic of a theory. The second informs us that even agents in ideal paraworlds do not perform with perfect regularity.

- (1) A moral theorist should recommend some set of advantage-reducing rules that benefit others at some cost to agents, but these rules should not require actions beyond their ability to perform or whose performance would usually seriously overtax them.
- (2) A moral agent who has acknowledged a set of advantage-reducing rules as *prima facie* obligations need not follow

¹² Murphy's 'collective principle of beneficence'; see Liam Murphy, *Moral Demands in Nonideal Theory*, 84 ff.

them to the foot of the letter in case unforeseen or special circumstances would render compliance in a particular case extraordinarily difficult.

To deny 1, one would have to be persuaded that agents can be obliged to perform actions that they cannot in fact perform, and that they can be required, not just seriously to overtax themselves on some occasions, but also to adopt a set of principles that have the usual and customary effect of seriously overtaxing them. To deny 2 one would have to be persuaded that some obligations are so rigidly fixed that no conceivable circumstance could justify or excuse non-performance.

Some theorists have held that there are duties that absolutely must be performed. Kant appeals in his *Lectures on Ethics* to the ancient notion of a fate worse than death. 'If . . . a woman cannot preserve her life any longer except by surrendering her person to the will of another, she is bound to give up her life rather than dishonour humanity in her own person.' Strangely, he does not suggest that, if I discover myself to be a sex maniac, I ought to kill myself to avoid dishonouring humanity by raping people. The notion that, if one cannot fulfil some obligation, even a non-moral purity obligation, one ought to choose death, is problematic in any case. It presupposes that the morally-worse-than-death roles can be identified and opens the door to the inference that I can be obliged to kill someone else to prevent that person occupying a role that is morally worse than death.

It might be suggested that 2 is false because the obligations to refrain from genocide, torture, and enslavement of others are absolute in a way that the obligation to refrain from non-performance of contract or lying is not. This substantive view is compatible with the principle. We can reject as ineligible for further consideration all paraworlds with any genocide, torture, or enslavement, no matter how well off on average their inhabitants are. To insist that some obligations are absolute is to express confidence that there are no unforeseen or special circumstances that could justify an exemption for an agent on an occasion.

Accepting 1 and 2 as general prescriptive principles then does not determine their application. Thesis 1 leaves it open what the criteria of serious overtaxation might be and leaves a theorist free to insist on a demanding set of obligations on the grounds that they do not

exceed human capacities. Thesis 2 leaves open how we are to identify special or unforeseen circumstances and how hard an agent should try to fulfil an obligation under adverse circumstances. How frequently can agents fail to perform a given action without its being the case that a certain moral rule is not embedded in their paraworld? A high-demand theorist will recognize few or no conditions as exempting agents from compliance with a given formula; a low-demand theorist will be more forgiving.

These results suggest that the reported philosophical tension between objectivity and selfishness—construed either as self-interest or as group interest—is a conflation of several different problems:

- (1) A theorist can have trouble deciding in a given context what kinds of sacrifices of advantage one ought to be required to make for moral reasons. And, an agent can have trouble simply deciding what to do when an important principle is at stake.
- (2) Varying circumstances, or previously unnoticed elements of the reality constraint, such as a disposition to partiality, once pointed out, can create difficulties for a rule a theorist has attempted to formulate. Difficulties can also surface when an agent is faced with unforeseen circumstances or acquires some new information.

For example, *N* acting as a theorist might have trouble deciding whether people can be lied to or killed when there exist strong personal or group, e.g. ‘national’, interests in their being deceived or dead. This difficulty affects her ability to envision a morally acceptable paraworld. *N* acting as an agent might have difficulty deciding whether to tell a lie or to sign up with the Selective Service Administration.

Or suppose *N* acting as theorist has adopted the pacifist principle that in a minimally morally acceptable world composed of people like us, no human ever kills any other human for any reason. Yet she has somehow failed to anticipate the question: ‘If *X*’s grandmother is being raped by *Y* and *X* has a gun, is *X* permitted to kill *Y*?’ and is thrown into confusion. Pacifist *N* acting as agent might find herself in the position of having a gun when her grandmother is being raped and be faced with the choice of sticking to or shrinking from her assumed obligation to refrain from killing under all circumstances.

A tension between subjective and objective perspectives can be said to exist when *N* as a theorist is trying to determine whether a principle whose imposition would provide large benefits, impersonally calculated, is too costly for agents. What might be called the Williams-Scheffler dilemma—should agents save one child of their own or many strangers whenever they are faced with the opportunity to do one but not both?—illustrates that tension. The same tension can be said to exist when *N* as an agent is trying to determine whether to perform an *ACT* such as letting two strangers drown and saving her own child. The tension between subjective and objective perspectives is wrongly construed as a conflict between situated living agents on one hand and detached theorists on the other.

The emotions of anguish, regret, and remorse are associated with the existential tensions of agents trying to decide what to do. These emotions do not normally affect theorists, who experience purely intellectual difficulties in reconciling values, even if there is room for a kind of intellectual worry. Yet the theorist's worry may be prompted by thoughts about what it would be like to be an agent in the circumstances she is considering, and tragic existential dilemmas involving a conflict between a principle judged to be right and a feeling are important even if infrequent in the lives of individuals. I judge that it is right to kill one to save twenty . . . but I cannot pull the trigger myself. I believe in euthanasia in hopeless cases . . . but I cannot endure my own sick child's being given a lethal injection. I am committed to marital fidelity and the principle of veracity . . . but I'm madly in love with someone else. I think people should give a reasonable amount to assuage global hunger . . . but I adore fine footwear and smart accessories. As a theorist, and as an agent, I have three choices:

- (a) As a theorist, I can insist that the principle is unimpeachable and that the obligation requires the suppression of certain preferences; as an agent, I can suppress them.
- (b) As a theorist, I can reformulate the rule in light of my new knowledge of the world, knowledge that at least one person (and probably more than one) has emotional dispositions, inhibitions, and preferences whose existence or strength

were not known to me before; as an agent, I can decide to act in accord with a less exigent rule.

- (c) As a theorist, I can ignore the problem, and as an agent I can act in accord with my feelings or out of a blind sense of duty without making any effort to reformulate the rule.

Williams suggests that there is nothing in virtue of which a choice of (c) versus (a) or (b) can be shown to be impossible for a rational person engaged in prescriptive theorizing, or to involve the violation of objective norms. This claim is correct in several respects. The authorless norms of rationality do not make policy (c) irrational and most moral principles we accept have not been confirmed and are not objectively binding. Yet in another respect, the claim that option (c) is always open to us is misleading. My choice of (c) exposes me to first-order moral criticism if ignoring a moral rule involves me in advantage-taking as an agent. It can leave me open to second-order criticism as a theorist if refusing to revise or endorse the rule can be construed as an advantage-maintaining evasion. For while my recognition that one ought to do *ACT* in *c* is not always sufficient on every occasion to motivate me to do *ACT* in *c*, the following propositions are fundamental elements of the theory of morals.

My recognition that performing *ACT* in *c* is obligatory can decisively motivate me to perform it.

I am morally at fault if it has been well confirmed that performing *ACT* in *c* is obligatory and I do not.

My moral standing, as well as my motivation, can be affected or determined by what is extramentally the case.

5.4. Counterweights to the Argument from Heavy Costs

Within the constraints imposed by the anonymity requirement, how do prescriptive moral theorists try to secure agreement on the preferability of their paraworlds and thereby confirm their prescriptions, elevating them as far as possible to the status of authorless norms? Chapter 2 suggested that they tend to advance simple rules for reasons of elegance and economy and Chapter 3 suggested that overly exigent

rules are avoided since it is more difficult to convince judges of their obligatoriness. The propensity to advance simple and elegant rules naturally conflicts with the desire to advance fully credible ones, applicable in any situation. Simplicity and elegance may carry the day: Kant's elementary rules—his stipulation of the duties to veracity, benevolence, self-development, and the preservation of one's own life—are greatly admired partly on account of the generality with which they are stated, though the first two duties at least are exigent and admit of exceptions or weakened interpretations.

However, it is not the case that less exigent rules are routinely preferred to rules that are more exigent. Prescriptive theorists can depend on certain intuitions that tend to reduce the force of the argument from heavy costs wherever it furnishes a *prima facie* reason for weakening an obligation. The costs of telling the truth or refusing to give way to suicidal despair may be heavy; nevertheless, we expect agents to undergo some degree of suffering to uphold the duties of veracity and perseverance. Informants typically regard some classes of difficulty of fulfilling a proposed formula of obligation as not significantly weakening its demand to be complied with.

Consider, for example, a typical 'duty of presence': the *prima facie* obligation to visit a sick relative in the hospital or to attend a dinner party to which one has agreed to come. The benefit actually conferred or the pain alleviated or spared might be minimal in either case, and this might be easily foreseen. The sick relative might even be in a coma or sound asleep and my contribution to the evening might be small. Getting to the scene might be expensive and time-consuming. Yet it is not considered right to let offset calculations determine policy where many duties of presence are concerned. Most of us have the motivating feeling that we are supposed to visit sick relatives and that we should go to parties once we accept the invitation. Existentialists might shake their heads over this admission of automatism with respect to certain classes of duty, but an observer who is simply reporting on moral phenomena has to acknowledge the existence of the feeling of being obligated to be somewhere and its role in governing our actions.

Agents, then, may be expected to and expect themselves to undergo some measure of trouble and to expend effort to fulfil even

burdensome and/or low-benefit obligations. This expectation holds outside the moral sphere—for not all duties of presence are moral duties—as well as within it. Personal grooming rituals, for example, may be exigent. Costly products may be purchased and applied in a time-consuming way, or trips to the hairdresser may be required at close intervals for little visible benefit. Yet the rituals may be significant elements of a person's everyday activity plan. They can increase his self-perceived aesthetic stature just as the fulfilment of certain presence duties increases self-perceived moral stature, and others, too, may be impressed by the performance of demanding routines that, as Freud speculated, signal reserves of psychic energy as well as of time and money.

The belief that there are duties that must be performed no matter how difficult the conditions, might be thought barely rational. People who take out a second mortgage to fly to a wedding are judged impulsive. Yet whether or not the circumstances ever warrant such actions, duties of presence and related conceptions of right conduct are related to natural principles of justice that are ubiquitous and that are not considered irrational or impulsive. Internalized, certain beliefs about appropriate distributions of advantage—the ill-gotten gains principle, the contract principle, and the culpable ignorance principle—reduce the weight of concerns about difficulty and burdensomeness in the minds of typical human beings. When cited by a prescriptive moralist, their effect, other things being equal, is to move demand levels higher, even when the simultaneous tendency of the argument from heavy costs is to move demand levels lower. That we have these normative beliefs, as well as a susceptibility to the multiplication of our burdens, can be considered an element of the reality constraint. Accordingly, the counterweight principle is as much an element of the descriptive theory of morals as the heavy costs principle:

Counterweight Principle

Subjects typically regard some classes of difficulty of fulfilling a proposed formula of obligation *ACT* for circumstances *c* as not significantly weakening its authority.

There are three principal types of situation in which a diminished sense of obligation and consequential non-performance brought on

by heavy costs tends to be regarded as moral failure or moral deficiency on the part of the agent who experiences it. In such cases, the agent is expected to assume a heavy burden for moral reasons and the costs of performing the action are not regarded as weakening the obligation.

- (1) *Ill-Gotten Gains Principle*: If the costs to the agent, though substantial, are perceived as resulting from the loss of advantages that were unfairly obtained in the first place, his obligation will be regarded as less reduced than it might have been were the advantages obtained fairly.
- (2) *Contract Principle*: If the agent is believed to have committed himself previously to a joint project with others who were counting on his cooperation, the costs to him of fulfilling the obligation, though substantial, will not be weighted as heavily as they would be had he not entered into the agreement.
- (3) *Culpable Ignorance Principle*: If the costs to the agent, though substantial, are perceived as deriving from correctable misperceptions or false representations, whether or not these are unique to him, his obligation will be regarded as less reduced than it would be if the costs were based in veridical perceptions and true beliefs

According to 1, notions of fairness, or correctly distributed advantage, interact with agents' welfare-centred considerations. According to 2, notions of prior mutual agreement do so as well. Principle 3 suggests that an objective representation of the world interacts with these considerations. The presence of such counterweights is not decisive since the logic of rules is non-monotonic: The renunciation of an unfairly won privilege or continued participation in a joint project might be so excruciatingly painful to an agent that it is judged to weaken what would otherwise be a strong obligation on her part.

Because they are proposed as principles of moral psychology, not as prescriptions about how we ought to think and judge, the above principles can be tested by intuitive methods, thought experiments of the familiar type. Readers who judge these cases differently are sources of important data for the theory of morals, as are readers who confirm the judgements given.

(1) *Ill-Gotten Gains Principle*

A connoisseur knowingly acquires a set of ancient statuettes from a smuggler who has stolen them from a museum in a foreign country. She delights in the statuettes and devotes her life to studying them, publishing a number of articles on their aesthetic qualities and their cultural significance. The theft is discovered some years later, her ownership is revealed, and the foreign government demands that she return them before her work is finished. The blow is heavy. It would be some consolation and perhaps allow her to continue to work if she could keep just one or two of the statuettes, sending the foreign government a few fakes. Is her obligation to return all the original statuettes overruled by the hardships and frustration of her life project that she will experience?

Most informants judge that it is not, on the grounds that the project itself should not have been undertaken with stolen statuettes, even unknowingly stolen statuettes. When an advantage or a resource should not have been possessed in the first place, most people believe that it ought to be surrendered as soon as this becomes known even if it is troublesome and distressing for the agent surrendering it.

(2) *Contract Principle*

X co-signs a loan for Y, who is about to take up a well-paying job and who needs to raise a certain sum for a down payment. Immediately thereafter Y suffers an unfortunate road accident and lies in a coma for six months waking up immobilized but lucid and able to talk. Meanwhile, the loan comes due. Y cannot pay and X defaults, forfeiting the asset. Y is upset. X insists that he was trying to be helpful in getting Y's loan approved and never anticipated having to pay. Y points out that X should have considered this eventuality before signing and had some back-up plan in place. X retorts that Y should have considered the possibility that he would go into a coma and had some back-up plan in place. Who is right?

Most informants side with Y. If one signs a contract, makes a promise, or knowingly volunteers for a dangerous or risky assignment, they think, one incurs an obligation that is not reduced by the eventual

costs to the agent, even should matters go other than precisely as anticipated. Some informants are inclined to read the ‘Gauguin’ case as analogous to this one. They see ‘Gauguin’ as involved in a joint project—family life—in which others are counting on his cooperation. People, they insist, have a duty to be aware of the possible costs to themselves before impulsively entering into joint projects, whether they do so out of generosity and goodwill or in the expectation of a reciprocal benefit. Even if they are not aware, they are still responsible for holding to the original terms.

(3) *Culpable Ignorance Principle*

X is a well-off person who pays high income taxes. The administration of her country is corrupt. A large proportion of her taxes goes into over-billed public works projects, re-election campaigns, and bad loans to students attending phoney educational institutions. *Y* is a well-off person who pays high taxes and believes falsely that he lives under a regime like *X*’s. Both *X* and *Y* feel equally burdened by the restrictions on their consumption by the taxation system obtaining in their separate countries. Who has done worse in maintaining an illegal offshore account that does not require him to divulge his taxpayer identification number?

Most informants judge that *Y*’s obligation to cooperate with the tax system is higher than *X*’s on the grounds that false beliefs that are corrigible in principle do not provide excusing conditions and that *Y*’s belief is merely self-serving. Yet acquiring a correct representation can be very burdensome to agents. A demoralized taxpayer could be in a situation in which no good information about the actual operations of his economic and political system is available, or in which it is extremely difficult to get this information. The obligation to be properly informed may be judged to be weakened by the difficulty of becoming informed. As the present argument is not normative—I am interested for the moment in what judgements members of moral communities spontaneously make—I am prepared to concede all these observations. When it is burdensome to get information, the obligation to get it tends to be perceived as reduced, and, when it is impossible to get a piece of information, subjectively felt costs based on false beliefs may be judged to nullify obligations. Nevertheless,

burdens generated by ignorance when better information is available are not ordinarily judged to reduce the strength of obligations.

The principal conclusion suggested by these thought experiments is that an observer's estimate of the degree of obligation a moral agent stands under varies with the information given. Information about the costs to the agent induces a fall in the estimate of degree of obligation; other information cancels out the fall that would otherwise be perceived. Costs to agents make no determinate contribution to our judgements about what ought to be done. It is sometimes reasonable and relevant to cite them in prescriptive argument and it is sometimes reasonable to insist that they are not relevant.

The thought experiments above were not intended to show, and certainly do not reveal, that specialist endeavours are not worthwhile, that contracts between consenting parties are unbreakable, and that wealth preservation is an ignoble goal. Yet many informants have little sympathy with the inhabitants of these briefly exposed world slices, though these creatures are manifestly trying to get on with their own projects, expressing their own choices and values. Imaginary cases may be described in the moral literature in such a way that the urgency of an attachment or a goal is foregrounded against some pallid restraining rules. Here I have simply experimented with the construction of three situations, employing the reverse technique of making subjective burdens appear somewhat flimsy against a more solid-looking background of obligation.

High-demand prescriptive moralists confronted with the argument from heavy costs are accordingly not helpless. One concerned with the problem of aid to distant strangers should face squarely the fact that much of his audience will perceive it as quite burdensome to send even \$100 to UNICEF and will judge, of themselves and of others, that their obligation to do so is low. The moralist need not deny that subjects' perceptions of their burdens are morally relevant; they are ordinarily judged to be so, and there is no higher court of appeal that can declare them irrelevant. At the same time, he can supply evidence that the good achieved is substantial and the burdens less than imagined. While conceding that the benefit provided by a charitable person is perhaps small in relation to aggregate need, he can show that it will be large from the perspective of the beneficiary. He can provide information to the effect that donations reach their

destination, that the conditions to which they are addressed are remediable, and that the specific amounts asked for are reasonable and helpful. He can point out that a year in which \$100 was sent to some organization that helps to get people on their feet is likely to be indistinguishable, in terms of the donor's satisfaction of his needs and desires, from one in which no money at all was sent. Indeed a year in which \$1,000 was sent might not be markedly different in retrospect from a year in which no money was sent.

Further, the prescriptive moralist might focus on reducing the effect of burden multipliers. He can propose strategies for keeping feelings of relative deprivation in check by giving a social context to charitable donations.¹³ He can propose new technologies of generosity such as small automatic withdrawals to mitigate agents' feelings of sudden dispossession and lost opportunities. He can try to supply information about the world economy that demonstrates that the experienced burdens are attributable to ill-gotten gains. Alternatively, he can strive to bring to light some implicit contract, against which the current feelings of resisting agents have no purchase. This advice to the prescriptive moralist should not be understood as cynical or manipulative. Rational argument in prescriptive moral philosophy cannot be understood otherwise than as the advancing of reasons and considerations that influence the setting of demand levels and the adopting of particular formulas of obligation.

As Peter Unger observes, most people throw away solicitations from UNICEF and similar organizations, without thinking they have done anything very wrong. Unger argued that the offence is grave, comparing it to driving past a bleeding stranger on the highway.¹⁴ The reaction of many members of their audiences to the challenges of Singer and Unger was that while their utilitarian arguments against selfish consumption were admittedly forceful and their concerns exemplary, they were not very persuasive. In ignoring what people are known to care about—the protection of their children and themselves from the more brutal aspects of existence, the cultivation of their talents and interests, and the beautification of themselves and

¹³ Beneficence doesn't have to be all or naught: reduce oneself to penury or do nothing. Compare Nagel on the difficulty of the 'leap' to another kind of life. *View from Nowhere*, 206.

¹⁴ Peter Unger, *Living High and Letting Die: Our Illusion of Innocence*, 9. See also Singer, 'Famine, Affluence, and Morality'.

their houses—they violated, it might be said, the reality constraint. That observation is correct. Nevertheless, Singer and Unger are advocates of massive advantage-reduction on part of the well-off and their critics are not. The critics react to the presentation of a paraworld in which well-off persons perform radical acts of renunciation with distaste. They do not find this paraworld appealing or think those agents are much like them. The critics cannot be proved wrong, but they can be asked to defend their claim that a *prima facie* obligation to provide a large benefit to those who are very badly off is defeated by the prospect of burdens that may not be very heavy, and that the sceptic suspects are the product of ignorance and ill-gotten gains.

Consider, finally, the claim that moral considerations do not have automatic priority over other considerations. This appears to contradict the claim that moral considerations are overriding. However, both claims are true. The claim that moral considerations are overriding can be taken to mean that we humans attach a surprising or impressive, and at times truly astonishing, degree of importance to reducing our advantages in favour of others and that this is a good feature of our world. The claim that moral considerations do not have automatic priority does not contradict this claim, for it can be taken to mean that we humans sometimes disregard moral considerations or refuse to take them into account and that this is also a good feature of our world. A promise to help someone move furniture can be broken in favour of a once-in-a-lifetime chance for box seats at the opera; established habits of service and sacrifice can be abandoned in order to have an emotionally meaningful life. Worlds in which everyone cares more about morality than anything else are unfit for our habitation.

What Williams refers to as 'the morality system' is experienced from time to time as burdensome. As he points out, its exigency instils in most of the audience for moral theory a diminished sense of responsibility. This diminished sense is, however, offset by our suspicion that the burdens of morality would appear to be less if we recognized the extent to which our advantages are the consequence of ill-gotten gains, or the violation of implicit contracts, or based in culpable ignorance. It is, however, effortful and psychologically costly to seek out and digest information about our privileges and

exemptions. Hence our ambivalence. Any honest person must recognize the existence of conflicts in him- or herself between impartial ideals of justice and the love of ease and luxury, between the desire for purity and integrity and the desire for experience and multiplicity. The world as I found it—as I stumbled into it—and its particular arrangements are not my fault. Nagel observes that the lack of a washing machine by the family next door is not his responsibility, even if he could have bought them one.¹⁵ And indeed, although anyone may intone a list of duties to us in the most solemn of accents, it is still up to us to reply that we do not care, or that we care about something else more. In a striking passage, Williams remarks that ‘We are not primarily janitors of any system of values, even our own: very often we just act, as a possible confused result of the situation in which we are engaged. That, I suspect, is very often an exceedingly good thing.’¹⁶

The claim that we are not the janitors of any system of values can be taken as a denial of the *prima facie* obligation to do whatever will maximize any particular value. Consider the following scenarios:

If *X* kills *Y*, it will prevent five other persons from killing five others.

If *X* tells a vicious lie to *Y*, it will prevent five other persons viciously lying to someone.

If *X* makes a slave of *Y*, it will prevent five other enslavements.

Unreconstructed consequentialists maintain that *X* is required to kill, lie, or enslave. It is better to do whatever results in the death of one than the death of five. Unreconstructed deontologists maintain that these actions are forbidden, since the prohibition on these actions cannot be lifted for the sake of any outcome. Williams’s remark suggests that these actions are neither obligatory nor forbidden. Someone faced with the opportunity to prevent a total of four deaths may perform an action unthinkable under other circumstances. Or not. We moral theorists cannot dictate to *X*.

Again, however, the meta-ethical posture of disqualification disguises ordinary world-level moral theorizing. If it is neither obligatory nor forbidden for *X* either to perform or not to perform the above actions, it is permissible to do *ACT*. This is interpreted in

¹⁵ Nagel, *Equality and Partiality*, 84.

¹⁶ Williams, ‘Critique of Utilitarianism’, 118.

our scheme to mean that, in morally good worlds, some agents perform *ACT* in *c* and some do not. Williams's remark suggests that this could be somewhat 'random'. Rather than arguing, as a world-level relativist might, that e.g., the police may kill one to save five but ordinary citizens may not, Williams indicates that agents could just decide what to do 'as a possible confused result' of the situations in which they found themselves engaged, some doing one thing, others another.

Williams presents himself at the same time as a theory-level relativist. In effect, we cannot, on his view, decide who is 'right'—the strict consequentialist who maintains that the cited actions are *always* performed in good worlds, the weak consequentialist or weak deontologist who maintains or concedes that they *sometimes* are, or the strict deontologist who denies that they *ever* are. Thus Williams subscribes to the world-level theory that confused and random action is often acceptable, as well as holding the entirely distinct meta-theoretical view that no general theory of agency can be formulated to deal satisfactorily with the dilemmas.

Williams's meta-ethical position is easily defended. The above statements and their variants, e.g. *It is permissible for X to kill Y if it will prevent five other persons from killing five others*, appear to be untargeted but nonetheless authored norms. It is implausible to suggest that this statement can some day, when we know more, and have thought more deeply into the matter, be confirmed or disconfirmed. The corresponding theories—weak consequentialism and its competitors—can each be expected to attract followers and defenders, as well as critics, but none can be confirmed. But what should we say to Williams's world-level claim?

It is unquestionable that we often act in a confused and improvisational manner. Agents stand under time and information constraints that virtually ensure that they will often act either in an intuitive and inconsistent way, or in a manner rigidly dictated by authority and convention, or sometimes in one way and sometimes the other. This can be considered a good thing in so far as it disburdens agents from time wasted on calculation or groping for principles and produces only slight moral harm. It is not clear, however, that this latter condition is mostly satisfied. We can therefore concede the meta-ethical point while continuing to press certain world-level points.

Moral theorists are, in some respects, like janitors—humble toilers whose work is never done. Someone has to do it, though. The burdens of world maintenance and world repair fall upon all our shoulders and dishevelled theories need attention too. We can all appeal for exemptions from the tasks to which we are least well suited, as from the obligation to furnish washing machines for all, but not from all maintenance tasks.

From the standpoint of the individual agent, there is only so much moral good any one of us can or wants to produce. Neither private citizens nor prescriptive moralists acting in the world establish the distribution of the global GNP; for the most part, they do not determine employment patterns or the availability of goods. And our internalized theory of responsibility takes us to be autonomous subjects and takes our actions to count more than failures to act. We all feel this: I am necessarily myself and only contingently a member of some group or other. What I do, I can be held responsible for but not the infinity of things I do not do. Argument and analysis can weaken these intuitions and show us that the situation with respect to agency is not so clear-cut. It can show us that we are not only individually self-interested creatures on the lookout for opportunities, but members of groups that wield collective social power in ways of which they may not be fully aware. In both capacities, we are able to formulate policies for ourselves to address those situations in which we hold anything from a momentary chance advantage over another to durable dynastic, political and economic power, and to think out policies for the collective, even if it does not always respond to expressions of authored imperatives and optatives. The cognitive illusions of perfect individuality, impeccability, and limited agency are powerful, and, like our intuitive theory of physical objects and properties, they are serviceable and resistant to re-education.

The moral theorist has an important part to play; he is not hampered by social powerlessness, as the agent is. In his prescriptive text, he can do what the ordinary agent cannot: control the distribution of the GNP and determine employment patterns and the availability of goods, so long as he respects the reality constraint and as long as his version of the idealism characteristic meets with agreement amongst a competent set of critics.

The following three chapters complete the shift from the descriptive to the prescriptive mode. The argument, it is hoped, will prove strong enough to elevate the qualified egalitarianism defended there from the status of a mere authored norm to somewhat greater objectivity. The argument that the sacrifices they require are too great for human beings as they are constituted by nature is frequently cited as a defeater of exigent moralities. It has produced a surprisingly resigned attitude towards socio-economic and sexual inequality in contemporary moral theory. Empirically minded persons may gloomily conclude that we are not the sorts of animals that are outfitted by nature to construct and inhabit egalitarian societies. Yet where moral theory is concerned, the post-utilitarian discovery of the significance of their personal well-being to agents themselves that centrally informs Williams's moral philosophy need not underwrite an ideological theory of exemptions and permissions. It is true that, as Williams maintains, utilitarians need to 'acknowledge the evident fact that among the things that make people happy is not only making other people happy, but being taken up or involved in any of a vast range of projects'.¹⁷ The immanentist defence of privilege embodied in the reference to the janitorial conception of duty should be rejected as incidental to the discovery in question.

¹⁷ Williams, 'Critique of Utilitarianism', 112.

The Division of the Cooperative Surplus

Societies in which it is possible to accumulate and store a surplus of goods tend towards a condition of social stratification and towards biases in the distribution of resources and the scope for agency that individuals within them enjoy. The provision of more goods and services in absolute terms to more people creates differently endowed and enabled economic classes. Despite the invention of efficacious techniques for growing crops, for converting raw materials into useful items of manufacture, and for economic forecasting and planning, the variance in levels of well-being across and within countries has increased and continues to increase.¹ Many human beings suffer from chronic parasitic and malnutrition-related diseases; others are well fed and remain healthy and vigorous into old age. Some live in cockroach-infested inner-city apartments with backed-up plumbing, others in luxury penthouses with soft towels and fragrant vases of flowers. Some have political and cultural influence and authority; others do not participate in public life at all. Security of person against crime and the brutality of soldiers and police, and freedom from surveillance and confinement are differentially experienced.

These qualitative differences supervene on basic economic facts. For the most part, countries and subcultures within countries in which incomes are low by surrounding standards are low on objective measures of well-being. Rich persons and rich nations are less vulnerable to luck than poor ones. The former enjoy 'budget redundancies' that enable them to weather the effects of poor production

¹ Variance is understood here as a measure of deviation from the mean that is not reduced or increased by changes of scale.

choices, natural disasters, and bets that fail to pay-off.² In short, ours is a high-variance world where the three basic goods of liberty, security, and happiness are concerned. The conditions of a life worth living are not met for many people who have not committed any criminal act, as well as for many people who have. In this respect, our distribution protocols fail to discriminate between persons convicted of the most squalid crimes and persons who have done nothing legally forbidden.

A curious visitor to our planet who considered the distribution of the components of well-being amongst all who cooperate to locate the materials for and produce these components might well write down the following questions in its notebook:

- (1) Does the liberation of human productive energies through technologies of material transformation and administrative organization require, entail, or cause a high level of objective deprivation amongst human beings?
- (2) Does the cultivation of human artistic and intellectual ability through the construction and operation of meritocratic institutions require, entail, or cause the subordination of women to men?

These questions are well worth pondering. ‘Perhaps the most significant contrast’, Margaret Ehrenberg remarks, ‘between forager societies and almost all others in the modern world is the equality between individuals found there, which contrasts with the very marked variations in status found in most other societies, particularly between women and men.’³ No one imagines that the deprivation of many is a sufficient condition of a high degree of well-being for a few or that the subordination of women is sufficient for the flowering of human creative abilities. It is easy to imagine a world in which deprivation and subordination occur without material progress. At the same time, it is clear that the deprivation and subordination of others contribute to the developmental goals many human beings, corporations, and national governments have historically set for themselves. Is this necessarily a moral problem? If so, what would an economically well-developed and complex world that was nevertheless just look like?

² See Douglas and Isherwood, *World of Goods*, 93. The notion of ‘budget redundancy’ is the economist Aaron Wildavsky’s.

³ Margaret Ehrenberg, *Women in Prehistory*, 52.

This chapter compares procedural approaches to understanding distributive justice—those focusing on how resources are acquired—with outcome-based approaches—those focusing on the resulting patterns of distribution. Both procedures and outcomes are shown to be regulated in just worlds. Contrary to the sceptical views of immanentists, we can make headway in stating the conditions governing a fair division of the cooperative surplus and in showing what greater social equality would actually come to. We can also appreciate why procedural fairness might have a positive relation to equality of outcomes without guaranteeing them, and why the theory of social justice is not, despite the fictional-worlds apparatus needed to articulate it, utopian.

Before taking up these topics, it will be useful to give some further backing to the claim that the notion of distributive justice is a centrally moral notion, i.e. one concerned with the relinquishment of advantage. This will help to lend support to the claim defended in the course of the chapter that the consent of the well-off to redistributive measures that worsen their condition is not a necessary condition of their legitimacy.

6.1. Is Variance a Moral Concern?

From the moral point of view, there is a *prima facie* obligation not to worsen the status of others to improve mine and indeed to improve the status of those who are worse-off by renouncing advantages I hold. These *prima facie* obligations are offset by obligation-weakening considerations, such as inefficiency and burdensomeness, which in turn are subject to override. Various formulas of obligation, of various degrees of exigency, are eligible for adoption as rules governing the distribution of valued goods and experiences. Some of the most stringent will quickly fall victim to the argument from heavy costs. Impersonally considered, they are too difficult for me or anyone else to comply with. The ideal of equality may be one such victim. Nagel, for example, argues that even economic equality, only one component of social equality, is too expensive for us, given our selfish interests and our competing values. The motive to acquiring more than others have is so powerful, he claims, that maintenance of

economic equality 'seems to require pervasive governmental control of individual life, serious denials of liberty, strict enforcement of general ignorance, and the absence of democracy'.⁴

Other philosophers go much further, suggesting not only that equality is an inappropriate goal but also that great disparities in outcomes are not necessarily morally objectionable. David Gauthier's description of the market as a 'morally-free zone' is shorthand for the widely shared view that, in certain departments of life, market-based notions of fairness can and should supplant the altruistic framework of morality in discussions of social justice.⁵ If the distribution of the components of well-being is justly regulated by markets, and if markets produce a condition of high variance, that condition is just and, a fortiori, morally acceptable. For it would be curious to maintain that a set of distribution rules generated results that were entirely just but morally unacceptable; such a notion of justice would be philosophically uninteresting. Gauthier's underlying position is best interpreted as the position that fair markets are sufficiently moral to ensure that they are interestingly just. Theories of just distribution, even if they are formulated without direct reference to moral considerations, proscribe theft, deceit, manipulation, economic blackmail, coercion, and other forms of advantage-taking. The issue, then, is not whether justice is a concept independent of morality because it plainly is not, but whether particular hypotheses concerning just distributions can be confirmed.

From the first-person standpoint, the mere fact that human beings experience different levels of well-being and deprivation appears to raise no questions about my personal moral status. A familiar observation is this: I can look on others' deprivations from an impersonal perspective, as would a visitor from another planet, and I can appreciate that human effort might be able to reduce a good deal of this misery if people were to do the right things. But the fact is, I don't hover over the world and I have little power to make anyone do the right things. I live in the world as a member of a number of advantaged reference-classes, as a citizen of a G7 nation, well educated, well paid, and in excellent health. Economic security, opportunities for meaningful work, for pleasurable consumption and self-expression,

⁴ Nagel, *Equality and Partiality*, 29.

⁵ David Gauthier, *Morals by Agreement*, 84.

companionship, and so on are all top notch. Preserving them matters to me—a lot.

From a metaphysical point of view, this good fortune is a lucky accident. The logicians tell us I could have occupied any other existing position in the world. That I am where I am, though not who I am, is a matter of luck. I did not choose to be born as a particular person. I competed fairly for the prizes offered without grabbing or elbowing others out of my way. Is the claim of the complement classes against me unlimited? Do not my interests, and my class interests, with which they may be bound up, my desire, above all, to get on with my life amongst my own people, generate their own claims, their own defences and exemptions? As Nagel observes, when we take up an exalted and purely contemplative standpoint, we can appreciate that no human is worth more or deserves more than any other human does.⁶ If a God were distributing the components of well-being, there would be no reason to prefer one human being to another, or to favour one country's inhabitants over another. However, no one is handing out goods from above. Their actual distribution depends on the summation of competitive and cooperative activities of individuals and nations, each of whom has a powerful interest in how things go for her, him, and it. The detached perspective, it seems, cannot be assumed as we go about our daily affairs. Whatever is displeasing to reason about a high-variance world viewed *sub specie aeternitatis* disappears when the world is viewed from here, from any subject position, not just the subject position of a privileged person. Williams's implicit assertion—for his parable has no force if not interpreted after this fashion—that moral criticism of 'Gauguin,' who represents the privilege of aspiration over the dreary ubiquity of need, is banal reinforces this point.

The psychological truth that we are self-centred and find it difficult to judge states of affairs impersonally is an important datum for moral theory. Yet we can give sufficient weight to the first-person standpoint in taking a high degree of self-centredness—perhaps even a greater degree of self-centredness on the part of the well-off than the badly-off—as a feature of the reality constraint when framing distribution rules for an ideal world. The fact that worlds with significant

⁶ Nagel, *View from Nowhere*, 190 ff.

levels of deprivation or subordination may be preferred to worlds with less deprivation by persons who imagine themselves to occupy the same favourable position in them has no theoretical significance at all.

Proper observation of the anonymity requirement in moral theorizing does not, however, necessarily favour the low-variance worlds produced by egalitarian distribution rules. As noted in Chapter 1, a world, or a world slice, consisting of ten happy people eating ice-cream cones is not morally better than a world of ten disgruntled people waiting for a bus. The former world has more hedonic content, but moral excellence and hedonic content are different notions. There may be nothing morally wrong in the world slice in which ten disgruntled people are waiting for a late bus; it need be no one's moral fault that the bus is late. By extension, there may be nothing morally wrong in a world in which millions of people have too little to eat or suffer painful and incurable eye infections; this too may be no one's moral fault.

One can easily imagine a world consisting of both deprived and well-off people that is morally equivalent to a world of only well-off people. If the well-off in the mixed world have not extracted their advantages from the deprived, and if they are not guilty of culpable letting-happen in refusing to ameliorate the condition of the deprived, the two worlds are morally equivalent. While I might on hedonic grounds prefer non-existence to the occupation of some roles in the mixed world, I need not deem it morally unacceptable.

Many deprivations, it will be observed, are unpreventable and irremediable by everyone, no matter how self-sacrificingly anyone wishes to act. An earthquake in a region where houses have been built cheaply to poor standards by incompetent workers might trap thousands of villagers, including the workers themselves, under piles of rubble where they die of thirst. They are much worse off than anyone not trapped; yet there might be no way to save them even with ample goodwill and exertion on the part of their neighbours. The disaster is partly a result of differential competencies, for the earthquake victims might be unskilled at building and too poor to seek the advice of engineers. An earthquake is not, however, a moral disaster unless someone stood to benefit from shoddy building practices, or some persons were able to help the victims, but had self-interested reasons for not doing so.

The emergence of the modern system of production, exchange, and compensation might be viewed as an episode in the history of the species that, like earthquakes, has been bad for some people without its being anyone's fault. It might be the case that there is nothing anyone in our mixed world is doing by way of promoting or maintaining his or her own interests that renders worse or fails to improve the condition of others. Indeed, a possible world can be imagined in which everything a person does to increase his own advantage benefits others ten times over. A high-variance world might have the interesting property that reducing variance by certain transfers from the well-off to badly-off improved their condition in the short run, but ultimately worsened it. Nothing anyone could do in this world would substantially improve the condition of anyone else.

We know nevertheless that behind the scenes of many natural disasters that were not brought about by advantage-takers—floods, epidemics, famines—as well as behind many non-natural disasters such as currency crashes, there is a clever or conniving human being or two, operating at the expense of a weaker party. And the differences in quality of life between individuals, subclasses, and whole countries that are barely hanging on and those that enjoy a surplus could be prevented by the renunciation of some advantage, an advantage in prices demanded or payments offered, for example. Our world does not seem to be one in which no moral action that reduces variance is possible because any limitation on the actions of the well-off and every transfer from well-off to badly-off is futile or self-defeating. If it were, there would be fewer anti-corruption laws, no progressive income taxes, no charities, public institutions or services, or rescue missions. Where action-hemming and redistributive proposals are concerned, political theorists position themselves at various points along the spectrum between the advantage-maintaining or -increasing centre and fringe hypermorality. Low-demand positions that take the advantages acquired by personal dominance—that is by the expression of aggression, charisma, beauty, intelligence, and familial prestige—to be inalienable are understood by their proponents to require careful defence. This suggests that questions of just distribution are tacitly understood as moral questions, and that advantages acquired by classes and nations require the same defence, if serious inequality is to be considered morally acceptable.

There are at least two further reasons beyond the allegedly accidental nature of deprivations and the alleged absence of direct harm or culpable letting-happen why one might conceive distributive justice and morality as unrelated topics. First, it might be argued that the theory of distributive justice is predicated on the assumption that individuals are struggling to increase or maintain their individual advantages with as little cost and trouble to themselves as possible. Questions of social justice concern the division of the cooperative surplus, the social product over and above that which could be produced by each of us working independently for our own sustenance. The question what regimen will come closest to giving us what we all want can be settled without reference to morality. To introduce a contrary action-governing principle, advantage reduction, is to miss the point of the exercise. Second, it might be argued that, while an ideal system of morality might permit individuals and groups to observe different moral rules, practising vegetarianism or eschewing it, or shunning marriage or institutionalizing it, a system of justice must impose the same rules on all participants who interact in a system of production and distribution. Its hallmark will therefore be that it represents a compromise between competing interests. Morality leaves open the possibility of moral choice in so far as few moral rules can be confirmed as right. Justice does not leave us any choices.

Regardless of the appeal of these arguments, a moral theory conceived as a set of principles for projecting a moral world will necessarily contain the elements of a theory of justice. Further, nations and legal units have inevitably to choose an imperfectly-confirmed theory of justice in the same way that individuals set their own levels for honesty, fidelity, or cruelty to animals, recognizing that these levels are open to criticism. Where moral rules typically represent reasonable compromises between the stringency of hypermorality and the non-moral interests of individuals, the rules of justice represent reasonable compromises between action-hemming restrictions believed to be necessary and appropriate and individual demands. Criminal justice, for example, restrains the criminal, but not for every moral failing, and not without a hearing, and the rules of retributive justice both moderate vengeful impulses and overcome the natural timidity that might otherwise be felt in the presence of powerful rule-breakers.

The rules of distributive justice seek to come to terms with the appropriative powers and desires of the able and vigorous.

Rawls's introduction of the difference principle is perhaps the most celebrated example of the employment of an advantage-reducing rule as a central element in a theory of justice. In Rawls's scheme, 'All social values—liberty and opportunity, income and wealth, and the social bases of self-respect—are to be distributed equally unless an unequal distribution of any, or all, of these values is to everyone's advantage.' Effectively, then, individuals are permitted to extract as much of the cooperative surplus as they are able, subject to various safeguards, provided they improve the position of the worst-off.⁷ Their freedom of action is thus limited by a prohibition that is surprisingly exigent, for many desirable goods and states—not only income and wealth, but opportunity and respect—are manifestly pursued in our world in ways that do not improve the condition of the worst-off, never mind everyone, and in fact exacerbate it.

Rawls justified the introduction of the difference principle by claiming that, while other conceptions of justice have prevailed in human societies, his does not conflict with the traditional notion he finds, e.g., in Aristotle, and is supported by reflection and by comparison of his with rival distribution protocols.⁸ Nor is Rawls's citation of an advantage-limiting rule an oddity. Both Gauthier and Robert Nozick, who make no presumption in favour of equality of outcomes, acknowledge limits to advantage-taking by deeming theft, violence, and parasitism illegitimate means of acquiring goods. They present ideal distribution schemes that might seem either utopian in their faith in natural equality, or else relatively low-demand, but they are not amoral. Gauthier's characterization of the perfectly functioning market as a morally free zone does not signify that ruthless advantage-taking is permitted in the ideal market, but expresses his optimistic idea that in an ideal market from which free-riders and parasites are excluded, additional moral rules regulating the distribution of advantage are unnecessary.

Rawls's principle can be criticized as unwarranted. Amongst the array of advantage-reducing rules governing various contexts that we could write and observe, why should we accept the difference

⁷ Rawls, *A Theory of Justice* (rev. edn.), 54.

⁸ *Ibid.* 6 ff.

principle and not a weaker or a stronger rule? It illustrates nevertheless the theoretical place in a theory of justice for an advantage-limiting rule. And if the principle is taken to apply not simply to the global problem—How is the social product to be distributed among the national (or world) population?—but to every local subgroup of the population, in every situation, it is a powerful—perhaps too powerful—advantage-reducing principle. On a local interpretation, in any practice, transaction, or mode of life in which *A*, *B*, and *C* participate, the only possible justification for a departure from equality of well-being as amongst *A*, *B*, and *C* will be that the worst-off member of the subgroup is benefited. Presumably, Rawls did not believe the population to have this interesting property of compactness under his theory of justice, though he evidently did believe that the difference principle ought to be applied globally, as well as nationally.

Theories of distributive justice are moral theories, in so far as they constitute hypotheses concerning permissible and impermissible forms of taking or maintaining advantage, and obligatory forms of distributing or redistributing the elements of well-being or their assumed preconditions, such as money, health care, liberty, or education. A non-moralized socio-economic system is one in which acquisition is wholly unregulated by moral ideation, whatever role is played by natural submissiveness or native partiality to kith and kin. A hypermoral social system is one in which the distribution is entirely regulated by compensatory principles. While all existing human societies are semi-moralized, a society of rival robber barons who live by plundering their neighbours approximates to the first ideal type, while a self-sufficient socialist welfare state that permits no one any luxuries approximates to the second.

We no longer imagine ourselves as living in a society of rival robber barons, though, only a few hundred years ago, many of our ancestors did. The consensus is that we rejected it after the Middle Ages for Hobbesian reasons. It was stressful and wasteful of human and natural resources. At the same time, the consensus is that we have also rejected the extreme levelling of the socialist welfare state for the reasons crisply summarized above by Nagel. This latter rejection raises the question whether normative political theory needs to be and ought to be concerned with outcomes at all. It suggests that

justice requires only the control of procedures—the suppression of robber-baron behaviour—and not any particular distributive outcome.

A procedural theory of justice considers only the mechanisms according to which the components of well-being are acquired and transferred by individuals, not the overall patterns those mechanisms produce. It regards any patterns produced by good procedures as just. To see why, despite their initial promise, procedural approaches fail to capture the betterness of just worlds and require supplementation by consideration of outcomes, it will be helpful to consider a basic prototype model of a procedural theory in some detail. Outcome theories of justice, by contrast, evaluate mechanisms according to the desirability of the pattern—typically, equality, or at least improvement in the condition of the worst-off—they produce. Though both approaches capture fundamental intuitions about justice, neither is self-sufficient.

6.2. Procedural Theories of Justice

Modern procedural conceptions of justice trace their origins to Locke. ‘Every man has a property in his own person . . .’, Locke writes. ‘The labour of his body and the work of his hands we may say are properly his.’⁹ Locke intended to show how the acquisitions of an individual—in his presentation, a tiller of the soil—are legitimate, inalienable, and entirely his to dispose of, if he has mixed his labour with resources available to all. Locke’s successors have adapted his acquisition model, according to which the well-being of the tiller depends on just two factors: his skill at tilling and the number of hours worked.

An extended model takes into account what the tiller can gain by exchange of the fruits of his labour. Suppose the tiller is able to reproduce by mitosis. After twenty years, he divides into two tillers, and each new tiller is able to split again after twenty years. The tiller-descendants, carrying their accumulated sacks of grain, spread out to find new land to till and the intrinsic quality of their plots of

⁹ Locke, *Of Civil Government: Second Treatise*, ch. 5 ‘Of Property’, sect. 27.

land is somewhat variable. The tillers themselves vary somewhat in their abilities, motivations, and personalities. Their holdings begin to vary as a result. Institutions such as banking and a property and grain market spring up over time, as well as new consumption opportunities, some of them prudent, some wasteful. Some personality characteristics and knowledge are transmitted from generation to generation of tiller, along with accumulated material resources; others are not.

Soon, there are thousands of tillers engaged in producing, exchanging, saving, and wasting, all of whom trace their ancestry back to Locke's original groundbreaker. The holdings of each of the original tiller's descendants depend on their skill at tilling and the number of hours they work, but also on their skill and the number of hours they work at buying and selling tillable land and managing their assets prudently, and even the skills and the number of hours their ancestors employed in educating their descendants in tilling, buying, selling, managing, and educating. Whether or not there are significant differences in the holdings of the tillers after many generations, one might suppose that each is entitled to whatever his labour and that of his ancestors has brought him.

Building on this idea, Robert Nozick originally maintained that outcomes involving large end-state differentials were acceptable so long as they represented a series of departures from an initially just condition effected by any number of intervening decisions and transactions that were themselves just.¹⁰ In terms of our model, provided there is no forcible seizure of tillable land and entitlements are respected, and provided buying, selling, managing, and educating skills are deployed without force or stealth (a moralizing condition, it might be noted), the distributions obtained will be just. Since everyone is subject to good and bad luck, lucky and unlucky events, even if they influence transactions, do not generate legitimate claims for redistribution.

On this view, the incidental benefits that accrue to Tiller Y from living in the same society as Tiller X if, for example, Tiller X has produced a technological innovation such as a wheeled cart that

¹⁰ 'The general outlines of the theory of justice in holdings are that the holdings of a person are just if he is entitled to them by the principles of justice in acquisition and transfer.' Robert Nozick, *Anarchy, State and Utopia*, 153.

others are able to replicate, is properly regarded as a matter of luck that neither increases nor decreases Tiller Y 's entitlement. There-being-wheeled-carts in Tiller Y 's habitat can be regarded either as a favourable background condition equivalent to a good climate and fertile soil, or as a heritable benefit that has been bequeathed to her compatriots.

But why should we accept the claim that the world will be just whatever the pattern of holdings turns out to be? An objection to the procedural theory just sketched is that historical processes of accumulated loss and gain undermine the connection between virtue and flourishing on which the intuition of justice originally rested; the theory is arbitrary, when taken diachronically. Suppose that A_1 and A_2 share a set of founder ancestors, but that, after 100 generations, the fortunes of various branches of the family have diverged and A_1 is well off while A_2 lives in poverty. Imagine that all holdings have been acquired and transmitted in a manner that is procedurally just. Further, no political and environmental events have interrupted or disturbed the acquisition and transfer of assets. No wars or conquests have interfered with the skill and effort of the ancestors, no rapacious leaders have differentially stolen from them, no benevolent ones have differentially endowed them with resources or powers, no plagues or spells of good weather have facilitated or ruined the efforts of some. All transactions involving the buying and selling of land and stock have been unforced and just. It is still the case that A_1 's present outcome depends on the competencies, hours worked, and decisions made by ancestors she has never met who happen to share some quantity of DNA code with her and it is unclear why her holdings are said to be just because they are a function of theirs.

The justice of their respective situations will be equally contentious whether the assets of A_1 and A_2 are acquired by mitotic division, or marital recombination, or are acquired through participation in a set of cultural practices that is handed down from generation to generation. Suppose again a common pair of founder ancestors and the subsequent isolation of two branches of the original family into two distinct clans. Each generation either contributes to the wealth-getting capabilities of its clan (e.g., by inventing machinery for extracting resources from the earth) or reduces its capabilities (e.g., by instituting

a badly thought-out taxation system). As is the case with the direct transmissions of assets, subsequent generations may undo the work of their forebears, or they may accumulate the advantage or disadvantage conferred on them. Wheel technology might be lost if several generations of fanatics decide that the wheel is contrary to the will of God. A_2 's present outcome still depends on the capability-reducing actions of other people, making her an unfortunate victim. If A_1 is lucky enough to have been born into a culture with democratic institutions, a low index of corruption, and excellent food-production technology, it is unclear why he deserves the benefits he enjoys relative to A_2 and why a theory of justice must confirm his possession of them.

Although it is sometimes assumed that Nozick must have intended to show that the enormous holdings of some persons in our world are just regardless of the deprivations endured elsewhere, this implication does not follow from his theory and did not represent his all-things-considered position.¹¹ Procedural theories based on acquisition and transmission are virtually worthless as practical tools for assessing existing high-variance societies. For to know that an existing distribution is just, on a procedural theory, we have to know that all transactions leading up to it were just. Though our suspicions may point one way or the other, it is impossible to infer anything about the justice of these transactions from an inspection of their results.

What are the probabilities, in case A_1 and A_2 are observed to differ greatly in their respective levels of well-being, that their current holdings are the result of just procedures for the acquisition and transmission of heritable wealth as sketched above? Three answers seem possible:

- (1) We can infer that the procedures have probably been just, since it is reasonable to expect that, after many rounds of legitimate accumulation and transfer, A_1 and A_2 will be much further apart than their ancestors were.
- (2) We can infer that the procedures have probably been unjust, since it is reasonable to expect that, after many rounds of accumulation and transfer, A_1 and A_2 will each possess close to the mean value of their combined holdings.

¹¹ 'Although to introduce socialism as punishment for our sins would be to go too far, past injustices might be so great as to make necessary in the short run a more extensive state in order to rectify them.' Robert Nozick, *Anarchy, State and Utopia*, 231.

(3) We can infer nothing about the procedures.

A priori, 1 and 2 seem equiprobable. On one hand, leveraging and compounding enable individuals to turn initially small advantages into larger advantages and to transmit them to the next generation. Polarization may at first strike us as likely in a fair acquisition and transmission game extending over many generations.

On the other hand, we know from experience that it is difficult for human families to preserve and increase their capital over a number of generations. Usually bad luck, politics, many children, or an irresponsible generation dissipate the family fortune. In a fair acquisition and transmission game, we might suppose, there must be regression to the mean. A second reason for suspecting that a polarized outcome implies unfair procedures follows from the assumption that transactions that are disadvantageous to one of the transacting parties will not be entered into by rational agents who choose voluntarily under conditions of adequate information. On the assumption that human agents are rational and that they are ordinarily well informed, it is difficult for a procedural theory to explain how significant variance can actually arise as a result of trade and exchange except through opportunism: the exploitation of others' bargaining weaknesses. Only on the realistic assumption that human agents are not fully rational, that they possess limited information, make decisions based on many factors other than self-interest, and that their choices are normally constrained, can it be inferred that they often enter into contracts that are disadvantageous to them. Polarized outcomes thus indicate that some people's cognitive and situational deficits have been turned to the advantage of others.

In short, we can infer nothing with confidence about the procedural justice of a society from the observed variance in well-being of its inhabitants. We cannot predict that fair procedures will tend to lead to strongly unequal distributions. Procedural theories of justice that focus exclusively on the share of the resources in the environment individuals can acquire by their own efforts and on their hereditary—i.e., familial and cultural—transmission accordingly afford either too much or too little critical leverage. Either the observed distributions lead us to the conclusion that the procedures that produced them must have been radically unjust, or the question of the justice of existing

arrangements turns out to depend on information shrouded in the mists of time. Yet it is worth trying to strengthen the procedural approach before tackling the question whether outcomes matter. This can be accomplished by focusing on the problem of a real-time division of a jointly created social product.

6.3. Basic and Symmetrical Cooperation

Productive activity is often depicted in social-scientific as well as in moral texts as solitary labour dedicated to finding, growing, or manufacturing objects useful to oneself and others. Though it would be more realistic to do so, economists do not begin their reflections with a tribe of humans or proto-humans composed of dominant and submissive individuals, and with differently specialized males and females, or with two tribes, one of which conquers the other and makes agricultural slaves of its members, or even with a village or urban society divided into working families.

Locke's introduction of his servant and his horse into the acquisition scenario suggests that he understood what he did not choose to explore and explain, namely, that the effective use of other living beings, not solitary labour, is the key determinant in the acquisition of wealth. For all elementary productive activities—gathering, farming, grinding, hunting, and building, as well as the most complex activities—were originally carried out by humans in groups, sometimes willingly and of necessity, but often under conditions of enslavement and principally for the benefit of others, and this has been well understood since the beginnings of scientific anthropology in the late nineteenth century. How an individual fares in the world depends less on her interactions with the non-human environment than on her interactions with other persons in cooperative ventures. A bank-teller does not perform the equivalent of digging turf for eight hours each day, extracting, as it were, his pay cheque from the available global resources. He is a participant in a cooperative venture aimed at generating and dividing profit. The apparent rationale for isolating the individual labourer in treatments of distributive justice is that it is a form of idealization that renders certain features perspicuous. Yet all such idealizations obscure the fact that nearly all

human labour is cooperative, and every neo-Lockean deserves to be queried as to the peculiarity of his starting point. If it is necessary to simplify why not begin with a stripped-down, idealized notion of cooperative labour?

Cooperative behaviour is sometimes identified with reciprocally altruistic behaviour, but not all reciprocally altruistic behaviour is cooperative, and not all cooperative behaviour actually benefits both parties. Some altruistic actions are non-reciprocal, like the care of mothers for children, and some non-reciprocal actions are cooperative. Cooperation does not require foresight, planning, or even communication regarding the cooperative activity. Mating, hunting, nest building, care of the young, and like activities are typical cooperative activities in the animal world. At the same time, cooperation is not any kind of autonomous activity on the part of two individuals that happens to mesh. The participants are aware of one another and adapt their actions to each other. Though the level of awareness and adaptation may be minimal in the case of two avian parents feeding nestlings, they do not try to push food down the throat of the same youngster at the same time.

In the human world, cooperation goes beyond a fixed repertoire, and it is appropriate to speak of cooperative activities as generating a surplus even in non-accumulating societies, for cooperation can reduce the amount of individual effort needed to survive or make survival more enjoyable than it would otherwise be. To be sure, cooperation does not always introduce efficiencies, or add utility, or even increase leisure. Generally speaking, however, a creature has no reason to cooperate if it is spared no effort and derives no benefit from doing so.

In any culture, humans need each other to survive. At a pinch, a Crusoe-like individual could build a hut, trap an animal, decorate a wall, nurse a child, and tend a garden, if she had already learned how to do so from watching others. In no culture, however, can people accomplish even these elementary tasks without being shown how to do so, and, in many cases, being helped to do so. The invention of tools, instruments, and institutional contexts for their deployment affords new and often unanticipated opportunities for cooperation. Large-scale building, farming, the practice of the arts and crafts, and trade require individuals to work together, and the escalation of

cooperation permits massive architecture, the production of organized bodies of knowledge and institutions of knowledge seeking, industrial manufacturing, and so on.

Cooperative relationships are rarely symmetrical with respect to the degree of autonomy of the cooperators, the eligibility of their roles, or the fraction of the cooperative product from 0 to 100 per cent that each cooperator receives. For it is frequently tempting to accept divisions that improve one's immediate prospects, even if one would do better in the long run to refuse to join into a project without a more equitable division of the product. The ease with which participation can be induced on terms unfavourable to one of the parties, and the room that the advantaged party has to make those terms less unfavourable, render these relationships open to moral criticism. Suppose 'basic cooperation' is defined as follows:

Basic Cooperation

Basic cooperation occurs whenever two or more individuals engage in concerted or coordinated actions that enable them to accomplish work that none of them could accomplish alone or could accomplish as easily, and when the cooperative product is apportioned between them.

Galley slaves and their overseer form a basic cooperative group, since concerted manpower performing coordinated rowing movements directed by the overseer enables a large warship to move across the sea, and no individual could make this happen by him- or herself. Getting the warship across the sea is nevertheless someone else's project, not the galley slaves'. The benefits of the cooperative product are partitioned, but most if not all the benefit goes to the shipowner. The subtask performed by the rowers is less eligible than the supervisory task performed by the overseer, for, given the chance, the supervisors would by and large not exchange places with rowers, while the rowers would exchange places with the supervisors.

Basic cooperation, with conditions added, attains the status of symmetrical cooperation:

Symmetrical Cooperation

Symmetrical cooperation occurs when basic cooperation occurs, and, in addition:

- (a) The work accomplished is equally important to both parties.
- (b) The material or psychological proceeds of the work, or the remuneration indirectly obtained for it, is divided between the parties according to the skill and effort manifest in the product, the performer of the task requiring greater skill and effort being better rewarded.
- (c) The material or psychological proceeds of the work are divided according to the eligibility of the subtasks they perform, the less desirable task being better rewarded.
- (d) Any indefinitely deferred benefits are equally shared in prospect and associated with real and timely efforts at realization.

The model represents the more diligent and more able worker as deserving a greater share of the product. It also represents the worker who is saddled with or elects the more disagreeable task as deserving a greater share. (These criteria are often in conflict.) It permits uncertain, symbolic, or emotional rewards to count towards satisfying the cooperators' claims, but it prevents delusory promises of reward from balancing tangible benefits. If a couple living in distant cities arrange their schedules and purchase plane tickets to maintain their relationship, the sustenance of their hopes for a joint future is a large part of the cooperative product, even if these hopes come to nothing. If A_1 enjoys free housecleaning, however, it is not balanced by A_2 's acquiring a happy but delusory expectation of marriage. The situation is similar if A_1 and A_2 cooperate to develop an online retail outlet and A_1 pays himself a handsome salary while A_2 acquires only stock options that prove worthless.

Most cooperative enterprises are asymmetrical. For, first, the success of most such ventures—achieving the intended output—is more important to one of the parties than to the other. (To keep matters simple, I shall treat multi-party cooperative relationships as though they were dyadic.) In the majority of cases, one party has more interest in the work being accomplished, and one or the other party may also enjoy the more eligible role, and may make off with the larger or more tangible fraction of the product. Any instance of cooperation by two or more persons affords opportunities for one to take advantage of the other while the activity is being carried out. Even where the benefits to A_1 and A_2 are immediate and real rather

than deferred and/or imaginary, the joint activity may saddle A_2 with the less eligible task, although the cooperative surplus is evenly distributed, or it may result in an unbalanced apportioning of benefits, although the effort invested is the same. Consider the following activities that can be carried out by two people:

- (1) A_1 and A_2 use a two-man saw to cut down a large tree.
- (2) A_1 and A_2 carry a heavy chest of drawers up a flight of stairs.
- (3) A_2 beats the egg whites and A_1 pours on hot syrup in a steady stream.
- (4) A_1 works in the stockroom and A_2 minds the till.
- (5) A_1 goes to the office and A_2 looks after the baby and cleans the house.

Assume that in each case both parties are aiming at a certain outcome, that they can effectively accomplish the intended tasks together but not alone, and that each receives some benefit from their cooperation. Nevertheless, even in cases in which the tasks performed are apparently mirror images of one another, one party may find it easy to exploit the other. In 1 A_1 can use light saw strokes on the return while A_2 uses heavy strokes, so that A_2 does most of the work of getting the tree cut down. In 2 A_1 can arrange matters so that she is at the top of the stairs and A_2 at the bottom so that A_2 does most of the work of getting the chest of drawers raised. As the separate roles of A_1 and A_2 become more differentiated, opportunities for asymmetrical participation or asymmetrical divisions of the joint product increase, and comparisons between the two roles become more difficult. In 3, A_1 can manoeuvre herself into a position where she performs the less tiring or more interesting subtask, such as pouring on the syrup; and in 4 and 5 A_1 can again fix the situation so that she performs the less tiring, more prestigious part of the joint work. The grander and more productive a cooperative venture becomes, the more likely it is to involve imperfect cooperation and the more imperfect the cooperation is likely to be. Building a pyramid—or a church—or staffing a textile factory are more likely to be basically cooperative activities than symmetrically cooperative.

The decision to cease competing as individuals and to begin to cooperate is often regarded as signalling the introduction of moral relations. In light of the forgoing, however, the introduction of some

level of basic cooperation has no unambiguous moral significance. It may mark the end of hostilities—or the beginning of organized warfare. The decision to join forces usually adds utility, but it rarely confers an equal benefit on both parties to the transaction, and it does not imply that both parties are equally motivated to accomplish the task they cooperate on, or that they work equally hard at it. Rousseau saw the loss of primitive autonomy as entailing the corruption of human society, and while his exaltation of solitude is difficult to take seriously, he was right to see interdependency as morally hazardous.

This point was anticipated earlier in the discussion of the evolution of cooperation in Chapter 1. If natural selection produces basic cooperation that is advantageous to a breeding group but not cooperation that satisfies some set of further conditions, it does not produce moral behaviour. Nor does it produce behaviour that resides in a morally free zone in the sense of being exempt from moral criticism. A_1 , in the examples above, is able to engage in basically cooperative activities while nevertheless gaining at A_2 's expense. While cutting down the tree, A_1 conserves her resources, forcing A_2 to expend more and, while minding the till, A_1 might acquire extra prestige at the same time as A_2 is deprived of social contact and is exposed to the risk of physical injury produced by lifting heavy boxes. Such situations are neither predicted nor accounted for by theories about the emergence of cooperation that regard individuals as equally powerful contractors, or that suppose that discrepancies in power cannot be maintained over the long run.

In the examples above, A_1 dominates A_2 , and A_2 submits to A_1 by engaging in basic cooperative behaviour with A_1 . Nearly all cooperative arrangements present tempting opportunities for one party to exploit the other by enlisting the other somewhat against his will, or without his being equally committed to the project, by assuming the more eligible task, or taking more of the cooperative surplus generated by joint effort than the exploiter's contribution to the project warrants, or by persuading the other to accept deferred or imaginary assignments of the surplus. But why does A_2 enter into a cooperative alliance with A_1 that favours A_1 ?

A_1 's social dominance may be either natural or acquired. She may have been born tall, with a loud voice, and an authoritative manner that makes others reluctant to question or challenge her. Or, though

born timid and retiring, A_1 may have set out to acquire a commanding manner and succeeded. The acquisition of wealth or titles may give her personality extra force. Or, she may have worked her way up into a position of leadership by completing a lengthy series of skill-testing exercises. If A_1 has been to cooking school and holds a diploma and A_2 has not, it will be seen as natural and just that A_1 should perform the more eligible tasks in the kitchen. It will be difficult in practice to decide whether any given manifestation of dominance behaviour is traceable to conventional or real, acquired or native attributes. Superiority can follow from the possession of purely symbolic indicators of excellence, or the possession of qualities deemed meritorious, or from sheer force of personality.

As Skyrms has recently pointed out, it is not always rational for self-interested A_1 to offer an unjust division of the social product to A_2 , and it is not always rational for A_2 to accept bad bargains even when his short-term situation is improved over what it would be were he to reject the offer.¹² Forbearing from domination and refusing to submit to attempted domination can be good *ESSs*; that is, strategies that do well when they meet themselves. But, where individuals are already sorted into the naturally or situationally dominance-prone and submission-prone, when they bear marks or carry flags to indicate the category to which they belong, and when the dominant can control the frequency and type of their own encounters, offering and accepting unjust bargains can become stable practices. (The dominant can, for example, agree to allocate subordinates, meeting and bargaining exclusively with them and not with each other: or they may bargain with each other on different terms.) The motives lurking in and pay-off to A_1 for offering the bad bargain are clear, but A_2 's motivations for accepting it need explanation. A_2 may have non-rational motives for cooperation, such as a love of service, or enhanced feelings of self-worth, or he may succumb irrationally to A_1 's social charisma. Otherwise, A_2 must believe either that the cooperative surplus will be fairly partitioned, or must believe that he will do better than by refusing to accept a subordinate position. Frequently, both types of motive operate. To explain why a free-living agent takes a position on the shop floor in a highly profitable company at a

¹² Brian Skyrms, *The Evolution of the Social Contract*, ch. 2.

very low wage, we might appeal to the agent's belief that (a) employment is more dignified than unemployment; or (b) the compensation is fair; or (c) although the compensation is unfair, there are no good alternatives.

That most cooperative enterprises are imperfect in the ways indicated does not imply that they are morally unacceptable. A_2 might decide to help A_1 , furthering a project that is mostly A_1 's or even A_1 's alone but that takes two persons to accomplish for minimal or no personal reward. If I answer your phone while you get a manicure, we together ensure that you get a manicure and that your phone is answered, but the first outcome may be a matter of perfect indifference to me and the second may not be very important. I might expect or hope for reciprocity in the future, but there is no reason to deny that there can be purely altruistic acts. But what is the difference between A_2 's helping A_1 and A_1 's exploiting A_2 ? Do secretaries help their bosses? Do auto workers help the manufacturer? If they participate in asymmetrical cooperation, are not helpers and are exploited, is exploitation always morally unacceptable?¹³ How to differentiate between helping and being exploited is a question that occurs to most of us episodically, for example when we take on some undesirable committee assignment or household responsibility, but it is a question that can be asked in connection with most cooperative enterprises.

To address these questions, we can take either of two routes: (1) We can try to supply a neutral definition of each term and then decide whether 'helping' is always morally creditable and 'exploitation' always morally wrong; (2) We can define 'helping' as 'morally creditable asymmetrical cooperation' and 'exploitation' as 'morally discreditable asymmetrical cooperation.' Neither route, however, terminates at a good answer to all questions. Attempts to provide a neutral definition of either 'helping' or 'exploitation' seem to fail. Building the normative implication into the definition, however, makes it harder to agree on criteria for evaluating actual cases.

By way of trying to create a neutral definition, we might begin by taking 'helping' to be A_2 's assumption of a disadvantage to create a benefit for A_1 and 'exploitation' as A_1 's putting A_2 at some disadvan-

¹³ See Alan Wood's useful paper 'Exploitation', in Kai Nielsen and Robert Ware (eds.), *Exploitation*, 2-26.

tage to create a benefit for herself. This procedure fails, though, to reveal a clear-cut difference. Or we might insist that 'helping' is the informed and voluntary award by A_2 of a benefit to A_1 and that 'exploitation' is the involuntary or uninformed award by A_2 of a benefit to A_1 . Unfortunately, this proposal does not provide a necessary condition for either helping or exploiting. Perhaps I have to force you or somehow induce you to help me wash the dishes, but your unwillingness does not always mean that I am exploiting you. Some jobs are exploitative although they are accepted voluntarily by informed persons. In any case, mentalistic criteria are notoriously difficult to apply. It might be suggested that an altruistic motive is present in cases of helping but not of exploitation. I might nevertheless be made to help without ever developing the motive, and the cultivation of love for and loyalty to her employer on the part of the employee does not ensure against her exploitation. Nor does the absence of a helping motive in an employee necessarily render her situation exploitative.

We can conclude that the role of the less-advantaged agent in a case of asymmetrical cooperation is often perceived to need a special explanation, since she has entered into a productive relationship with another in which her real share of the cooperative surplus is not proportional to her contribution, and agents who are both rational and reasonable are predicted not to do this. The explanation may be that she is gulled. However, she may also enter into such a relationship when she is as well informed as her counterpart. Two divergent patterns of explanation are then available and we employ the one that appears best to fit the situation: 'Model helping' is voluntary, cheerful, accompanied by an altruistic motive, and appears morally praiseworthy in the flow of advantage from the one with a surplus of resources to the needier party. 'Model exploitation' involves reluctant performance, constrained choice and resentment, and appears morally discreditable in the extraction of a benefit from a weaker party and its award to an advantaged one. These typical markers do not always cluster together. A secretary is likely to be cheerful and to have genuinely altruistic motives; she would not last long in the job if she did not have these attributes. At the same time, she is likely to be at least occasionally resentful of her place in the world and to feel her choices to be highly constrained.

According to the view under consideration, the goodness of a world with respect to its distribution protocols is a function of the degree to which its transactions occur at the symmetrical, as opposed to the basic, ends of the cooperation continuum. But it is not obvious that this is so.

Consider a world of perfectly symmetrical cooperators. They have our ordinary endowments—that is to say, as individuals, they would often prefer to be basic cooperators when it would be advantageous to them to do so. Yet, this being a *moral* paraworld, they control their impulses—or have created institutions that ensure that these impulses are controlled—in ways we suppose it possible for human beings to control theirs or to create institutions. They do not offer exploitative contracts, though they occasionally offer others help. We have reason to think that the world their actions produce will not have the extremes of well-being and deprivation we observe in our world. Accumulated advantage and disadvantage are virtually ruled out by the absence of privilege, for in this world, the better rewards offered to skilled and effortful productions are offset by the lower rewards offered for the performance of desirable tasks. Exploitation is minimized since workers cannot be paid off in emotional terms, with the promise of going to heaven, for example, and since the work must be important to both parties. The world under this description appears to be reasonably just.

Now suppose that this world is constrained by having no more resources than ours does and that its inhabitants have the psychological profiles of our existing persons. Many familiar tasks are not done in this world. These are tasks that are easy to arrange for someone to do under conditions of basic cooperation but difficult to organize under conditions of symmetrical cooperation. Large-scale agriculture may not have a place. Coal mining and factory production probably will not. Palaces, cathedrals, and monumental tombs are less likely to be built. Reflecting on conditions in such a world, a hard-headed realist might insist that we do not have an overall moral preference for the fair division of the cooperative surplus.

A decisive shift away from basic cooperation in the direction of symmetrical cooperation in many industries, the realist might venture, would fundamentally alter the material conditions of life for all of us. Indeed, he might add, it is the worst-off who benefit most from

the ubiquity of basic cooperation. Cheap food, cheap energy, clean washrooms in public buildings, affordable television sets and sportswear, are all by-products of exploitative labour relations and at the same time apparently vital to the well-being of those who cannot afford to pay a premium for these basic goods.

We cannot, however, take the argument that the reality constraint precludes more symmetrical arrangements very seriously. Confirmation from disinterested parties of the provocative thesis that our preference for a procedurally just world is constrained by our desire to see everyone well looked after, with sufficient food and respectable clothing, has not been forthcoming. Yet the objector is right to signal the importance of considering overall outcomes in assessing the justice of a world, not only procedures. It will be useful accordingly to give some attention to outcome theories of justice, specifically, to egalitarianism and the notion of equality that egalitarians consider definitive of a just world. The Rawlsian principle just invoked to the effect that one ought never to alter the procedures of a society in such a way as to worsen the position of the worst-off needs evaluation against procedural and outcome-based theories of justice.

6.4. Equality

If a willingness—in some cases an eagerness—to enter into basically cooperative relationships is not only a result of, but also maintains and generates, significant social inequality, it seems that by making a society more symmetrically cooperative we might be making it more egalitarian. Equality need not be considered, in this case, as a distinct component of justice. We could simply regard it as a by-product of a well-ordered society. And since equality of outcomes is easier to assess than the symmetry of cooperative endeavours, we could take equality as a good approximation for justice.

Egalitarians will insist, however, that more is required. Monitoring spontaneously arising cooperative transactions, they might insist, is not enough to ensure justice. Often, the reason A_2 is worse off than A_1 is simply that A_2 is unable to take part in certain kinds of cooperative enterprise open to A_1 . Moreover, when deprivations are acute and easily relieved by transfers from well-off to worse-off, then, no

matter how they are generated, there is moral fault in neglecting them.

These points can be conceded. Egalitarianism is nevertheless a problematic position. Why is equality good per se? What should be equalized? Resources? Satisfaction of basic needs? All desires, wishes, and velleities, as far as possible? Every day? As Walzer argues, if an equal distribution of goods amongst the population of the world were effected at 9 a.m., by noon the distribution would have become unbalanced. Some would have frittered portions of their endowments away, others would have mislaid them, and others still would have gambled on them and multiplied their holdings a thousand times, or compounded them by arbitrage, or have employed their wiles and talents to collect portions of others' endowments. Further, while reducing the variance in incomes has a powerful tendency to reduce the variance in well-being with respect to certain forms of consumption and participation, it is insufficient to reduce the social variance between healthy and sick, female and male, vital and depressed, and even between dark and fair. If God awoke one day and decided to make everyone equal, he would quickly realize, it might be argued, that he had no clear idea where to begin and what to do.

Egalitarians need not be fazed by such arguments. They may insist that we have a good idea what an omniscient and omnipotent being might do to realize not only economic but broader social equality. After ensuring that everyone had enough food in the refrigerator, well-baby care, and education to the level provided by a good private high school, God could add additional janitorial staff and tennis courts to every public housing complex. He could landscape highways, get rid of pole lights and billboards, and redecorate grocery shops, airports, and bus terminals where most people spend a lot of time. God could build more branch libraries, bring down the price of theatre tickets and canoe rentals, and put more books on tape for the blind. He could improve the taste of ordinary tap water and introduce universal daycare and community policing. He could package potato chips and doughnuts in tiny, beribboned boxes and price them at \$10 the ounce, thereby helping to equalize the health status of rich and poor. God could place mothers of schoolchildren on every National Security Council; for that matter, he could place schoolchildren on every National Security Council. In each case, God would be

repairing deficiencies and trimming excesses that result from the differential possession of expert knowledge, social power, and authority by individuals and groups with respect to one another. He would be working on making an egalitarian world.

But what are God's rules of thumb in the above scenario? How does he know what changes to introduce?, the sceptic will persist in asking. Will he also introduce free hard-core pornography for all, and addictive drugs? Intuitively, an egalitarian world is one in which everyone's proper interests, i.e. those interests of his that are neither perverse, arising out of a pathological state or process, nor based in delusion, are satisfied to the same extent. Yet many sceptics doubt that any such set of proper interests can be identified without prejudice. Immanentists may go so far as to deny that there can be meaningful interpersonal comparisons of well-being. Walzer insists in a Pyrrhonist spirit that 'There is no set of primary or basic goods conceivable across all moral and material worlds—or, any such set would have to be conceived in terms so abstract that they would be of little use in thinking about particular distributions.'¹⁴

The theory of proper interests, immanentists charge, presupposes an inappropriately zoological perspective on human life. The good life for parakeets and poodles, though it extends beyond what is required for biological subsistence, does not vary much from place to place. In the human world, by contrast, the region between the requisites for biological subsistence and culturally mediated desires and aspirations is not occupied by a set of needs that could be written down in a manual. People and communities, immanentists will insist, should be left alone to order their own priorities by experiencing the effects of their own choices and preserving or modifying their orderings in response. Apart from trying out various preference orderings, personally and communally, and experiencing their effects, there is no method that will reveal which orderings are best.

In defence of the egalitarian, it can be shown that proper interests are not as variable and mysterious as they are made out to be. The starting point in this case will be a priori reflection on the notion of a decent life that acknowledges at the same time the empirical

¹⁴ Walzer, *Spheres of Justice*, 8.

plenitude of actual or possible needs, desires, and preferences.¹⁵ To that end, it is profitable to distinguish between three categories of the components of well-being: the known, the presumptive, and the speculative and doubtful.

The known components of well-being are the biological requirements of survival. There are certain prerequisites of life and reproduction, including breathable air, potable water, and a certain number of calories. Doing without them is impossible, and anyone who does not have access to them is objectively deprived. While the lack of shelter, clothing, security of person, and assistance when ill may be compatible with survival given a favourable environment and plenty of good luck, they are usually necessary.¹⁶ These needs are often designated as fundamental, or basic, or urgent in Scanlon's¹⁷ sense and their remedies count amongst the first-tier goods. Their provision is, however, insufficient to ensure a decent human life. As Partha Dasgupta remarks, 'The meeting of these needs is a prerequisite for the continuation of one's life. Their fulfillment makes living possible. [But] for life to acquire worth, for it to be enjoyable, other sorts of goods are required. . . . This suggests that, roughly speaking, there are two tiers of goods and services.'¹⁸

While there is some dispute about the precise composition of a list of second-tier goods, there is general consensus that the following belong on it: opportunities for affiliation, for raising children; mobility; the right to refuse a proposed marriage partner or escape from an actual one without excessive social and economic penalty; freedom from harassment and derogation; some variety and pleasure in food, drink, furnishings, and ornaments; and the opportunity to engage in meaningful work, to advance one's knowledge and understanding, and to participate in decisions regarding the state or other civic unit in which one lives. The needs for these goods can be

¹⁵ The problem of objective deprivation and its grades was introduced into the philosophical literature by T. M. Scanlon, in 'Preference and Urgency'. Later contributions to its solution on which I draw below include Amartya Sen, 'Equality of What?'; James Griffin (ed.), *Well-Being: Its Meaning, Measurement and Importance*; Martha Nussbaum, 'Human Functioning and Social Justice: In Defense of Aristotelian Essentialism'.

¹⁶ An overview of the literature on primary goods can be found in Partha Dasgupta, *An Inquiry into Well-Being and Destitution*, 36 ff. See also Braybrooke, *Meeting Needs*, and his follow-up article, 'The Concept of Needs, with a Heartwarming Offer to Utilitarianism'.

¹⁷ Scanlon, 'Preference and Urgency'.

¹⁸ Dasgupta, *Inquiry*, 40.

termed 'presumptive' in the recognition that, while individuals can have decent lives without them, by and large, they are necessary.¹⁹ Presumptive needs are interpersonally and, to a large extent, intergenerationally presentable. They correspond to the grievances that appear repeatedly in letters home, songs of labour, loss, and deprivation, fairy tales and tragic novels; and these go well beyond material deprivations.

There is also a 'third tier' of goods, the speculative and doubtful components of well-being.²⁰ These are objects and states that are aspired to but that have a partial and tenuous connection with well-being; their pursuit is risky in a way that the pursuit of the presumptive components of well-being need not be. Third-tier goods may be keenly desired, but, obtaining them, the individual frequently feels no better off than he or she did before, once the initial moment of triumph and elation has passed. It is usual to think in this connection of luxurious or frivolous consumer goods and services such as high-performance cars, real and costume jewellery, plastic surgery, hand-tooled leather accessories, and so on, but there are other consumption states whose relation to well-being is equally speculative and doubtful. Being in a position to pursue scientific research in a well-outfitted laboratory, having an art collection or a spectacular portfolio of stocks to leave to one's children, possessing the leisure to write a novel or the money to design and build a house are consumption states that strike us as highly desirable, though we know at the same time that they may bring either ample, rich, and deep satisfaction or little more than trouble, anxiety, and disappointment. Most needs allegedly satisfied by exotic sexual experience or mind-altering drugs are of this nature, but so are many intellectual, financial, and artistic pursuits. The elation felt on attaining the desired state is often evanescent, and disappointment at falling short of expectations or ideals may be enduring. Unanticipated side effects are intensified when persons venture outside the realm of the presumptive components of well-being. The destructive potential of fame and achievement in a meritocratic society is intense, as our glances at the tabloids and our perusal of the biographies of the clever and powerful

¹⁹ For a fuller treatment along recognizably Benthamite principles, see Braybrooke, *Meeting Needs*.

²⁰ Richard J. Arneson discussed similar non-basic pursuits in a number of articles, including 'Perfectionism and Politics'.

remind us. The pursuit of third-tier goods can lead to isolation and loneliness.²¹

The good life is one in which the known and presumptive components of well-being are present, one that is not marked by unusually many or unusually heavy deprivations induced by restrictions on consumption, expression, affiliation, activity, and participation, or by insult and derogation that reduce social honour. Deprivations experienced within these categories appear to represent the main categories of diagnosable and presentable grievances, grievances individuals feel justified in expressing to other persons whom they know, and which they expect to be acknowledged and respected as legitimate. Agents will also tend to grieve the lack of third-tier goods when they possess all the known and presumptive components of well-being and are aware of the availability of these other more speculative and doubtful goods. Prescriptive egalitarianism rests on the highly plausible assumption that unless there is a specific reason—such as punishment or education—to deprive a particular human being of something he needs or wants, one should not do so, and unless there is a specific reason—such as punishment or pedagogy—not to furnish a person with what he wants and needs, one should do so. In good worlds, according to the egalitarian, people are never arbitrarily deprived and successful efforts are made to supply their needs and wants.

Now, a world of symmetrical cooperators is not necessarily an egalitarian world, even if we should expect to find a much greater degree of compression in it. The differential skill and effort of some individuals and their willingness to apply themselves to tasks considered less desirable should result in their receiving quite a bit more than average of the cooperative surplus. Hard-working and capable undertakers, for example, might do very well in such a world, while aspiring artists, musicians, and actors, who insisted on the more eligible subtasks, would fare poorly. Persons who engaged in solitary, hedonistic activities—lotus-eaters—would receive no portion of the cooperative surplus and would have to subsist as best they could. For a strict egalitarian, the resulting variance would be an unacceptable outcome. On her view, human beings should not differ in their possession of any of the components of well-being, regardless of their contribu-

²¹ Robert J. Lane, “‘The Road Not Taken’: Friendship, Consumption, and Happiness’.

tions to the social product, unless they are being justifiably punished or appropriately educated, in which case the deprivation is usually temporary and for their own good. Appropriate distributive and redistributive protocols are needed to ensure that the resources needed to acquire the components of well-being flood the population to the same level. In practical terms, this implies a series of transfers from persons possessed of skill, cunning, and accumulated advantage to less favoured ones.

Strict egalitarianism, then, as a prescriptive position, can be defined as follows. All components of well-being should be distributed as evenly as possible, regardless of what individuals contribute to the cooperative ventures in which they take part. Strict egalitarianism is in some ways a morally appealing position, but, for reasons that will be explored more fully in Chapter 7, it is not an entirely tenable position. Though it captures our sense that outcomes matter and are never guaranteed by procedures, it fails, as one might predict, to satisfy our basic intuitions about the existence of a procedural component to justice.

6.5. Justification and Consent

Theories of justice ought, one might suppose, to be formulated in accord with the anonymity requirement. Yet some versions of contractualism seek to do away with this idealizing apparatus, substituting agreement ‘on the ground’ as a condition of real obligation. Nagel insists in this connection that while ‘differences in bargaining power carry no moral weight in themselves . . . they can be given authority to determine results within a system that is legitimate by a standard of acceptability that is not the result of bargaining power’.²² Accordingly,

Justification in political theory must address itself to people twice: first as occupants of the impersonal standpoint and second as occupants of particular roles within an impersonally acceptable system. This is not capitulation to human badness or weakness, but a necessary acknowledgment of human complexity. To ignore the second task is to risk utopianism in the bad sense . . . The requirement of dual justification is a moral requirement.²³

²² Nagel, *Equality and Partiality*, 39.

²³ *Ibid.* 30.

For simplicity, let's concentrate on the implied claim that *if* one proposes to worsen the condition of the well-off by introducing new regulations or distributive protocols, it is necessary and *morally required* (not merely prudent, or politically expedient) to secure *their* consent as potential occupants of certain roles.

This claim appears to be different from Scanlon's claim, discussed in Section 5.1, that a policy acquires legitimacy by being acceptable to all parties who would be affected by it, assuming those parties to be not merely rational and self-interested, but also reasonable. Rather than appealing to Scanlonian idealization, to what reasonable people could not in good conscience reject, Nagel seems committed here to the view that the objections of even unreasonably self-interested people to redistributive proposals can render them illegitimate. This appears to be the force of the requirement of the second, personal justification allegedly owed to all, if the second requirement is to be functionally distinct from the primary requirement of impersonal justification. As Aristotle²⁴ pointed out, it is the weak who favour equality—and, he might have said—who specially object to being parties to basic cooperative enterprises. Numerous well-off persons and nations do not favour greater interpersonal or international equality or controls on the kinds of cooperative arrangements they may enter into. Is their reluctance to relinquish privilege and social dominance indefensible? Why should I, as a well-off person with many desires and interests, support redistributive policies that would reduce my opportunity to consume new and valuable goods and experiences? Nagel seems here to express Aristotelian doubt that the interests of the weak ought to carry the day regardless of what anyone else wants.

Yet the rationale for claiming that transitions to worse-for-someone conditions are morally illegitimate is hard to understand. If a given distribution scheme satisfying the anonymity requirement has been confirmed, what further justification from situated perspectives is now required? Nagel is right to insist that there is an intrinsic conflict between objective and subjective perspectives and that solutions to social problems must be satisficing and not perfectionist. However, a theoretical justification for dispossession that they are willing to accept is not owed to all existing persons, including the most

²⁴ Aristotle, *Politics*, 1318b.

grasping, socially insulated, and ignorant; this would be supererogatory. The consent of the existing class of well-off persons to it cannot be a necessary condition of the intrinsic justice of a system.

At the same time, it is right to say that a theory of justice must take the existence of the well-off and their interests into account. To rescue Nagel's notion of a dual justification, distinctions need to be made, first, between a distribution protocol and a redistribution protocol, and second, between the role of the theorist and the role of situated agents in a political context. A distribution protocol is a rule or set of rules for dividing the cooperative surplus in an ideal world in which no prior institutional structure is envisioned as having given rise to certain memories and expectations. A redistribution protocol is a rule or set of rules for reapportioning the cooperative surplus in a world in which it is already being allocated in certain historically determined ways, giving rise to memories and expectations that have psychological force. Transition costs now need to be taken into account.

When a theorist projects a distribution protocol, the interests of the existing well-off and the existing badly-off do not influence the choice of a theory. For the distribution protocol is simply the set of practices observed in the envisioned paraworld and no one there is antecedently better off or worse off. The existence of some wealthy and powerful persons is arguably desirable from an impersonal standpoint. As Bentham observed, persons who are much wealthier than others contribute to the aggregate utility. They make possible libraries and ballet troupes, Christian Lacroix ball gowns, and other culturally and aesthetically valuable objects and they furnish gainful employment by investing in wage-paying resource-extraction sites and factories. On the other hand, the wealthy and powerful produce these fine things by organizing basic cooperative enterprises and by resisting calls for greater equality.²⁵ Morally ideal worlds, we should conclude, can have very wealthy persons and nations in them, but only if they refrain from certain activities.

²⁵ Walzer's notion of 'complex equality' is a *prescriptive* theory that asserts that social dominance in the possession of one social good *ought* to be balanced by relative deprivation in another good so that such hegemony is not possible, not a *descriptive* theory of 'how the world works'. His *descriptive* claim that 'no social good ever entirely dominates the range of goods; no social monopoly is ever perfect' is true but ignores important effects of aggregation. *Spheres of Justice*, 11.

Designing a redistribution protocol is a different task altogether. How do ideal agents behave when they go about redesigning their tax system; or reforming their welfare, insurance, or pension plans; or offering subsidies; or imposing new tariffs? Assuredly, they observe moral constraints on the use of force, even to secure moral ends. They take into account the psychological costs of dispossession in avoiding too aggressive or too suddenly implemented policies. A good redistribution protocol will minimize disappointments and violate existing positive expectations as little as it can.

The politician's or voter's role, meanwhile, is different from the role of the theorist. To the extent that a theory of just distributions or just redistributions has been confirmed, those who are able to put it into practice ought to do so; a strongly confirmed theory deserves more effort than a weakly confirmed one. How much effort does an unpopular theory, or one disliked by the powerful and influential, deserve?

Suppose, for example, that the redistributive theory that there ought to be a tax on excess consumption (income less subsistence-plus-savings) of 90 per cent turns out to be reasonably well confirmed. Still, nothing will happen unless a politician advances a bill and voters support her. Since the theory has been reasonably well confirmed, I ought to vote for a politician who will advance the bill and she ought to advance it. The obligation to do either of these things is diminished *prima facie* by the difficulty of doing so. It might be very hard for me to vote for a high consumption tax if I am a rich person with expensive tastes and hard for a politician to advance the policy if there are many powerful others like me. Yet the diminished obligation is offset by ill-gotten gains, the contract principle, and even culpable ignorance. All things considered, the politician should strive to win my support for the policy and I should support it.

If I do give way and support a policy that is not in my interest, my consent does not contribute to the philosophical legitimation of the policy, though it is needed to make the policy actual. The philosophical legitimacy of the proposal depends on its being confirmed, not on its popularity, even if there is reason to think that confirmed proposals tend to collect followers. Justification is not, therefore, dual, though Nagel is correct to suggest that there are two processes: first, determining the goodness of the policy, assuming everyone complies with

it, and second, determining the acceptability of getting everyone to comply with it. The notion that a tax bill would not be legitimate if all actual well-off persons, or even all actual well-off persons deemed to be psychiatrically *compos mentis*, did not agree to it is plainly untenable. The policy would be politically legitimate if voted in the right way, and the policy would be philosophically confirmed if competent judges with a good but impartial understanding of just how bad it would be for the well-off endorsed it.

The claim that one can worsen people's conditions without gaining their consent for an overall benefit need not, as is often feared, open the door to the utilitarian abuse of individuals. We are not licensed to make slaves of other people for the overall benefit of the collective. The reason is not that they will not consent to this—for they might. Rather, it is that after long debate and much experience, we seem to have confirmed the proposition that worlds with slaves are impersonally worse than worlds without them. That those who are now slaves do not mind being slaves does not make a proposal to free the slaves illegitimate on the grounds that their consent and so universal agreement cannot be secured. If *I* would not consider my role as a well-adapted, consenting slave acceptable, I have reason to insist that no world in which slaves exist is even minimally morally acceptable. Nevertheless, it can be just to impose unwilling deprivations.

According to the externally imposed hardship and relative deprivation principles, burdens that shift a subject downward out of his normal reference class are typically experienced as heavier than freely chosen or naturally happened-upon burdens that are shared with others. While it might be a terrible thing for an eminent university professor were he forced to become a worker in a day-care centre, it is not necessarily bad to decide to be a worker in a day-care centre or to find oneself in that role by chance. A redistribution protocol should be misery-sparing and not driven by vengeance, but a distribution protocol that assigns someone to the role of childcare worker is not unjust.

However, the fact that a university professor would regard this as a terrible fate gives us some information about the nature of the work and its ordinary compensation. Service jobs, including nursing, rubbish collection, janitorial labour, childminding, and housecleaning have historically been poorly compensated. These

jobs are not held in low esteem because they are low skill, for they may require considerably more knowledge and expertise than clerical work or retail sales, but because they involve contact with taboo materials, such as blood and excrement, the products of putrefaction, and dirt. It is human nature to avoid contact with these materials, though individuals obviously differ in their levels of fastidiousness. These tasks are at the same time of profound importance to the collective, and where motives such as love and altruism do not suffice, economic coercion ensures that service work is done. When their importance is not recognized and appropriately compensated, it is bad for anyone to be a service worker.

Paradoxically, demoralization and resentment over perceived social inequalities are most profoundly experienced by persons who compare themselves with others whose income and mode of life is only marginally different from their own. They are rarely felt by persons one might otherwise consider to be objectively deprived. People's estimates of their own and others' conditions, their explanations of why they occupy the roles they do, and their estimates of the reliability of their own social judgements tend to be erroneous. W. G. Runciman found that groups that have clear relative deficits of income, status, and social power typically do not perceive themselves as unfavourably positioned. 'People's attitudes to social inequalities seldom correlate strictly with the facts of their own position. . . . Dissatisfaction with the system of privileges and rewards in a society is never felt in an even proportion to the degree of inequality to which its various members are subject.'²⁶

Two opposing conclusions can be drawn from Runciman's finding that objectively deprived persons are unable to perceive their deprivations and are disinclined to consider the reward system of their societies unfair. First, it can be inferred that the moralist ought to care only about subjective happiness or unhappiness, not about objective deprivation or statistical equality in the possession of third-tier goods. It is conceivable that the most stratified society with the fewest opportunities for social advancement will be the one in which people are subjectively the happiest. Attempts to reduce

²⁶ W. G. Runciman, *Relative Deprivation and Social Justice*, 3.

stratification require aggressive ‘consciousness-raising’ and may burden subjects with knowledge that subtracts from the happiness otherwise to be found in their lives. Alternatively, it can be inferred that distortions in the perception of their own state tend to affect the badly-off. This observation in no way entails that it is not better for people when their objective deprivations are reduced, or when their access to the doubtful and speculative components of well-being is not enhanced.²⁷

Robert Axelrod and his co-respondents have shown that two players in a game of Prisoner’s Dilemma who act according to the simple programme known as TIT FOR TAT, in which cooperative actions are reciprocated and defections are punished, experience more total reward than players who adopt other strategies. ‘Once cooperation based upon reciprocity gets established in a population, it cannot be overcome even by a cluster of individuals who try to exploit the others... The establishment of stable cooperation can take a long time if it is based upon blind forces of evolution, or it can happen rather quickly if its operation can be appreciated by intelligent players.’²⁸ Axelrod warns further that envy in games that are not zero-sum is counterproductive:

People tend to resort to the standard of comparison that they have available—and this standard is often the success of the other player relative to their own success. This standard leads to envy. And envy leads to attempts to rectify any advantage the other player has attained. In this form of Prisoner’s Dilemma, rectification of the other’s advantage can only be done by defection. But defection leads to more defection and to mutual punishment. So envy is self-destructive...

A better standard of comparison is how well you are doing relative to how well someone else could be doing in your shoes. Given the strategy of the other player, are you doing as well as possible...?²⁹

²⁷ Runciman studied English manual labourers, whose access to consumer goods such as central heating, telephone service, private schools, fur coats, and foreign vacations was, as one might expect, more limited than that of white-collar workers. Runciman, *Relative Deprivation*, 93. The subjective feeling of being well off can even increase, he found, with objective losses. *Ibid.* 23.

²⁸ Robert Axelrod, *The Evolution of Cooperation*, 189.

²⁹ *Ibid.* 111. Nozick takes a similarly dim view of envy in *Anarchy, State and Utopia*, 239ff. Rawls ‘assume[s] an absence of envy’ on the idealizing grounds that ‘it is generally regarded as something to be avoided and feared’. *Theory of Justice* (rev. edn.), 465, though cf. 470. For a different perspective on envy and ‘levelling down’, see J. Harvey, ‘Justice Theory and Oppression’, 176 ff.

One might be tempted to infer that exploitative strategies must have been eliminated long ago and that the envious are likely to drive down their own holdings. This inference would be a mistake, however. Perhaps we are moving towards better cooperation, but only slowly, under what amount to the blind forces of social evolution. Perhaps we do not act very rationally with respect to cooperation, because we are still more moved by emotion than by intelligence. Another possibility is that exploitation works as a strategy in our world, and can resist invasion by symmetrical cooperators for a long time. It does not work as a strategy in the games Axelrod describes, because the echo effect, by which defection is punished and punishment is retaliated against, produces a downward spiral in total return. But exploitation does work for one player when the other player is unaware of the defection, or cannot punish it, or is subject to retaliation involving multiples of his own punishing force.

Envy is not always rational, but it is arbitrary in an unfair division of the cooperative surplus to blame the envious rather than the greedy for bringing about the spiralling deterioration of relations. Like other irrational emotions, envy can send a signal that provokes a better response and thereby shortens the duration of an unintelligent exploitative episode.

6.6. Immanentism and the Argument from Inevitability

Both procedural theories and outcome-based theories of justice have a representational aspect. Their articulation and use presupposes that we can compare existing conditions with ideal distribution procedures or end-states. This assumption is taken by some critics to be misguided; it is considered an example of the confusion inherent in the very notion of a moral theory. According to immanentists, the patterns of deprivation and sufficiency that we observe globally are the product of historical processes within which individuals have shaped meaningful lives for themselves as best they could. There is no suprahistorical, non-local standard by which we could judge the absolute justice of any existing distribution, and no standard by which to judge whether a given scheme for the division of any quantity of

the social product is fair. Even the term 'distributive justice', an immanentist might insist, is misleading, for it calls to mind the image of a divine distributor or philosopher king who could get the distribution of goods and experiences right. But there is no such thing as getting it right. We have no idea what 'getting it right' could possibly mean. Immanentism is well expressed by Roger Scruton, who offers a version of the disqualification thesis (internal): 'Goods do not come into the world unowned, and rights are the work of a complex history of human interaction. Free ourselves from that history and from the long experience of conflict and cooperation that is resumed in it, and we simply have no conception of our rights, and no ability to assign them.'³⁰ Scruton's position is not that our actual, inherited arrangements and whatever system of compensation for our labours has emerged must be optimal. He does not claim, as invisible-hand theorists did, that the best outcome is secured by the free play of particular selfish interests. Rather, he thinks that we have no timeless conception of the right and the just that would enable us to judge existing institutions and the holdings of persons within them objectively. If we are disqualified from speaking of the whole, we are not in a position to think about reconstructing it according to formally elegant criteria. Alterations in the social fabric must be motivated by needs, initially perhaps somewhat inchoate, that surface in the immanent situation. The state to be aimed at is, he says, agreement between negotiating parties, each viewed as rational by the other, not rectification of a situation historically delivered and now imagined from a detached vantage point to be problematic.³¹

This combination of scepticism and commitment to procedure is somewhat unstable. A consistent Pyrrhonist would suspend the evaluative impulse fully, avoiding all endorsements, rather than accommodating a few elements of Kantianism in the appeal to the rationality of the participants or of social contract theory in the appeal to consent. In Scruton's account of the capitalist ownership of the means of production, as in Walzer's account of exchanged women, there is an invitation to the reader to agree that a situation is just

³⁰ Roger Scruton, 'Contract, Consent and Exploitation: Kantian Themes', 215. There is something 'oppressive and irrational', he says, 'in the attempt to state a complete vision of society'. Nozick made essentially the same claim earlier in *Anarchy, State and Utopia*, 160.

³¹ Scruton, 'Contract', 226.

because of the patient's aware and consenting role in it, even while the history of institutions in all their entrenchment and opacity carries most of the weight of their justification. However, the immanentist need not retreat from his basic claim that we are unable to judge the justice of practices in which we ourselves are entrenched or in which others are as entrenched as we are in ours.

The immanentist reminds us that existing employer–employee relations which govern distribution in local contexts are not the result of a well-reasoned adoption of a contract by beings who might just as well have been set down on earth five minutes ago. He might well have added that the manufacturing or service employee is not a new kind of human—a rational contractor—who appeared on the world's surface at the dawn of the industrial revolution, just when the 'natural slaves' whose farming and quarrying originally brought wealth into the world had disappeared.³² He is the descendant of that creature and retains some of his characteristics. At the same time, the wage labourer is delivered from the condition of being a serf in modern dress by the very fact that he does negotiate. Something radically new has, after all, come into the world in the form of the labour contract, whose evolution, as Henry Maine argued, required the dismantling of durable Roman structures.³³

Because the labourer is not a serf, the question of fair compensation arises as a concrete problem, and with it the need for the philosophers' unhistoricized terminology reflecting the most adequate conception she is capable, in her own moment, of offering. In that case, either current practices are the summation of a set of individual negotiations which appropriately benefited both of the parties consenting to the contract, or they are not. If the former is true, coercion has never played a role in shaping the contours of modern employer–employee relations; it has been fair negotiation all the way. Alternatively, current practices are merely those we happen to have ended up with. In that case, it would be miraculous if our present institutions were just.

Observing agricultural labourers or textile workers on and off the job on their compounds, few of us could distinguish free and slave varieties by observing what they did and how they talked, and the

³² Kenneth Lux discusses the transition and the origins of the idea that laziness is an inherent trait of the labouring classes in *Adam Smith's Mistake*, 174–5.

³³ Henry Summer Maine, *Ancient Law*, p. lxix.

same goes for a variety of occupations. Slavery involves the performance of work that is repetitious and uninteresting to most human beings. So does most specialized wage labour. The labourer improves his condition relative to what it would have been had he refused to enter into a cooperative contract, just as the slave does. To be sure, the wage labourer has exit options. No one will chase him and return him to the workplace in manacles if he turns in his resignation. He can go on strike. Nevertheless, his bargaining position retains traces of his old subservience. The employer has both sticks and carrots. He has something to offer—higher wages, better working conditions, more benefits—and he can make himself an object of trust and loyalty, as well as an object of fear. He also has reserves, and, frequently, options, notably the option to invest elsewhere, that improve his bargaining position. The labourer can threaten the employer; but he normally cannot offer him more than he already has if the workplace is efficient. Outside the negotiating situation, the economic underclass can threaten the rich both in their enjoyment of life and in their pocketbooks. They can enact criminality and sabotage, and present expensive health problems; at the same time, the underclass cannot easily offer to improve the situation of the rich man, rendering its negotiating power relatively weak.

Wage labour can be considered a partially moralized version of ancient slave labour. Certain advantages have been transferred to the originally less-advantaged party. The wage labourer in a modern society has the freedom to move even if he does not exercise it, and some freedom to change occupations. The advantages the employer is in a position to extract have been reduced, for the employer is required to pay a bankable wage, not only to serve self-interest by providing for the worker's subsistence. In some cases, legal rights and educational opportunities acknowledge no difference between the employer and her children and the labourer and her children, in accordance with the difference-erasing, generalizing tendencies of moral progress. These transfers have always been hard fought. Progressive legislation was historically resisted by the well-off, even when it was evident that liberal policies did not lift all boats and pushed some beneath the waves.³⁴ A fundamental question of

³⁴ See Lux, *Adam Smith's Mistake*, 177.

distributive justice is whether given instances of wage labour are sufficiently moralized to be deemed just. Have they been purified of their objectionable elements or do they still retain the mark of their origins? Given that the modern inheritors of the old privileged classes continue to extract labour from others for their own benefit, to what extent is the *prima facie* moral requirement not to exploit others weakened by overriding considerations?

Immanentists warn that conscious planning is often dangerous or at least utopian in the pejorative sense.³⁵ Nagel writes: 'Attempts to create a classless society have spectacularly failed the test of moral transformation so far. . . . It is no use to assert that we all ought to be working for the common good and that this requires the abolition of private property in the means of production. If the personal element of most people's motivation cannot be shrunk enough or the impersonal element expanded enough, a system of comprehensive private ownership seems doomed to degenerate under a combination of stagnation, nepotism, etc.'³⁶ Some twenty-five years ago, the socio-biologist E. O. Wilson expressed himself in a forthright manner, citing antitheory considerations and the heavy costs argument against greater social equality. 'We believe that cultures can be rationally designed. We can teach and reward and coerce. But in doing so, we must also consider the price of each culture, measured in the time and energy required for training and enforcement and in the less tangible occurrence of human happiness that must be spent to circumvent our innate predispositions.'³⁷ Both the philosopher and the biologist appear to be writing from an empirical perspective, signalled by such neutral terms as 'motivation', 'test', 'measure', 'innate', and so on.³⁸ This can produce the misleading impression that it has been scientifically proved that a condition of radical inequality is not only natural, but also of great antiquity, and not only of great antiquity, but also good for us.

³⁵ 'A theory', according to Nagel 'is utopian in the pejorative sense if it describes a form of collective life that humans, or most humans, could not lead and could not come to be able to lead through any feasible process of social and mental development.' *Equality and Partiality*, 6.

³⁶ *Ibid.* 27. The illocutionary intention behind Nagel's book is not to provide a philosophical defence of inequality but to make plain the psychological obstacles in the way of its reduction.

³⁷ E. O. Wilson, *On Human Nature*, 148.

³⁸ 'Societies are constantly trying to beat people into shape because they stubbornly fail to conform to some preconceived pattern of human possibility. . . . Political theory is. . . an empirical discipline whose hypotheses give hostages to the future, and whose experiments can be very costly.' Nagel, *Equality and Partiality*, 29.

As noted earlier, even beneficent policies may have unwanted side effects that make direct attempts at amelioration through social engineering risky. Perhaps God can raise the minimum wage without raising unemployment or inducing inflation, or provide more janitors to housing complexes without raising property taxes, but humans cannot. Alleviating a deprivation here causes another one there. Raise the price of T-shirts and fresh produce and global variance will be reduced but the domestic poor injured. Lower the costs of prescription drugs for some elderly women, and the pension funds of other elderly women will collapse. Provide an elite education to ghetto dwellers, and they will abandon their communities. Build more libraries and gardens with tax money, and the nurses will respond by going on strike. Critics impressed by these difficulties may perceive the *prima facie* obligation to reduce deprivation as weak.

The observed level of social stratification must represent a set of stable strategies, immanentists allege. If it did not, other systems would have invaded and replaced them, but experiments in the regulated distribution of the social product have turned out badly and have either disappeared or are on the verge of doing so. The historian can point to a series of revolutions and reforms and counter-revolutions and counter-reforms as indicating that real differences in people's abilities and characters will consistently work to undermine anyone's airy preference for symmetry over asymmetry.

People differ greatly, immanentists allege, where ability, ambition, and effort are concerned. Some command and take, others obey and give. These patterns replicate themselves at many locations in the social world, the implication is, because we, like other social species, function well and feel well within stratified systems. The presence of others below us makes us feel proud and secure, while the presence of others above us raises our level of aspiration. Fear of falling into a lower socio-economic class or hope of ascending into a higher one provide the essential motivation needed to accomplish the world's work. Moral Pyrrhonists do not usually object, though they may, to such unnatural impositions as the extension of the franchise to all adults, the taxation of incomes, the provision of social security for the old and ill, or antidiscrimination laws. They worry nevertheless about three egalitarian experiments: steeply progressive taxation, the

strengthening of trade unions, and affirmative action. These experiments were aimed at and in some cases succeeded in equalizing the distribution of resources, reducing disparities in the power of workers vis-à-vis owners, and distributing members of ethnic groups and men and women into professions and occupations in proportions corresponding better to the proportion they compose of the class of the talented and gifted. There is a widespread perception that the experiments were well conducted, that they have run long enough, and that their outcomes refute the hypothesis that greater equality is possible as well as desirable. They are perceived as having failed to reduce socioeconomic inequality, as having worsened human relationships, and as consuming time, effort, and money that would have been better spent elsewhere.

The philosopher seeking to falsify the thesis that stratification and high variance are at once natural, inevitable, preferred by us, and impersonally preferable is not, however, at a loss. He can point out that, wherever slave or semi-slave economies and status-based legal systems are replaced by presumptions of equal entitlement and more symmetrical contract relations, real deprivations have been reduced. He can collect evidence to show how social evolution may nevertheless be resisted and undermined by those who fear that they have something to lose. He will not dispute that there are powerful inequality-producing tendencies resident in humans or that the most respected minds in every generation from Plato and Aristotle onwards are likely to regard variance in outcomes as both inevitable and just. At the same time, folk history, with its litany of grievances against the powerful expressed in songs, fables, and novels, and political history, with its periodic uprisings, does not lend support to the view that inequality is everyone's underlying preference. Reputable economists increasingly point to the collective advantages, impersonally measured, of constraining income disparities.

Critics of high-variance outcomes perceive social institutions as working to maximize differences in income and influence by forcing intrinsically quite similar people through an obstacle course in which the loss or gain of a few points at each level can have dramatic consequences. They acknowledge variations in ability, ambition, and effort, but insist that they are too small to account for observed

outcomes and irrelevant to the human entitlement to most goods. They regard experiments in social engineering intended to counteract the ordinary workings of old institutions and historical drift as largely successful. The problem, they allege, is that they are not seen to be successful. The futility-of-redistribution argument is taken to be persuasive, and the successes of public policy advantage-redistribution measures are rarely trumpeted. More newsworthy are labour disputes, preventable deaths under national health programmes, and the rejection of meritorious job applicants for reasons extraneous to their qualifications. We do not often read about the superior productivity of the unionized trades and the greater self-respect of their members, or the substantial difference that state-financed medicine makes to the health of the poor and even the middle-class citizens of a country, or the personal satisfactions and opportunities to contribute to knowledge that affirmative action programmes have brought to thousands of people. It is helpful in making their case if egalitarians can show that the benefits of low variance have been underestimated and the costs overestimated, that current economic privilege reflects the ill-gotten gains of the past, is based in culpable ignorance, or involves the violation of an implicit social contract.

It is salutary meanwhile to realize that, even if opportunism and self-interest are frequently theorized as the source of all dynamism in life, morality in the sense of advantage reduction is ubiquitous in our competitive world. The existence of teachers, doctors, nurses, fire-fighters, and public transport workers has moral significance, even if the persons occupying these social roles are not conscious of the moral dimension of their work and think of it in terms of professional status, personal satisfaction, or pecuniary rewards. The provision of opportunities for individual ambition need not conflict with the existence of institutions whose *raison d'être* is the compensation of inequalities and power imbalances. Thus, one popular objection towards social levelling—that it is necessarily destructive of autonomy and hostile to talent—is removed.

To summarize, A_1 and A_2 cooperate for Hobbesian reasons—because conflict is expensive and they want to increase their productive capacity—but their decision to cooperate rather than compete does not make their relationship morally adequate. The initial moment of cooperation may announce the beginning of their moral

problems and our global interdependency has multiplied ours. The human willingness to engage in basically cooperative relationships results in a condition of high variance in well-being and objective deprivations for many. A reasonable theory of distributive justice must therefore incorporate some features of procedural approaches. Just worlds regulate and police their cooperative arrangements to dissuade agents from making and accepting the bad bargains they otherwise would.

Worlds whose cooperative enterprises are largely symmetrical as defined above are unlikely to contain strongly polarized populations of very well-off and very badly-off persons. Just as one hundred tillers on one hundred identical plots of land who consume all that they sow and reap are not likely to vary enormously in their levels of well-being before they begin to exchange, bank, educate, and plan, one hundred pairs of symmetrical cooperators helping each other sow and reap may not vary much in their individual levels of well-being either. However, if it is further posited that persons do not merely lend their skill and effort to cooperative tasks of a given degree of eligibility and claim rewards accordingly, but also bring accumulated resources—tools, capital, allies and acquaintances, prestige—that bring entitlements of their own, even a well-regulated world will begin to take on the familiar aspect of high variance. These resources increase the productivity of the cooperative unit. At the same time, they enhance the effects of innate skill and effort. If A_1 and A_2 have exactly the same cognitive abilities and are both willing to work ten hours per day, but A_1 can bring a factory or a fine old name into the cooperative enterprise and claim a much greater share of the joint product on that account, A_1 and his like may project themselves onto a different plane of well-being from A_2 and his like. Outcome approaches to social justice become more, not less relevant, as the procedural aspects of cooperation become more complicated.

But is equality in the possession of some or all of the components of well-being a feature of all just worlds? Strict egalitarianism conflicts with the intuition that worlds that do not discriminate at all with respect to the performance of individuals within them are worse than worlds that do discriminate. The strict egalitarian is committed to the position that persons who are wasteful, destructive, and corrupt, even

while remaining within the letter of the law, ought to have exactly as good lives as persons who add increments of beauty, truth, and justice to their environments.

Chapter 7 argues that just worlds are regulated with respect to outcomes as well as with respect to procedures. They are egalitarian, except at the margin. The distribution of the doubtful and speculative components of well-being is not uniform in just worlds exhibiting mild scarcity, though it does not favour some groups over others. This position is, it should be understood, considerably more stringent than the prevailing immanentist interpretation of the role played by considerations of meritorious performance in distributive justice.

The Role of a Merit Principle in Distributive Justice

Suppose there is a given quantity of desirable goods and states, the components of well-being, to be distributed according to some set of rules. Define *U* as a condition of the world in which needs are universally met. Define *M* as a condition of the world in which the components of well-being can properly be said to be a function of merit; i.e., the more meritorious are better off than the less meritorious, in proportion to their merit. Suppose we have the opportunity to write down a distributive protocol that will determine the allocation of the components of well-being to everyone. Which protocol should we choose if scarcity obtains to the extent that the complete satisfaction of all their wants cannot be experienced by all? Two very different, though not logically inconsistent, general answers suggest themselves:

- (1) *Merit priority*: We ought to choose one amongst the various protocols that realize *M* even if need-satisfaction *U* does not result.
- (2) *Need priority*: We ought to choose one amongst the various protocols that realize *U* even if reward-for-merit *M* does not result.

Of course, we might prefer a protocol that realized neither *M* nor *U*, and libertarians like F. A. Hayek have insisted that we should be uninterested in realizing either.¹ Credible libertarians do, however, seem to subscribe to some form of merit priority, as will be shown below; and it is unclear why we would not be attracted by the

¹ According to Hayek, 'The values [men's] services will have to their fellows will often have no relations to their merits or needs.' *Law, Legislation and Liberty*, ii. 72.

possibility of realizing at least one and possibly both of *U* and *M*. This chapter will argue that need priority is obviously true and merit priority obviously false. However, this does not settle the case for egalitarianism. Justice does not require that no one's life be better in any respect than anyone else's, only that no one's life be better than anyone else's in some respects. Meritorious performance can confer extra entitlements at the margin of well-being, provided suitable limitations are recognized.

The original dilemma—merit v. need—is soluble. It is the setting of an appropriate threshold for 'needs' that is controversial and there are reasons for thinking that the threshold should be set fairly high. Further, outcomes resulting from the deployment of a merit principle must be unpatterned with respect to various social classes if the conditions of a just world are to be met. In summary, semi-meritocracies are defensible institutions, but only under conditions that are not met in our world.

7.1. Two Distributive Norms

Merit priority and need priority reflect the manner in which two distinct distributive norms appear to govern all human economies. One is the norm of granting resources in one's possession, or under one's direct or indirect control, to those who need them, regardless of whether they have done anything to earn them. The other is the norm of distributing resources according to merit, regardless of need. The non-meritorious deserve deprivation, it is thought, just as the wicked deserve punishment, while the meritorious and virtuous correspondingly deserve fulfilment and reward. These norms may be termed the 'need principle' and the 'merit principle'. Merit is understood for present purposes to include the passive possession of qualities deemed admirable, such as physical beauty or insight into the structure of elementary particles; the active performance of tasks that require strength and effort, such as drilling or digging; and the exercise of coordinated specialist knowledge and effort, such as athletic, musical, or organizational ability that normally goes by the name of talent. I make no apologies for the whiff of theological rectitude and officers' deportment clinging to the term, for my

intention is to question the role of merit in a normative political theory that is consistent with our current empirical knowledge and democratic leanings.

The dual norms are expressed in institutions that are distinct but that coexist in the wider society. Meritocratic institutions distribute money, authority, and prestige on the basis of accomplishments and the outcome of selective and competitive encounters. Simultaneously, compensatory organizations including taxation bureaux, charities, and welfare programmes distribute the components of well-being on the basis of perceived need. The appearance of the two norms is not a special feature of modern market economies that have a monetary surplus available for distribution. Even if the extra privileges accorded to merit are minimal in subsistence cultures, the need principle never enjoys absolute hegemony. Honour and regard can be considered as elements of well-being that are somewhat scarce and that are differentially awarded in response to perceived merit.

The question whether both principles are ethical ones is controversial. A need principle might be introduced on non-moral pragmatic grounds, for example, as a condition of maintaining a workforce fit for exploitation. Additionally, many particular instances of alleged merit-based desert have no moral significance.² Ordinarily, however, the supplying of benefits to others at some cost to oneself and the appropriate distribution of rewards and punishments are seen as ethical performances.³ Yet the merit principle and the need principle are frequently in conflict. Many of the unmistakably neediest—the handicapped, the elderly, the impoverished inhabitants of badly ordered regimes, criminals, addicts, slum-dwellers, and the mentally ill—appear to those who enjoy some degree of control over resources to be deficient in valued attributes such as consideration for others, beauty, sound judgement, and industriousness. To award them resources is to violate the merit principle. Conversely, many persons held to be meritorious—film actors, athletes, CEOs of corporations—are not needy, and to reward them is to violate the need principle.

Interest in the two principles and recognition of their potential for conflict is a feature of modern political discourse. Rawls famously

² Though cf. George Sher, *Desert*, 7, *et passim*.

³ By e.g. J. S. Mill in *Utilitarianism*, in *Collected Works*, x, 241 ff.

argued that persons do not deserve their native endowments and therefore do not deserve the benefits that flow from them, including the opportunities further to accumulate advantages.⁴ Rewards for merit can at most have an instrumental role in his scheme: 'All social values—liberty and opportunity, income and wealth, and the social bases of self-respect—are to be distributed equally unless an unequal distribution of any, or all, of these values is to everyone's advantage.'⁵ Yet earlier scepticism over the merit principle articulated as well by Nagel⁶ has been countered by the rearticulation of meritocratic theories of desert. Liberalism, in its classic formulation, is a theory according to which individuals are permitted to accumulate as much by way of resources as their ability and effort allow, subject to certain background constraints. Rawls's claim that 'the democratic interpretation of [the principle of equality of opportunity and the difference principle] will not lead to a meritocratic society',⁷ has been deemed surprising, and he has been sternly rebuked as having abandoned the basic commitment that makes liberalism credible.⁸ At the same time, the need principle has remained at the centre of welfarism and is focal in capability theories.

Imagine a highly talented, hard-working pianist A_1 and a not-as-talented, less hard-working pianist A_2 . A_1 practises hard and productively for four hours every day; A_2 rarely practises. A_1 and A_2 complete for a full scholarship to a prestigious school of music. A_1 and A_2 exist in each of W_1 , W_2 , and W_3 . In W_1 , a perfectly meritocratic world, A_1 receives the scholarship and the subsequent benefits of the training he undergoes. In W_2 , an imperfectly meritocratic world, less meritorious A_2 wins the scholarship because of his good looks, or a mistake in the computer-generated list of finalists caused by a stray cosmic ray, though in W_2 such outcomes are somewhat unusual. In W_3 , an antimeritocratic world, A_2 wins the scholarship and has the better outcome for the same reasons and such outcomes are the rule because

⁴ Rawls, *Theory of Justice*, 89.

⁵ *Ibid.* 54.

⁶ Nagel took a stance similar to Rawls in 1973, arguing that 'differential abilities are not usually among the characteristics that determine whether people *deserve* economic and social benefits'. 'Equal Treatment and Compensatory Discrimination', 354. More recently he has emphasized the instrumental reasons for rewarding merit in *Equality and Partiality*, 102–3.

⁷ Rawls, *Theory of Justice*, 91.

⁸ See esp. Samuel Scheffler, 'Reactive Attitudes and Liberalism', in *Boundaries and Allegiances*, 82–3 f.; see also Sher 'Effort, Ability, and Personal Desert'; as well as *Desert*, 109 ff.

the more meritorious candidate is never preferred in W_3 . The merit-desert intuition is that W_1 is the just world, and that W_2 and W_3 are normatively inferior worlds, W_3 being even worse than W_2 .

The intuition that an injustice is done whenever the more meritorious are not favoured by a reward system can be expressed as follows:

The Merit-Desert Principle

If A_1 works harder than A_2 and/or has more talent enabling her to produce more goods and services of higher quality, A_1 deserves to enjoy a higher level of well-being than A_2 . It would be unjust if A_2 were to obtain as much as or more than A_1 . Merit does not merely have a tendency to attract reward in our world, but deserves it.

Belief in the corresponding principle as an axiom of distributive justice is fostered by reflection on examples of its contravention. In justice-conscious democracies, patronage systems that award the most lucrative and visible positions to cronies are regarded as grossly unfair and inefficient. Individuals are often outraged when they are passed over for a raise, a promotion, or a social honour in favour of colleagues they perceive as less industrious and accomplished, or whose natural assets and talents, considerable though they may be, seem irrelevant to the performance of the job in question. People's sense of fairness is such that they may prefer lower aggregates of utility to certain discrepancies, and this preference has always been considered a legitimate point against utilitarianism. Not only is envy a natural reactive emotion, it seems reasonable to distinguish between 'justified' and 'unjustified' resentment in situations of differential reward.

The merit-desert principle seems to operate ubiquitously and effectively and to correspond to a deeply held, non-revisable intuition about justice. We are constituted to expect that the exercise of our capacities will bring us gratification, and the habit of attending to the relationship between exertion and reward is useful to creatures who need to learn from their own mistakes. The belief that those who do not forgo short-term opportunities and resist temptations deserve to reap what they sow reflects the view that knowledge of what Bentham termed 'natural sanctions' is an important element of education. As Melvin J. Lerner remarks in his study of ordinary

people's beliefs about justice, 'In order to plan, work for, and obtain things they want, and avoid those which are frightening or painful, people must assume that there are manageable procedures which are effective in producing the desired end states.'⁹

The merit-desert intuition is so widely shared, the defender of merit priority might argue, that the principle of equitable but unequal distribution it generates is one on which everyone can agree. While I am not a talented piano player or an inventor of great proofs, and while I freely acknowledge that my chances of being one in any world in which my position was randomly assigned are no greater than the actual frequency of such persons in our world, my attachment to the intuition and to its particular applications may be just as great as that of the meritorious agent whom the realization of the corresponding distribution rule would benefit. A person of little merit might prefer to live in a W_2 or a W_3 world, it might be argued, but this could only be for selfish reasons, and could not constitute an all-things-considered rational preference.

It is the apparently non-discardable nature of the intuition to which Samuel Scheffler refers when he questions Rawls's view that 'the assignment of benefits and burdens in accordance with a conception of merit or desert . . . [requires] justification by reference to something putatively more fundamental: either to the utility of such assignments or to their placement within a larger institutional framework.' This position, he thinks, may be 'insufficiently insensitive to the role of . . . relevant practices in giving expression to our reactive attitudes.' Scheffler regards this discordance as potentially capable of relegating Rawlsian liberalism to a marginal position in political philosophy; it is too much at odds, he intimates, with our normal ways of acting and thinking about ourselves and others.¹⁰ A related argument is that to fail to reward merit is to refuse to participate in an appropriate, characteristically human mode of valuing. Offering financial rewards to the meritorious is, on this view, an extension of the natural human tendency to show appreciation and admiration for outstanding and even for creditable performances. That these rewards

⁹ Melvin J. Lerner, *The Belief in a Just World: A Fundamental Delusion*, 9.

¹⁰ Scheffler, 'Reactive Attitudes and Liberalism', 23-4.

can be used to purchase the components of well-being is precisely what makes them appropriate tokens of valuation.¹¹

For a defender of merit priority, the relationship between the intuition and the principle is as follows: *Because* reasonable persons mostly share the merit-desert intuition, the principle *ought* to regulate the distribution of the components of well-being. Some meta-ethical account of this inference is obviously called for. Normative intuitionists amongst its defenders will accept the inference on the grounds that it is a product of our direct epistemological access to moral reality. A more cautious meta-ethicist will base her acceptance of the inference on the grounds that the intuition is an expression of what psychologists term ‘consensus reality’, the set of beliefs, descriptive and prescriptive, that humans by and large agree on and believe one another to agree on, adding the further premises that our theories of distributive justice ought to be descriptive of consensus reality, not revisionary where it is concerned. In our own terms, the intuition incorporates, confusedly, a belief about psychological reality and a hypothesis regarding an ideal rule.

Though defenders of ideal markets as paradigmatically just systems officially reject the notion that levels of well-being ought to have any predetermined relationship whatsoever to specific qualities of individuals,¹² their commitment to the value independence of a market is clearly only partial. According to the founder-author of this tradition, Hayek, remuneration is tied to the perceived value of services provided.¹³ And, while there is no necessary connection between remuneration and guarantees of well-being, it is empirically observable that the probability that an individual’s well-being is above threshold levels increases steeply with increments of income above \$0 before levelling off. Market systems that are operating properly tend to ensure good outcomes for persons who produce goods and services deemed valuable, and while in theory these persons could be anyone producing anything to any standard at any speed, in practice these

¹¹ As Young remarks (without endorsing the view), ‘Many supporters of desert have argued that to fail to accord people their deserts is to fail to treat them as autonomous beings, as people who can actively and purposefully intervene in the world.’ ‘Egalitarianism and Desert’, 335. Cf. Malthus, *An Essay on the Principles of Population*, 95. Abolishing relief, Malthus says, will restore ‘liberty and freedom of action’ to the peasantry of England.

¹² Hayek, *Law, Legislation and Liberty*, ii. 72.

¹³ *Ibid.*

rewardees are individuals perceived as producing well-made and useful or valuable items in sufficient quantity. Liberty, in other words, does not only disrupt some patterning with respect to outcomes; it produces predictable patterning of its own. Perhaps because they recognize this peculiarity of the founder doctrine, subsequent market theorists have been more frankly meritocratic in their appeal to Lockean models of entitlement generated by virtuous individual productivity. Nozick's celebrated discussion of the former professional basketball-player Wilt Chamberlain's ability to extract resources from his fans by exercising talent and effort overtime is an imaginative extension of the idea of an entitlement that flows from ability and effort that are widely acknowledged and admired.¹⁴

Developing the Lockean idea of a natural entitlement to that which flows from one's endowments, David Gauthier introduces a thought experiment involving sixteen islands, on which there are sixteen different Robinson Crusoes.¹⁵ Each possesses one or the other of two polar traits: clever/stupid, strong/weak, and energetic/lazy. Eight islands have abundant resources, eight have scarce resources. All possible combinations of Crusoes and islands are realized. Each Crusoe employs her native endowments to exploit the resources of the island. No Crusoe has grounds for complaint, according to Gauthier; each does as well as she can given her ability, levels of effort, and the environment at hand. The level of well-being that each enjoys is not open to criticism.

Gauthier asks us now to imagine that the Crusoes acquire mobility and can leave their islands but can travel only in one direction, making certain kinds of trade impossible. Each Crusoe can now visit or try to settle on the next island downstream, or move further along. Gauthier asks whether it could be wrong for each Crusoe to use his own resources and those of any island he can reach for his own benefit. He concludes that this cannot be unjust, unless a Crusoe violates the proviso that one may not seize the products of another's labour without providing compensation. An invading Crusoe₁ is not permitted to make things worse for the resident Crusoe₂ than they would have been had he not turned up, but commits no injustice by not making things better. If clever, industrious Crusoe₁ arrives on an

¹⁴ Nozick, *Anarchy, State and Utopia*, 159 ff.

¹⁵ Gauthier, *Morals by Agreement*, 219 ff.

island where lazy, incompetent Crusoe₂ is failing to eke out a living despite an abundance of resources, there is nothing to criticize if Crusoe₁ does not share and waxes fat in the new territory.

Though according to merit priority, few if any of the products of the cooperative surplus require distribution on the basis of need in order for justice to be done, this is not to say that its defenders deny the moral relevance of the concept of need. Hayek acknowledges the need for a guaranteed minimum income for the non-able-bodied. Nozick describes as ‘reasonable and intelligible’ patterns of transfer including gifts, bequests to children, and charitable donations.¹⁶ It is safe to say that no one—not even the most Malthusian of living theorists of distributive justice—disavows the need principle entirely. (Malthus retracted his claim that the poor had no inherent right to survival.) Accordingly, the reader should not be distracted by verbal issues. To the objection that Crusoe₁’s failure to help needy Crusoe₂ is wrong, a defender of merit priority might respond that it may be morally wrong without being unjust. Gauthier notes that some categories of needy person are ‘not party to the moral relationships grounded by a contractarian theory’,¹⁷ but this leaves open the possibility that they are party to some other set of moral relationships. Nozick’s expectation that private charity would continue to operate to supply needs unmet by the merit principle reflects another view of the relationship of the two principles. All in all, while the denomination of rules as ‘justice-related’ or ‘contract-based’ v. ‘moral’ or ‘non-contract-based’ may be important in some argumentative contexts, this partitioning issue has no role to play in the present discussion. The relevant question is whether *some* kind of wrong is done when the need principle or another merit-disregarding principle fails to operate or is overridden in a particular context. When considering the distribution protocols of an ideal world administered by a perfectly just philosopher king, we are trying to compare the outcomes for different categories of person according to different norms, and the names by which the norms are denominated are not germane.

Though the first-person experience of the merit-desert intuition and its violation may be attended by sharp visceral feelings of justice

¹⁶ Nozick, *Anarchy, State and Utopia*, 160 ff.

¹⁷ Gauthier, *Morals by Agreement*, 18.

and injustice, comparable to those associated with retributive satisfaction or frustrated vengefulness, the intuition is not at all precise, and neither is the form of the principle to which it allegedly gives rise. The following considerations put its depth and universality into question:

- (1) Other principles besides merit are available for distributing not only basic goods such as food, water, and shelter, but also non-basic goods such as companionship and attention to persons whom we know. It is unclear why these principles are not available for distributing goods of all sorts to persons we do not know.
- (2) The merit-desert intuition dissipates when it is considered in isolation, purified of certain assumptions regarding overall sharable benefits. Its alleged status as a widely shared intuition of reasonable people or as a fundamental normative principle is accordingly dubious.

(1) It is a well-confirmed moral proposition that babies and elderly parents should have not only their needs for food, water, and shelter met, but should also be provided with the conditions of a decent life, even if they do not yet or no longer exercise their talents in an effortful way or display competence. We not only feed our pets but also try to ensure that they have pleasurable experiences, even though they do not apply their talents consistently to anything. Even prisoners guilty of heinous crimes who are held to lack merit altogether are given more than what they need to sustain life; they are allowed to watch television, for example, and this is not considered wrong. Homemakers, still a common occupational group in many parts of the world, fit the Lockean paradigm poorly, since their labour is not evaluated and compensated on meritocratic principles and they are often economically dependent on others. Defenders of merit priority are not inclined to deprive them of all the constituents of well-being; though they may well maintain that women's lesser status is appropriate in virtue of their lesser tendency to engage in competitive exercises. The burden of proof can be shifted to the defender of the merit principle who is required to explain why it should determine the allocation of resources outside, say, the family. The answer to this question should be non-circular, i.e., it cannot consist

in the re-presentation of the merit-desert intuition as uniquely suited to the public realm.

The defender of merit priority can readily concede that it is right to ignore the merit principle when one stands in a particular relationship of friendship or kinship to a needy person. According to the terms of the two-Crusoes story, a just state of affairs *may* obtain when an able and industrious individual who is able to extract resources from the environment (and being able to negotiate a particular wage settlement on the basis of one's merits can be considered as a kind of resource extraction, as we have learned from Nozick) refuses to accept a decline in his level of well-being to assist a non-meritorious individual in need. Gauthier certainly does not insist that an unjust state of affairs obtains whenever a meritorious individual *does* accept a decline in his standard of living to assist another. Nor does he deny that it would be morally wrong to fail to assist an elderly parent, one's own baby, etc. His account accordingly leaves open the question why, if it would be morally wrong (though not unjust) to fail to meet the needs of family members by forgoing the personal consumption of some reward for merit, it is not morally wrong (though not unjust) to fail to meet the needs of others by forgoing either the personal consumption permitted by a merit reward or simply the reward itself.

Several responses are available to the defender of merit priority. One is that there is no *natural motivation* to meet the needs of persons outside the nuclear family or outside some other special related group of kith, kin, and familiar animals and that the wrongness of failing to provide for intimates is related to its abnormality for most humans. The motivation to assist needy strangers is, however, susceptible to cultural, ideological, and ethical influences; what we take to be natural v. abnormal behaviour in this regard is not fixed. Another possible response is that the provision of the components of well-being to these 'exceptional' non-meritorious persons and animals is not inconsistent with the priority of the merit principle and is actually an instance of just reward. Babies and children will someday belong to the economic system as producers, the elderly once played their role, and homemakers and caretakers make the present productivity of the mature and the future productivity of young possible. And if criminals are fed, warmed, clothed, and afforded education and

entertainment at public expense, the defender of merit priority might insist, it is only because we can reasonably expect that some proportion of them—we do not know which individuals—will one day become deserving members of society. Certain social services, such as assistance to the blind and handicapped, might be deemed appropriate because they help them to function as consumers and producers in the marketplace.

Characteristically, however, defenders of merit priority worry about ‘utility sinks’—those persons whose irremediable past and present non-meritorious conduct, combined with an institutional readiness to transfer benefits to them, is believed to threaten the entitlements of the meritorious.¹⁸ A defender of merit priority will not avail himself of the observation that the non-meritorious are human beings who are in all relevant respects *the same as* the meritorious or that these persons *might have been* meritorious had they had different life histories since such generous allowances produce a *reductio ad absurdum* of the merit principle. Yet it is unclear why the principle should not be subject to such a *reductio*.

The defender of merit priority who concedes that it is not unjust and is indeed morally required to use funds one could have spent on oneself to feed one’s own non-meritorious baby is therefore obliged to explain why it would not be wrong to fail to meet the needs of non-meritorious strangers and would be permissible not to meet them. Suppose that Crusoe₂ is failing to sustain herself on her resource-rich island before Crusoe₁ turns up, and that she is gradually expiring. There might be various reasons for her poor performance. She might have numerous small children. She might be lame, or prone to dizzy spells and blackouts. Or she might be a hedonistic lotus-eater who lacks motivation. Whatever the reason, suppose she is economically incompetent. She lacks industry, ambition, and skill at finding and using the abundant resources of her island. Gauthier’s position is that there is no affront to justice if Crusoe₁ lands on her beach, meets her, and does not help her, so long as he does nothing to worsen her condition. He can gather up the fruits of the island for sale on the next one, while Crusoe₂, with her flawed body or bad character, heads towards her inevitable demise.

¹⁸ Gauthier, *Morals by Agreement*, 18.

Now one can agree with Gauthier that as long as Crusoe₁ and Crusoe₂ are alone on their respective islands, no questions as to their respective entitlements can arise. Nevertheless, it is unclear how the observation that an individual's level of well-being will be observed to vary with his labour on a desert island and that there is nothing wrong with *that* can be transformed without a great deal of argument into a fundamental thesis of social justice. As two members of a social species whose members are interdependent, Crusoe₁ and Crusoe₂ form a society even without formal agreement, and new considerations enter the picture. The existence of Crusoe₂ is prejudicially construed as merely another environmental condition experienced by Crusoe₁ on the new island, and the existence of Crusoe₁ as merely another environmental condition experienced by Crusoe₂. Even on a low-demand normative theory, it is wrong for Crusoe₁ not to feed and clothe Crusoe₂, particularly as it is not at all difficult for him to do so, given his cleverness and resourcefulness.

The difference between A_1 's culpably *doing something bad* to A_2 , and A_1 's innocently *doing nothing to stop something bad happening* to A_2 is mediated by a third concept: that of A_1 's culpably *letting something bad happen* to A_2 . The mediating concept is applicable because A_1 and A_2 can observe each other's actions and are aware of each other's existence and condition, and because direct or indirect means of affecting each others' condition exist and are known to them. Disregard for the needs of strangers, when one is aware of them and has direct or indirect means of supplying them, is unjust, as defenders of need priority have traditionally maintained.

(2) Even if consideration is restricted to able-bodied productive labourers, the merit-desert intuition is less robust than it might seem. It does not easily survive transportation outside the familiar scripts and scenarios of meritocratic competition such as contests, tournaments, raises and promotions, and political power struggles, in which more than simple desert may be at stake. Imagine a village in which the components of well-being are awarded to villagers in proportion to the quality and quantity of the products they manufacture, by which their merit is assessed. The products themselves may be returned as components of well-being to the producers, but the circulation of goods is not a necessary feature of the model. Suppose that the quantity of resources is also fixed: when the stock runs out, no

more will be distributed. For simplicity, we can begin with a village in an array of worlds W_1 , W_2 , and W_3 . Rather than digging resources from the ground and consuming them, the inhabitants manufacture small clay figurines, which they smash on an altar, and they receive the components of well-being in return from the gods. (The model captures better certain features of modernity.) The needs of the villagers are equal, indeed identical. All goods received can only be personally consumed and are not sharable. In a W_1 world, those who can produce the most and best figurines receive more by way of water, food, clothing, shelter, amusements, educational opportunities, trusty friends, and so on, and those who perform poorly receive fewer of the components of well-being or goods of inferior quality. In W_2 , the merit principle is partially followed, so that well-being is positively correlated with the feature of being a good sculptor, but certain disruptive factors such as luck, class, and sexual discrimination make the correlation imperfect. In W_3 , well-being is negatively correlated or uncorrelated with the ability to produce well-made figurines quickly.

It is not obvious that in this set of worlds, in which a unique skill is valued, the W_1 world is normatively superior to the others. Why should a person whose talent for producing figurines is limited and who does not work as fast have a smaller supply of drinking water, less wholesome food, and so on? Nor is it clear why it is preferable to suffer deficiencies because of lack of merit than to suffer deficiencies because of sexual discrimination. Because the stock of goods will eventually run out, being assumed to be fixed, neither meritorious A_1 nor incompetent A_2 will enjoy immortality. But if meritorious A_1 has accumulated reserves, he or she will experience a longer life than A_2 , as well as a happier one. And because the components of well-being are stipulated to be personal and cannot be shared, the argument that A_1 deserves a reward for acquiring additional resources that can increase the welfare of her society is not available.

It might be agreed by the defender of merit priority, following Hayek with slight revision, that, under the circumstances, A_1 has no surplus of entitlement over A_2 . This is so, he might maintain, because A_1 does not deserve any reward at all. For $A_{\pm 1}$'s productions are of no use to others in her society either directly or indirectly; they are simply requisitions of the gods and, contrary to the hypothesis, they

cannot plausibly be considered to be meritorious. Alternatively, following Dworkin,¹⁹ he might argue that the society of villagers is unjust because it has failed to equalize opportunities by allowing its inhabitants to display merit of various sorts depending on the skills they natively possess, or by providing basic training in the manufacture of figurines. These admissions, though, amount to substantial concessions on the part of a defender of merit priority. The requirements that only genuinely useful products of some durability are eligible to be considered as expressions of merit, and that either special training must be given to those who have no intrinsic gift when it comes to producing such items, or that there must be sufficiently many products that are regarded as valuable to ensure that everyone has sufficient talent and effort to produce at least one type of valued product, are quite stringent. Moreover, their introduction appears to be ad hoc and undermining of merit priority. For the problem the imposition of the extra requirements is designed to address is the problem of the needy; how to arrange matters so that they will have sufficient resources.

It might be argued that A_1 sacrifices leisure and endures boredom and physical discomfort to achieve her level of productivity, and therefore deserves compensation and extra rewards. Suffering through hard labour, it might be maintained, is itself meritorious. This claim is somewhat doubtful. As Nagel suggests, the possession and exercise of a talent may be intrinsically rewarding and external rewards will only further enhance A_1 's well-being.²⁰ Even if A_1 gets tired and bored making figurines expertly, she does not endure tiredness and boredom to help others. Her suffering seems devoid of moral significance.

To experience the uneasy feeling that equal awards of the components of well-being to A_1 and A_2 would be unjust, we have to build up a representation of A_2 as a shirker failing to pull his weight. This involves our conceiving resources as devolving upon A_1 and A_2 jointly, and conceiving both parties as able to determine the quality and quantity of their output. Incompetence will appear under these

¹⁹ Dworkin argues that individuals should be insured against the initial risk of being born without talents; all such deficiencies ought to be compensated at the starting gate. 'What is Equality?', pts. 1 and 11. The merit principle nevertheless plays an important role in his framework, in so far as a non-basic level of well-being is not assured as an outcome.

²⁰ See Nagel, 'Equal Treatment and Compensatory Discrimination', 357.

conditions as a wilful vice that deserves punishment and competence as deserving of reward. The position that *effort* is the only component of merit intrinsically deserving of reward, in so far as innate talent, though not its development, lies outside our control, has several defenders.²¹ Such a position flows naturally from consideration of the Lockean digging scenario,²² in which a person's nutritional status, shorthand for his overall well-being, is entirely contingent on the effort he puts into digging his fields, digging being considered paradigmatic unskilled labour that cannot be evaluated for expertise, grace, style, creativity, etc.

The chief objection to this limiting case version of merit priority is that the notion of effort required is an elusive one, located between a pure metaphysical will whose operations are empirically undetectable and empirically detectable striving that is a function of accidental cerebral organization and environmental influences. A secondary objection is that the intuition that effort deserves reward seems to fail in worlds in which the pure trying activity rarely or never produces results. (There could be a world in which those who do not try very hard produce the best quality and quantity of items and vice versa.) Ultimately, effort seems to be intrinsically deserving of reward only to the extent that we believe that as good, or nearly as good, sharable products can be produced by perspiration as by inspiration. The connection between the sheer goodness of a non-sharable product and the entitlement to the necessities of life is already weak and the intrinsic connection between either hopeless or successful effort to produce a non-sharable product and entitlement cannot be any stronger than it is. However, we cannot rule out the possibility that it is the effort component of merit and not the talent component that ought to count in any defensible application of a merit principle, whether instrumental or non-instrumental.

Summarizing, the merit-desert intuition is best regarded as an intuition concerning appropriate distributions of the components of well-being in well-defined situations in which the products of the meritorious are socially useful and can be bestowed upon others, or where the resources received for the products are at least sharable,

²¹ For Nagel, effort 'being a manifestation of the will, is the most personal or internal factor, and uniquely suitable to be regarded as the individual's personal responsibility'. *Equality and Partiality*, 106.

²² Locke, *Of Civil Government: The Second Treatise*, ch 5. 'Of Property', sect 27.

where the expression of merit is costly to individuals and entails sacrifices on their part, effectively *for* the good of others, and where we have a clear notion of what it is to pull one's weight and not to burden others. All these conditions are lacking in the figurine scenario. Further, the intuition is apt to operate more powerfully with respect to a condition in which the quantity of resources available for distribution is not fixed and is influenced by the reward scheme chosen. This condition was also left out of the model. The conclusion that seems warranted is that any obviousness resident in the merit-desert intuition is a function of certain additional assumptions, assumptions that are not clearly independent of the need principle.

7.2. Instrumental Considerations regarding Merit

The intuition that merit ought to determine the level of well-being is likely to materialize, then, when A_1 's talent and industry are conceived as influencing the amount of resources available to A_1 and A_2 jointly, and when A_1 is viewed as able to share or bestow them or even as required to do so. The use of a *reward* for capable A_1 's making resources available to less competent A_2 , or an *incentive* to encourage A_1 to bring down from heaven more of a sharable resource, now seems appropriate, providing respectively a non-instrumental and an instrumental justification for the use of a merit principle. Yet neither type of newly available justification supports the thesis of merit priority. For the intuition behind the appropriateness of rewards and incentives is that a world in which the needs of both the meritorious and the non-meritorious are met is better than a world in which the needs of only one or the other are met. And that is simply the thesis of need priority. For merit priority to be true, W_1 worlds have to be preferable to equally resource-rich W_2 and W_3 worlds in which resources are sharable but incentives are not needed to induce the talented to draw them down, or dig them up, as the case may be.

As noted earlier, some adherents of the merit principle take it as standing in no need of independent justification. It is this class of adherents who privilege reactive attitudes as genuine reasons for doing something whose political influence Scheffler worries

about.²³ The need principle, however, does not seem to stand in need of independent justification either; both are on an equal footing as basic commitments. The most compelling defence of merit priority is not posed by its proponents who maintain that the merit-desert principle is a fundamental and unrevisable prescriptive intuition but by those who maintain that history has shown that the deployment of the merit principle, ignoring needs, barring some few exceptions, is the most effective way to promote universal well-being and to minimize deprivation. This position is subtle, in so far as its defender implicitly concedes need priority but maintains that the universal satisfaction of needs cannot be aimed at directly but only achieved as a by-product. On this view, merit priority provides the proper operating principle, even if the satisfaction of needs universally is the goal.

Many reasons commonly advanced for the ubiquitous employment of a merit principle are instrumental in just this respect. It is better to have more competent pilots, bureaucrats, builders, and police officers than incompetent ones, and if rewarding persons in these categories according to their merit encourages competence, that is a purely instrumental reason to do it. A new good—competent piloting—is thereby brought into the world merely by organizing, or rearranging the reward structure, if not ‘for free’ then at least (it is hoped) for less than any utility subtracted from others. Disproportionately high rewards, such as CEOs’ salaries, create focuses of responsibility and accountability, preventing their organizational diffusion and this is seen as a general benefit. Similarly, one may overreward to encourage literary competency and productivity by giving large prizes on the basis of small and subjective differences in the quality of a product, for example, a novel. More good literature is made available to all through this kind of stimulation, thereby addressing the population’s need for good literature.

Yet these examples fall short of showing that a systematic commitment to merit priority maximizes the good that is available to all. It is impossible to make a retrospective study to decide whether ‘history’ really confirms the hypothesis of merit priority. The claim

²³ Scheffler, ‘Reactive Attitudes and Liberalism’, 31.

to obviousness of the historical thesis can nevertheless be tested in a simplified model.

Consider a mining operation in which there is a fixed amount of gold known to be underground and labourers are set to work to dig it out (the Lockean digging scenario). In each work cycle, the amount dug up is distributed according to protocol. Suppose the labourers are linearly ranked according to their merit; in this case, the ability to dig gold. The hypothesis is that the most gold will be brought out of the mine over some specified interval if the labourers are rewarded in a series of repeated cycles in decreasing order of merit. Since there is no unique way to fix the scale, and since a highly compressed distribution will be virtually identical to an equal one, the hypothesis should perhaps incorporate the specification that differences in payment between individuals or classes of the population are either (depending on exactly which version of the hypothesis is being tested) large by some objective measure or distinctly noticeable to the recipients—quite small objective differences might be strongly registered by them. A sub-hypothesis is that the reward scheme maximizes output through the production of positive and negative incentives. After receiving the first round of rewards, each labourer is influenced to increase or decrease his digging, or to keep production at the same level, and, according to the hypothesis, even if some labourers are induced to decrease their digging, their decrease is more than compensated for by increased productivity in others, which now merits a larger reward.

Stated in this form, the hypothesis is hardly uncontroversial. The assumption that low-merit labourers are by and large stimulated to produce more by receiving small rewards and do not defect to even lower levels of productivity, while high-merit labourers are by and large stimulated to produce even more by receiving high rewards and will otherwise defect to lower levels of production is gratuitous. How much an individual currently on Tier n of the payment hierarchy will be induced to increase or decrease his or her production by the addition or subtraction of $\$x$ of reward is an unknown. If the labourers are ascribed both biological limitations that fix their maximum digging abilities and their minimum nutritional needs and further psychological complexities, designing a schedule that produces maximum total output becomes a complicated task. When they

are able to observe each other's compensation levels and assess each other's merit independently of the 'objective' assessment, numerous motivational factors will be brought into play—including feelings of envy and feelings of solidarity—that will influence how they react to incentives. And rewarding the meritorious who do not need the incentive is inefficient, as is failing to reward those of the non-meritorious who are highly sensitive to incentives.

In short, if the aim is simply to maximize the quantity of gold taken from the mine by playing on the assorted psychological propensities and biological capacities of the labourers, there is no reason to think that a decompressed merit-based scheme must perform better than any of the multitude of schemes that depart from strict observance of the rule that the more gold one dug in the last cycle, the more one is to receive in the next cycle. The hypothesis that the system that maximizes the gold output of the mine will be one that compensates according to merit in this manner *might* be true, but if we ask why anyone is inclined to believe that it is true in the absence of proof, the answer is likely to be the following: The protocol is simple and is judged to give better results than other simple schemes, e.g. equal division. However, if a merit protocol is to be regarded as just on the grounds that it does the *best* job of meeting needs, its simplicity is irrelevant. If more complex schemes that work better are available, the rationale for the simple one collapses. Turning back to the real world, it should be apparent that if the quantity we are aiming to maximize through incentives and disincentives is aggregate well-being, not simply output or throughput of raw materials, the task becomes even more complex. (Increasing output can, for ecological as well as psychological reasons, reduce well-being.) The belief that a linear reward scheme will maximize that quantity is probably false.

In some domains, for example, in the civil service, introducing meritocratic competition and selection procedures in the place of patronage has opened up a livelihood to the middle and lower classes, and enhanced public service, since more competent persons now fill these roles. In other domains, the use of competitions and the introduction of differential rewards for merit has no significant beneficial effects. The efficacy of incentives varies with institutionally created expectations, and they may be ineffective where individuals'

productivity is relatively inelastic, or where they are not motivated by money. ‘Performance-based pay’, as Nancy Folbre points out, ‘may actually lower quality.’²⁴ When a formerly uncontrolled vocation, such as medicine or teaching, becomes organized under meritocratic principles, self-selection becomes less important in determining the composition of the profession. As a result, the values of the profession can change. This may, but need not, increase the total amount of good available for distribution. Voters would be reluctant to see high political offices filled through objective tests of knowledge of history, political theory, urbanology, finance, and so on that one would normally consider as qualifications for those jobs, and rewarded with salaries commensurate with the power and responsibility they entail. The assumption in force is that formal meritocratic selection procedures and top-notch compensation would not increase the collective good. This assumption may be and probably is seriously wrong. The point, however, is that people are evidently prepared to be accepting of self-selected leaders; those who desperately *want* the office. We infer from their desire a strong motivation to perform the duties of the office assiduously.

Though we might prefer politics to be more meritocratic than it is, these examples show that we cannot assume that the introduction of a meritocratic reward system always increases the total amount of good available in a given domain. There are other ways to increase the total amount of good available. Compensation that follows historical patterns, even where it does not reflect merit, enables people and organizations to plan efficiently, and giving disproportionate rewards to the noisy and overlooking the complacent creates the background conditions for signalling genuinely unjust states of affairs, even if the noisy ones are no more meritorious than their quiet counterparts. And persons whose qualifications render them scarce in our world are awarded a substantial premium for their participation in cooperative enterprises regardless of the merit they are held to possess. This efficiency saves us time we would otherwise have to expend worrying about the definition of merit. The cleaner’s success in purging the ward from staphylococcus bacteria with a mop and bucket might make more difference to patient mortality than the surgeon’s super-

²⁴ Nancy Folbre, ‘Roemer’s Market Socialism: A Feminist Critique’, 67.

skilled laser ablations. Yet we do not even attempt to reward the cleaner appropriately for her life-saving activities. For cleaners (and teachers of introductory logic courses) are plentiful, and surgeons (and philosophers of quantum mechanics) are not.²⁵ The incentive is not an incentive to perform better, but simply an incentive to choose one profession rather than another.

For structural reasons, then, it is all-things-considered good policy to award benefits that are not precisely keyed to merit and to tolerate their being awarded to others.²⁶ As Cohen argues, individuals who are correctly informed as to the large-scale consequences of their individual decisions to reward those whom they admire may change their values or reallocate their funds.²⁷ Increasingly sober views of executive compensation provide a recent illustration of the influence of information on values and preferences in this respect. In some cases, the total good available for distribution appears to be increased by the introduction of meritocratic incentives, but without being widely or evenly distributed. An increase in the total amount of good available may not mitigate widespread deprivations. Undoubtedly, extra competence is produced by a highly competitive medical system under which doctors compete to be top-rated highly paid specialists in prestige areas such as cardiology. Yet its benefits are consumed by relatively few persons. 'Community based' health care is offered by more numerous, lower-paid, and self-selected rather than specially enticed and competitively selected persons and arguably does a good job. Finally, there are good instrumental reasons to under-reward the meritorious by diverting the components of well-being to those identified as non-meritorious. The State recognizes that the industrial or post-industrial economy, which its laws and courts legitimate, is cyclical in its employment patterns. The least competent workers must be maintained during downturns so that

²⁵ Cohen presents some new arguments against the provision of incentives in Rawlsian fashion. *If You're an Egalitarian . . . ?*, pp.124 ff.

²⁶ Consider the following list of real, 'institutional' determinants of compensation: Formal qualifications earned, e.g. test scores, degrees, publications; natural advantages, e.g., beauty, height; seniority; prestige of work role, e.g., manual, white-collar; historical patterns of reward in the work location; scarcity/plentitude of role-fillers; degree of moral/legal responsibility for outcomes; history of labour disputes; return to investors; employer's overhead; funds lost, wasted, or diverted to corrupt purposes; state's overhead; state's provision for present and future non-earners. Their positive relationship to merit is either partial or non-existent.

²⁷ G. A. Cohen, 'Robert Nozick and Wilt Chamberlain: How Patterns Preserve Liberty'.

they are available as a reserve. And in a world in which either real merit is properly rewarded but very unevenly distributed, or evenly distributed but improperly rewarded, the frequency of violent revolution, terrorist acts, criminality, sabotage, slummification, and the risk of disease will be much lower when low- and non-earners are sustained and supported.

It might be suggested that, in a paraworld in which prudential motives operate powerfully, fewer transfers of goods from the well-off to the badly-off are required to assure that needs are met. And a world in which people are motivated to support themselves and do so is morally better, it might be urged, than a world in which they are the passive recipients of the bounty of others. Accordingly, transfers are not the only way to reduce variance. If we could induce people to overcome their bad habits, indolence, and demoralization, educate themselves, choose better leaders, and plan and save in ways that are more rational, we would not have to make transfers. If we apply the right kinds of pressures, it might be argued, incomes and holdings will come to exhibit less variance without the need for heavy-handed redistributive impositions.

In a zero-sum economic game, however, the end result will be the same, and there are not always clear advantages to trying to reduce inequality by moral suasion v. direct transfer. Inculcating prudential behaviour in a resistant subpopulation may be coercive and costly. While exhortation and encouragement are free, or nearly so, and relatively non-invasive, they may have little effect on that subpopulation. Truly efficacious measures may turn out to be more expensive and psychologically distressing than providing the intransigently imprudent with the components of well-being directly. Further, redistribution often does foster prudential behaviour in the non-resistant. The possession of some of the components of well-being such as education, housing, and childcare with more in prospect tends to increase the motivation to defer gratification and assume responsibility.

In summary, it has never been shown that there is a strict and positive correlation between the depth and scope of a society's commitment to meritocratic principles and its ability to meet the needs of all its members. Rawls has been criticized for maintaining that 'the democratic interpretation of [the principle of equality of

opportunity and the difference principle] will not lead to a meritocratic society'.²⁸ Criticism of his opposition to the merit principle has, however, been based chiefly on the affirmation of the principle, not on a posteriori comparisons. A meaningful comparison might consider two populations in which the variance in merit (however it was established) was the same, one of which was wage-compressed and the other not, controlling for variables unrelated to compression per se such as military expenditure and corruption. The aim would be to determine whether the 'historical' hypothesis that meritocracies have less objective deprivation was true.

Instrumental arguments presupposing need priority for the deployment of merit principles, then, are no more powerful considered in isolation than arguments from pure entitlement or desert that ignore need. Given that the amount of resources available for distribution can be viewed as a dependent variable rather than a fixed quantity, the distribution system selected is one but not the only factor that determines it. The quantity R of the components or preconditions of well-being that can be made available in a given world is a function of many interacting variables, including its temperature and vegetation, the food preferences of its inhabitants, their muscular strength and motivational levels. A given reward structure may be one of these interacting variables, enabling us to speculate that:

$$Wa f(x_1, y_1, z_1 \dots) = 1R$$

$$Wb f(x_2, y_2, z_2 \dots) = 2R$$

where Wa and Wb are distinct worlds and the x and y variables represent different conditions that influence the availability of the equal 'underground' or potential resources of a world. Wb in the example has twice as many resources available for distribution as Wa . For some pair of worlds taking the values $1R$ and $2R$, the *only* difference between them might consist in the presence and absence respectively of a given merit principle. But we cannot give a *general* answer to the question: 'How much does the employment of a merit principle—e.g., one that rewards exactly proportionately to merit—contribute to the size of R ?' It depends on the world. In a world in which there is no variance in merit, it will contribute nothing; in a

²⁸ Rawls, *Theory of Justice*, 91.

world in which there is considerable variance but the population is not moved by monetary incentives, it will also contribute nothing.²⁹ In a world in which ripe fruit is easy to pluck from the tree and desires are few, it may contribute nothing as well.

Whether, given the conditions of *our* world, we should prefer to see more or fewer meritocratic policies in operation depends on the potential resources it contains and how long they will last, how long we want them to last, the distribution of merit in the population under consideration, on the motivation and emotional tendencies of its members, and on their signalling and signal-reading abilities. The level of variance in well-being in the three types of worlds is independent of their meritocratic or antimercitocratic character type. A meritocratic W_1 world in which everyone's degree of merit happened to be the same would have low variance of outcomes, but so would some W_2 and W_3 worlds. And a state of high variance in well-being may obtain in worlds of any type. While a W_1 world in which merit is distributed equally and signalled reliably would distribute the components of well-being equally and would give no one cause for complaint, a W_1 world in which there are significant numbers of persons who possess less merit than others, or in which merit is signalled very unreliably and is difficult to detect, has disadvantages for both the meritorious and the non-meritorious.

7.3. Detecting and Assessing Merit

Although Hayek's version of the just protocol requires individuals to produce goods and services that appear valuable to gain rewards, he presumably does not regard it as just to reward individuals who only appear to be producing valuable goods and services. Only some purveyors of illusion are considered to be justly rewarded even under the most libertarian of regimes. An evening at the theatre is good, but not an Initial Public Offering for a fake technology.

²⁹ If the value of R supervenes on assignments to the variables, there has to be some additional qualitative difference between the two worlds if $f(x, y, z \dots) = 2R$ in both cases but the worlds have different reward schemes. The difference might simply correspond to the inhabitants' views of what constitutes acceptable body size, or their adaptation to a coarser diet, or their ethical beliefs. I take this to be, generally, Cohen's point against Rawls's instrumental defense of incentives in *If You're an Egalitarian . . .*, n. 28.

The contrary claim that the merit-desert intuition is indifferent as between real and apparent merit can be tested by rewriting the principle as follows:

As long as A_1 is *believed* to work harder than A_2 and/or is *perceived* as having more talent enabling her to produce more goods and services of perceived higher quality, whether or not she actually does, A_1 deserves to enjoy a higher level of well-being than A_2 . It would be unjust if A_2 were to obtain as much as or more than A_1 .

Written in this form, the claim is clearly unacceptable. A wine producer who accumulates a fortune before it is discovered that his synthetic product is fake and fraudulently labelled is viewed as deserving of going to jail. It is considered irrelevant that consumers previously estimated his talents as a vintner to be considerable, were willing to purchase the product, and initially deemed it of high quality. It would be implausible to insist that, while the consumers' preferences changed after they acquired new information about the wine and they no longer desired it, the producer is entitled to the proceeds from what he was able to market before they acquired their new preferences.

Accordingly, if the error rate in detecting actual merit (in the simplest case, ranking two candidates in order) is, say, 50 per cent, it may be *true* that the more meritorious deserve more of the components of well-being than the less meritorious, but the corresponding distributive *principle* will nevertheless be unusable. For, if the error rate is this high, then, no matter what anyone's intentions and efforts are, we cannot ensure that merit *is* rewarded or make our world a W_1 world. In this case, the insistence that the principle is fundamental serves only to promote a false sense of rectitude.

But surely we take great care to ensure that our selection procedures are just and that the tests and competitions we use to assess relative merit really do so? While we might sometimes fail to get the ranking right, it seems incoherent to suppose that our merit-assaying methods are irremediably no better than chance. One might be sceptical that the winner of a tennis tournament on some given occasion is really the best player, but the sceptical statement that, generally speaking, tennis tournaments fail to result in rankings of tennis-playing excellence that are better than chance seems not just

false, but impossible. One can reject facile statements such as ‘Tennis-playing excellence is whatever tennis tournaments discriminate’ while still insisting that the sceptical statement could not be true under any conditions. To vary a familiar line of argument, if a sequence of presented actions and evaluation decisions in some possible world does not do a reliable job of ranking tennis players, it is not a ‘tennis tournament’. This confidence is however unjustified. There are at least four difficulties with the deployment of a merit principle in real-world contexts:

- (a) Sweep limitations
- (b) Misattribution
- (c) Custom and convention
- (d) Mimicry

(a) *Sweep limitations*

To qualify as a bona fide tennis tournament, a sorting procedure need only do a reliable job of ranking the tennis players who entered the competition. So it is easy to imagine, not only a world in which tennis tournaments fail to result in rankings of tennis players that are better than chance, but also a world in which the rankings are considerably worse than chance. All that is required for the condition to be realized is that the vast majority of excellent players do not take part in the tournaments. This can happen for various reasons: they do not know that a tournament is occurring, they do not have the wherewithal to get to the tournament, or they lack information about their own level of ability. It might be objected that showing up for the tournament is a component of tennis-playing merit; the lazy player who stays in bed or misses the bus to the match is less good as a tennis player, but this claim has little credibility outside some very specific circumstances.

In any selection procedure, we do not want to know who is the best in the world for the position being filled. We want to know who can be plausibly designated as the best, given our unwillingness to spend more than a certain amount of time and trouble collecting and evaluating candidates. There are potentially more of the meritorious in many categories than there are slots to fit them into. Every high

school has persons just as good at what they do as Bob Dylan and Anaïs Nin, but, unfortunately for them, the need for persons in these roles is somewhat limited.

(b) *Misattribution*

Merit is customarily ascribed to individuals in a way at odds with the real determinants of success. The nomination and cultivation of talent and the incitement to effort are optional cultural work in which the subject, his helpers, and his competitors all participate. Her mother loves music or hates it, and the prodigy's talent waxes or wanes accordingly, though not necessarily just in the way one might expect. Talent and effort attract notice only when they are coordinated, and their coordination depends on external opportunities; the talent has usually to be noticed by someone other than the talent-possessor in order for the possessor to receive instruction and the means to develop it. So, although we tend to regard the thanks given by the successful to parents, spouses, managers, editors, etc. as pro forma and conventional, in fact a given display of coordinated talent and effort is attributable only by convention to the one deemed deserving of the reward. When merit is not displayed, it is likely that circumstances and other persons did not cooperate. Thus, the same 'intrinsic' quantity of coordinated talent and effort will lead to different outcomes depending on circumstances, and if the rewarders do not adjust their judgements accordingly, their rewards are, by the merit-desert principle, unjust.

Further, as social scientists have shown repeatedly, humans consistently overestimate their ability to judge performance and promise objectively.³⁰ In attempting to determine who contributed most to a joint task (and, correspondingly, who failed to pull their weight) and should be rewarded for it, observers are biased in favour of the contributor they most like and admire for extraneous reasons and biased against the contributor they least like and admire for extraneous reasons.³¹ Beliefs regarding performance are adjusted in order to

³⁰ On evaluation biases, see Richard Nisbett and Lee Ross, *Human Inference: Strategies and Shortcomings of Social Judgement* and Virginia Valian, *Why So Slow?*, 125–44.

³¹ Lerner, *Belief in a Just World*, 34.

satisfy the preference that good things happen to people judged to be good overall.

(c) *Custom and convention*

It seems obvious to us that some people are better than others at some things because they have an aptitude for them and apply themselves to them, and also obvious that it is better in our world to have an aptitude for some things (e.g., plasma physics) than others (e.g., weaving lanyards). The more meritorious are those who are better at the things it is better to be better at. Yet we recognize that there is a strong element of conventionality or historicity in the determination of what is a valuable production. It would be unsatisfactory to define a talent as any competency that can generate substantial monetary rewards since the sense of ‘can’ is unclear. Any well-honed practice could, in some possible world, generate rewards, but some talents, such as skill at direction-finding in the desert, or knowing the names of thousands of plants and animals, are such that, in our world, they cannot attract large monetary rewards. And contra Hayek, meritorious performances, according to consensus notions of reality, are not concerned with the provision of the most utilitarian necessities; the butcher and baker do not enjoy the most lucrative positions in our society. Those who gratify our non-utilitarian tastes are best rewarded, notably athletes, entertainers, popular authors, and theorists of the arcane and other-worldly. The meritorious name of God is more frequently mentioned in connection with astrophysics than in connection with plumbing or needlework.

(d) *Mimicry*

In our world, rewards are given on the basis of the principle of comparative advantage, applied by a judge who has some discretion over the size of the reward. Consider a simple employer–employee relationship. The employee’s earnings reflect what the employer believes or perceives the worker’s marginal contribution to the enterprise to be. We might suppose that comparative advantage at least approximates real merit, and that employers are reliable judges of comparative advantage. The selection procedures of the reward-

givers are aimed at producing reality-reflecting rankings. Problems requiring insight are asked on standardized tests and interviews probe for weaknesses and deficiencies in a candidate's knowledge. The objects being judged in such situations are not inert, however, like tomatoes being evaluated at an agricultural fair. The potential employee is a strategizing human being, whose aim it is to achieve the rewards of being perceived as meritorious whether or not he is so. The candidates in a meritocracy are each participants in an arms race against their reward-givers. The competence to *mimic* merit—to display the surface features of merit that will facilitate one's selection—may be obtainable more cheaply and quickly than actual merit, frustrating the judge's best efforts and forcing them to devise newer and more subtle tests.³² In such strategic situations, there is no reason to believe a priori that reward-givers have a faculty for detecting mimicry that is superior to the candidates' (and their allies') faculty for mimicking, just as there is no reason to suppose a priori that the predator's ability to detect tasty butterflies that merely look poisonous should outstrip the prey's ability to engage in successful deception. Successful mimicry—the ability falsely to appear more competent, experienced, creative, and promising—than the competition must be pervasive in meritocracies, for the greater the rewards offered for merit, the greater the incentive to deception. To make the distribution of the components of well-being highly dependent on attributes that not only can be simulated but that we have excellent reason to believe are often simulated is a policy that cannot be supposed to ensure just outcomes.

As Hume remarked, 'So great is the uncertainty of merit, both from its natural obscurity, and from the self-conceit of each individual, that no determinate rule of conduct could ever follow from it.'³³ Nevertheless, it is just as much a distortion to maintain that no activities and products are really meritorious as it is to maintain that whatever is deemed so or is socially rewarded is so. Rawls asserts that 'there is no set of agreed ends by reference to which the potential social contributions of an individual could be assessed'.³⁴ This seems

³² Examples of simulation include 'indicator chasing' and the results of 'professionalism seminars' and 'grooming'.

³³ Quoted by Hayek, *Law, Legislation and Liberty*, 62.

³⁴ Rawls, *Political Liberalism*, 276.

correct, but the observation that we do not have perfect unanimity and that we do have substantial disagreement does not exclude overlapping consensus on one hand and the meritocratic pursuit of discrete sets of values on the other. *Sub specie aeternitatis*, merit appears to be painted or projected onto persons and things in a way that is intrinsically relational and underdetermined by their intrinsic properties. The same can be said, though, of all valuation terms; the atomic constitution of two pictures does not determine which is beautiful. A reasonable person will accept that there are more and less meritorious enterprises and productions, recognizing simply that our ability to make reliable judgements—judgements we would continue to endorse even after severe critical analysis of our methods and assumptions—is much worse than we normally assume it to be and that we are subject to both random and systematic error.

7.4. Objective Deprivation and Thresholds

According to the results of the last three sections, merit exists, though our capacity to detect it reliably is defective. Further, we have some reasons, both instrumental and non-instrumental, to endorse some version of a merit principle, as well as some instrumental and non-instrumental reasons to ignore merit in making distributions. It is now time to return to the original dilemma.

As noted in Section 7.1, merit priority is universally agreed to be false under some interpretations of the term ‘need.’ Despite occasional claims to the effect that it is illogical or irrational not to schedule rewards on the basis of merit,³⁵ no credible political theorist subscribes to merit priority under every possible construal of need. Political theorists agree that basic needs, the requisites for biological survival such as fresh drinking water, medical care, shelter, and adequate caloric intake ought to be met universally, whatever else is done with the world’s resources. The need principle is not, in fact, deployed even under this basic interpretation, since funds that could, in principle, be used to meet needs are used instead to reward merit.

³⁵ Nozick appears to find a logical connection between merit and reward, assuming the premiss (which he takes as self-evident) that ‘people are entitled to their natural assets’. See *Anarchy, State and Utopia*, 225–6.

Yet commitment to the principle is not lacking in the sense that no credible theorist has posited that there is no *prima facie* right to survival on the part of an already-born person, or any duty to assist, and that the world as we observe it to be is just. This is tantamount to agreeing that we ought to follow a protocol of need priority at least with respect to basic needs. Consequently, even if the use of some form of merit principle is acknowledged to be just—and it has not yet been established that this is so—only a combined protocol such as the following with need priority at its foundation can be considered just.³⁶

Combined Proposal

In any condition *R* of the world that generates a surplus, resources will be directed first to the meeting of needs instantiating *U*. The surplus, above and beyond what is required to do so, is to be distributed according to perceived merit that is determined in competitions that are as controlled, and free of bias as it is possible at any time to make them, thereby instantiating *M*. If no surplus is available, *U* should be instantiated.

The interesting controversy, accordingly, concerns the proper interpretation of the concept of ‘need’ as it appears in the statements of merit priority and need priority. For we do face a choice (or, more strictly, an array of choices) between adopting a minimal and adopting a maximal conception of well-being. For any given level of resources *R*, which protocol is preferable?

- (a) Define the concept of ‘need’ strictly. Satisfy needs universally and then distribute the relatively large amount that may be left on the basis of merit, with some fine-tuning to reduce undesirable incentive effects.
- (b) Define the concept of ‘need’ generously. Satisfy needs universally and then distribute the relatively small amount that may be left on the basis of merit, with some fine-tuning to increase desirable incentive effects.

³⁶ As is stated e.g. by Roemer, ‘None shall consume luxuries while deprivation for others continues to exist’, *Theories of Distributive Justice*, 202; and in David Braybrooke’s *Principle of Precedence*, which states that some persons can be required to give up goods they do not need, though not goods they do need, to satisfy others’ needs, ‘Concept of Needs’, 61. The question what to do when radical scarcity does not obtain is not, however, addressed by them.

The strict conception of need refers to the biological requirements of survival—or what are sometimes termed ‘first-tier’ goods—calories, water, clothes, shelter, and treatment for acute infections and chronic debilitating conditions. It is unclear, however, why one would recognize a natural entitlement to first-tier, but only to first-tier, goods, unless one believed that so long as people are provided with first-tier goods they are almost always able to construct a decent life for themselves. Assurance on this score is lacking: many humans endure just such a marginal state of existence, scarcely preferable to not being alive at all. As observed in Chapter 6, a decent human life requires a ‘second tier’ of goods: some variety and pleasure in food, drink, furnishings or ornaments, the opportunity to engage in meaningful work and to advance one’s knowledge and understanding, along with opportunities for affiliation, mobility, some choice of mates, and freedom from harassment and derogation. The generous conception of need, then, incorporates second-tier as well as first-tier goods.

A fundamentally unjust world is one in which it is the case both that objective deprivation—the failure to possess the known and presumptive requisites of a decent human life—is widespread and that successful merit-mimics, or only a small subgroup of the truly meritorious, enjoy access to the doubtful and speculative components of well-being. The image of the morally corrupt state embodies both features. And this suggests an approach to the solution of the original choice problem.

Having already accepted need priority on the basis of the obvious unacceptability of merit priority, we should choose the generous conception of need for two reasons. First, the strict conception, according to which only basic biological needs have an automatic right to satisfaction, is arbitrary since it is too weak to support the notion of a decent human life. Second, our ability to detect real as opposed to apparent merit and to reward the entire human population for merit is so constrained that it is appropriate to limit the amount available for merit awards.

Accordingly, in a just world containing more than adequate resources to meet them, the satisfaction of known and presumptive needs will be independent from meritorious performance. Merit may nevertheless permissibly entail the award of additional goods and states that persons believe will benefit them, whether or not the

possession of these things will necessarily render them objectively better off, effectively matching the dubious aspects of merit determination with the dubious aspects of the pursuit of the good life. Perceived merit can entitle us to rewards in the form of access to the speculative and doubtful, though not the known and presumptive components of well-being. For if we can know that it is just for differential perceived merit to entail *some* differential access to the components of well-being, and if there is no rationale for allowing it to entail differential access to the known or presumptive components of well-being, the only remaining possibility is that it is just for it to entail differential access to some other category of the components of well-being, namely the doubtful and speculative.

The argument for rewarding merit cannot rest on either intrinsic or instrumental defences of merit principles; these have been shown to be too weak to support even a combined protocol. It can only rest on a preference, based on the recognition that there is some degree of virtue in those inconclusive defences, for merit principles over rival distribution principles. For, in a condition in which we have more than is required to meet first- and second-tier needs universally, the five alternatives available are the following:

- (1) Distribute the surplus at random.
- (2) Distribute the surplus to historically privileged groups and individuals.
- (3) Distribute the surplus equally.
- (4) Distribute the surplus according to a Rawlsian ranking in which the relatively worst-off are first, and the relatively best-off are last.
- (5) Distribute the surplus in whatever way maximizes future surpluses.

The use of a merit principle seems preferable to each of 1–5, once needs have been met. It is preferable to the other proposals since it makes some concession, rather than none, to the force of the merit-desert intuition and to the observation that the introduction of meritocratic principles can sometimes improve overall well-being.

Egalitarians who favour protocol 3 maintain that the whole remainder of *R* ought to be divided equally, once needs are satisfied. But it is unclear why, if the satisfaction of a need can be set at a generous level,

the purely formal preference for equality ought entirely to override a merit principle. While the Rawlsian protocol in 4 is appealing in a condition of scarcity, it is unclear why, once needs are universally met, the relatively worst-off, who are no longer in a condition of objective deprivation, should continue to be favoured. The protocol would permit a small number of the least hard-working and competent to enjoy the most subtle and complex pleasures of civilization if it turned out that the relatively worst-off person (who might already have an entirely decent life) was thereby made better off.

Protocols 1 and 2 might be permitted to operate in a marginal way in a just world to gratify needs other than the need for social justice—lotteries for the sake of entertainment, tax exemptions for the monarch for the sake of feudal deference—but they cannot be considered as principles of social justice. And the protocol in 5 merely postpones and so evades the question.

Once needs are generously met, departures from equality require no special triggering condition. Provided the known and presumptive components of well-being are provided to all, there can be no objection in the name of justice to allowing those deemed meritorious (a few misidentified and mimics among them) their racehorses, diamond tiaras, fancy chronometers, 200-foot yachts, and other such items, as well as opportunities to use expensive cyclotrons, collect rare books, or construct artificial paradises of their own making, provided we are doing as well as we can with respect to correcting the known deficiencies of our assay methods. The merit-desert principle can be sustained with this amendment.

Several objections can still be made to the combined proposal as interpreted and justified above:

- (1) The protocol is exigent. It requires that, in a general 'leveling down', the meritorious accept significant reductions to meet the overwhelming needs of the non-meritorious. There will be little or no surplus to distribute on the basis of merit if needs, generously defined, are met.
- (2) The protocol is self-defeating. Only by refusing to meet the known and presumptive needs of the comparatively non-meritorious and by withholding from them the requisites of a decent life, can we stimulate them to develop meritorious

behaviour that will allow them possession of the known and presumptive components of well-being by increasing the total good available.

- (3) The proposal is inconsistent. Access to the doubtful and speculative components of well-being consistent with that of one's social reference group is one of the known and necessary components of well-being.

Objection 1 is commonly heard, but not very compelling. Suppose that the level at which it is possible to enjoy a good human life, is set, generously, at 10 per cent of the world's combined GNPs. My rough-and-ready calculations in Chapter 5 suggested that thirty million top income earners (who are not of course the most meritorious persons) could bring about this result by redistributing less than 10 per cent of their salaries. This is hardly 'levelling down'. It reflects a curious conception of the distribution of merit; namely, the idea that there is a tiny meritorious elite suspended above a huge utility sink of persons who are a net drain on the productive capacities of society.

Objection 2 is relevant in a condition in which potential resources are enough to meet needs but in which the condition of the world is such that they are very hard to get 'out of the ground' and the motivation to work is low. The hypothesis advanced is that, in all worlds in which non-adjustable social, psychological, and material conditions are similar to ours, a world whose regime withholds first- and second-tier goods in order to reward those at the top of the perceived merit scale better with third-tier goods will contain *fewer* persons whose first- and second-tier needs are not met. There might be unknown 'laws of nature' in virtue of which this hypothesis turned out to be true, but it cannot be seen to be true on mere inspection. It is rendered doubtful by data indicating that societies with large income variance also have a good deal of objective deprivation.³⁷ If, contrary to expectations, the hypothesis

³⁷ There is historical and current evidence that rapid growth that increases the Gini coefficient of inequality induces not only relative but, for the worst-off sectors, absolute deprivation. 'Growing numbers of people after 1780 could not afford food', according to Timothy M. Smeeding and Peter Gottschalk, 'Cross-National Income Inequality: How Great Is It and What Can We Learn from It?', 5. The USA has the largest per capita income in the world; inhabitants in the lowest 10th decile of income distribution are nevertheless 'at severe risk of poor health, subsequent poor education performance and diminished achievement'. *Ibid.* 18. The USA also has the greatest real income inequality amongst the seventeen OECD countries studied.

were true, it would be more reasonable to take it as implying that the realization of a just world is hampered by an adverse environmental condition, than to take it as implying that justice requires punishing the deprived.

Objection 3 confuses the task of designing a distribution protocol with the more complex task of designing a redistribution protocol for a world in which persons have pasts and memories and expectations based in their pasts. Consideration here is restricted to a set of individuals who are not yet fixed in social circles and classes and who have no memories or expectations. No one can suffer a psychologically distressing 'fall' out of her former socio-economic class or claim to be made objectively badly off by suffering depression or demoralization on that account. The design of a just redistribution protocol is an important but distinct task that has to be guided by a decision as to which distribution protocol is most just.

7.5. Statistical Equality of Outcomes Required

A further objection to the combined protocol is that it is too permissive of invidious discrimination. Under the combined protocol, everyone undergoes risk, the risk of being born with too little by way of native endowments to have access to the speculative and doubtful components of well-being. Individuals undergo further risk on account of the gap between real and perceived merit, for their merits may not be appreciated. Now, life is risky in many ways and a just world does not need to exclude all forms of risk, but only unjust risk. So we might wonder what is required for the risk of diminished access to the speculative and doubtful to be fairly distributed.

Recall that our ordinary conception of merit floats between a conception of merit as a special expertise that happens, contingently, to command respect and to attract rewards, and as an objectively valuable attribute. It is clear that some groups—characterized by language, nationality, gender, 'race', or other features—possess more of the special expertise that commands respect and attracts rewards: there are more Japanese physicists than Congolese. But is it conceivable that some groups have more merit, considered now

as an objectively valuable attribute, than others? Suppose persons possessing a certain rare blood type have discovered a set of obscure tasks at which they excel and which others with more common blood types can barely perform. Under what conditions can the rare blood group claim to possess more by way of humanly valuable qualities than other people? Unless the majority values equally performance on those tasks they cannot for some reason do, it is difficult to see why the members of the rare blood group should be considered objectively meritorious. Of course we can attempt to fix the reference in this world of 'objectively meritorious attributes' and then project a world in which the majority, for some reason, lacks them and does not care about them and a small elite cares about them and possesses them, but this does not undermine the point. Though not everyone is personally musical, musicality can be regarded as an objectively meritorious attribute only to the extent that competence with respect to and interest in music is well distributed through all human populations. It follows that a quality like 'intelligence' can only be considered objectively valuable if intelligence is well distributed and consistently prized throughout human populations. If this is not the case, greater intelligence does not entitle anyone to a better life; if, however, intelligence is distributed and valued across human populations, Nagel's suggestion that a country that rewards the intelligent better than the unintelligent is more unjust than a country that practises sexual or racial discrimination seems unconvincing.³⁸ On this view, valuing white skin over dark or masculinity over femininity is no worse than valuing intelligence over dull-wittedness.

Suppose, in a world rich in resources, the combined protocol is the basis of distributions. We can envision a condition of the world—call it non-correlation—where membership in a particular group—ethnic, linguistic, age-bracketed, etc.—carries no predictive power with respect to outcomes where the possession of all three tiers of goods is concerned. Under non-correlation, some persons of true worthiness will be edged out by those simulating merit, some will receive referred credit for the competency and effort exercised by others, some will benefit from historically conditioned definitions of merit that have little rational basis, and others will just fail to be in the

³⁸ Nagel, *Mortal Questions*, 99. Nagel denies that statistical equality as between racial groups and genders with respect to outcomes is a requirement of justice, *Equality and Partiality*, 89–90.

right place at the right time, missing the whole competition. At the same time, membership in any group will not increase or reduce the risk of failing to obtain possession of the components of well-being. This world can be considered just, with the error rate of its meritocratic procedures approximately matching the error rate of individuals' judgements of what actually improves their lives.

We can also envision a condition of the world in which degree of risk is strongly correlated with properties such as ethnicity, gender, parental income, place of residence, and so on. This world is inevitably unjust. For, if we are considering merit as an objective property, statistical inequalities of outcome among different groups can only arise through discrimination, i.e., failure of merit assays or equal social investment to develop talent, with respect to particular groups. If, on the other hand, we are considering merit as special expertise contingently possessed by some groups of humans, inequality of outcomes does not imply discrimination, but it does imply an arbitrary conception of human excellence that ought to confer no particular entitlements. This is not to say that every instance of statistical inequality of outcomes in worlds like ours is morally intolerable. It is not morally intolerable that African-American basketball players are more respected and rewarded than Japanese basketball players. In a just world, however, statistical inequalities ought to balance out from group to group.

To be sure, any individual's membership in a reference group can be construed as a matter of luck. It might be argued that the distinction between the correlated world and the non-correlated world is accordingly illusory. Both are equally just, for the risk that any individual assumes, prior to his chance assignment to one or another category of person, is equal. Those who take merit priority seriously are perhaps disposed to regard luck as a force that can be ignored on the grounds that it does not discriminate. On this view, while the world would be in some sense better if luck were incapable of rendering the fates of two persons of equal merit different, the existence of good and bad fortune does not affect the choice of a distribution scheme, since everyone is equally subject to luck and since we cannot by definition control it. Even when luck makes the outcomes of two persons different by casting them into two different reference classes, we can and must ignore it. And it might be argued

further that, just as no individual can be made objectively badly off by suffering a decline in the availability of one of the speculative and doubtful components of well-being relative to others, provided he continues to enjoy the known and necessary components of well-being, no group can be made so. But the recognition that luck is ineradicable and that some percentage of both good and bad fortune is undeserved is itself a reason for insisting on some degree of compression in outcomes.³⁹ It does not follow from the truism that no one is invulnerable to luck that everyone is equally vulnerable to chance events; the well-off are often insulated by their reserves. If it is right to compensate for undeserved misfortune, it may also be right to subtract from undeserved reward. And this point applies not only to the luck of events but, as Rawls argued, to the luck of situation and endowment. Justice does not consist merely in the assumption of equal risk in an unassigned position, and the argument that what is not necessarily wrong for an individual cannot be wrong for a group fails for related reasons.

Consider a group of free-living Greens in a condition in which known and presumptive needs are universally satisfied. The Greens pursue the speculative and doubtful components of well-being in competitive encounters, along with a group of Reds, with equal, though limited, success for members of both groups. Some environmental change occurs, inducing mild scarcity, and the speculative and doubtful components of well-being as well as the known and presumptive components grow less accessible. Under a new regime, the Greens are persuaded, or persuade one another, or are recruited to farm the land for the Reds so that the known and presumptive components can be supplied to all. The Greens cease to take part in meritocratic competitions, or their competencies come to be viewed, conventionally, as non-meritorious, or they lose the ability to signal merit when faced with competition from merit simulators. Consequently, the Greens renounce, or are required to give up, or simply lose the inclination for the further pursuit of the speculative and doubtful components of well-being, though all their lower-tier

³⁹ The view that radical contingency ought to lead to compression in *judgements*, i.e. to non-judgementalism, was advanced (though not defended) by Nagel in 'Moral Luck', in *Mortal Questions*, 24–38. This is obviously a different conception of moral luck from Williams's; one would not for example expect Williams to favour compression of outcomes.

needs continue to be met, and this to the same high standard as before.

Here, the mere fact of a discrepancy between the condition of the Greens and that of the Reds seems to render the situation unjust, not just in view of the losses of the Greens of their former status, of which they may retain some memory. The burden of meeting known and presumptive needs ought, it seems, to require equal relinquishment of the pursuit of the speculative and doubtful components of well-being by the Greens and the Reds.

Such a scenario would actually be realized in a society that denied access to the speculative and doubtful components of well-being to members of certain identifiable categories by running tournaments to which they were unable to come, by misattributing their successes to others, by deeming their productions non-meritorious, and by accepting fewer mimics from their category than from its complement through the exercise of greater vigilance. For such a society might do a better job of meeting known and presumptive needs universally than a rival system by selectively diverting labour towards agricultural production, childcare, and domestic maintenance. This society would be unjust, not because some persons would enjoy third-tier goods that others did not—for this result is tolerable when it happens ‘by chance’—but because some groups would enjoy an advantage in their pursuit.

Even in the event that the known and necessary components of well-being, generously conceived, can all be furnished for less than the cost equivalent to entire social product, the proposal that merit can justly influence distributions of the remainder is still flawed. For a merit principle to have a valid role in distributive justice, even as a complementary principle, human capacities to define, attribute, discriminate, and determine merit must be rendered sufficiently reliable that the threat of the exercise of privilege by some groups and the assumption of excessive risk by others is less than it is under alternative systems in which the realization of *M* is not an explicit aim.

Consider the position of the Count Raymond de la Cisternas in Matthew Lewis’s novel *The Monk*. The young Raymond is setting off in disguise as a ‘private gentleman’ and receives some words of advice from his father:

Tis true that as the Conde de las Cisternas you would have been received with open arms: and your youthful vanity might have felt gratified by the attentions showered on you from all sides. At present much will depend upon yourself. . . . You must lay yourself out to please. You must labour to gain the approbation of those, to whom you are presented: They who would have courted the friendship of the Conde de las Cisternas, will have no interest in finding out the merits, or bearing patiently with the faults of Alphonso d'Alvarada.⁴⁰

A person seeking to make his way in the world is judged by others—by members of the public, by authorities, and by official assessors of merit. Dark-skinned people, small or short people, those with certain foreign accents, strange mannerisms, or high voices, are, like the disguised Raymond, dressed as beings of lesser capability in our world, and they must endeavour, despite their unprepossessing garments, to win the approbation of those in a position to help them advance in life. Few persons vote against, fail to select, dismiss, or ignore someone because he or she is a member of a group they do not, for irrational reasons, like. It happens nevertheless that when merit is assessed, some categories of persons do not fare well. It is not necessary for them to fail that someone in a position of authority is against them; it is sufficient that there are excellent reasons for the selector to be in favour of someone else. The noblemen Raymond meets need not discriminate against him if they fail to assist him on his way. But they may have no interest, as his father warns him, in determining his merits and no reason to overlook his faults. If he tries to impress his capabilities on them, he may, in his humble clothing, appear merely froward, and if he gives way to frustration or loses his temper when his underlying noble qualities pass unnoticed, he may be judged even more unsuitable.

It is easy to say 'Not once have I harmed someone because I thought their race, or sex or nationality inferior, and if others of my sort have done so in the past it is not for me to pay their debts.' But one might interrogate oneself as follows: Amongst those on behalf of whom I exerted myself, how many were not of the dominant classes? Amongst those whose faults I bore with patiently and whose merits I troubled to find out, how many of them were?

⁴⁰ Matthew Lewis, *The Monk*, 95–6.

Moral Equality and ‘Natural’ Subordination

‘When [a man] is in a cooperative and benevolent relation with a woman, his theme is the principle of abstract equality; the concrete inequality to which he can otherwise attest is not posited.’¹ Simone de Beauvoir’s remark has not lost its pertinence with regard to modern moral discourse. General moral theory discusses the obligations and entitlements of persons without regard to sex or gender. This discursive posture may express the presumption that between the ordinary run of philosophers and women there exists a cooperative and benevolent relationship. Or it can be interpreted as evasion.

Aristotle, Rousseau, and Kant, amongst others, took note of the division of labour and modes of life between the sexes. They considered the status of men and women relative to one another to pose interesting questions and believed themselves responsible for explaining the special obligations of women and the special privileges of men. Where the old philosophers explained women’s subordination as best they could—and it would not have occurred to them to do so by reference to women’s preferences or choices, as opposed to nature’s plan for them—modern theorists of distributive justice are disinclined to take either an analytical or an evaluative position on the character of women’s lives.² Whether it is motivated by unease, uncertainty, or indifference, or entirely innocent, this is an oversight. The old philosophers were right to appreciate that men and women stood in a curious relationship to one another where the distribution of the components of well-being was concerned. We should aim to

¹ Simone de Beauvoir, *Le Deuxième Sexe*, i. 27–8.

² See the criticisms of Susan Moller Okin, *Justice, Gender, and the Family*, 9, also Richard Rorty’s remarks in his review of Annette Baier’s *Moral Prejudice*, ‘Why can’t a man be more like a woman?’, 3.

recover their interest in explaining and evaluating this noteworthy feature of our world and its relationship to our underlying natures. We need not reproduce their descriptive errors or their unacceptable prescriptions.

8.1. Male–Female Relations in Moral Philosophy

It might seem that the reassignment of the topic of women's social inequality to sociology and gender studies has a benign explanation. Moral and political theory are concerned with relations between rational, reflective agents, not between persons occupying certain social roles. Women, one might think, are included as moral subjects in all reasonings concerning justice and goodness and have precisely the same obligations and entitlements as men.

Yet the claim that gender relations are not a subtopic of distributive justice because women are full moral and political agents is unsatisfactory. The fact that women are rational and reflective and are now widely recognized as such (assuming this was ever seriously doubted) does not entail that existing relations are just, or lie outside the purview of philosophy proper. We do not regard the rich or the poor, the gifted or the ordinary, as irrational and unreflective. We are certain that they have the same rights and obligations as 'ourselves'. Yet we are willing to speak of their interests as a class, and to ask specific questions of the form: What should the wealthy relinquish to ensure a given level of well-being in the poor? What limits to their aspirations must the less talented accept in a spirit of realism?

It might be ventured that, while we know that the interests of the rich and the poor are in conflict, we know the interests of men and women to be in harmony. But empiricists will be curious as to how this harmony of interests has been established. Are not males and females genotypically and phenotypically distinct subclasses that constitute necessary and exploitable resources for one another, which they cannot ignore? The existence of another sex that it can recognize by means of sensory cues is a feature of every dimorphic animal's environment, a potential resource that its genes are disposed to employ to further their own advantage. Literature and the media of popular culture bring into relief the combination of fascination and

indifference, dependency and antagonism that characterizes social relations between the sexes. Few of us enjoy, in this realm, the substantial psychological reserves that make an easy Humean grace a component of our demeanour at all times.

Where subhumans are concerned, we cannot make too much of how the two sexes treat each other, however solicitous of the welfare of individual animals we might be. Each will hold its own, and if one sex has a shorter lifespan, or eats less or receives more wounds than the other, this is not a matter of moral concern anymore than it is if female praying mantises eat male praying mantises. It might seem that we should take the same hard-headed attitude towards the statistical inequalities of modern societies. If an inequality is not caused in large measure by the actions of, and produces no advantages for, a favoured class that they ought to forgo, it isn't a subject for the theory of justice. Being born without an arm is usually, though not always, an unfortunate condition. Yet others do not benefit from the condition and one cannot claim that it is unjust.

Being born a female might be considered just such a natural misfortune. Though it is not as grievous as being born without an arm, being born in this condition reduces one's chances of becoming a significant and personally recognized contributor to civilized endeavours. At the same time, it exposes one to excess risk of beatings, insolvency, and social disgrace. This can be considered mere bad moral luck. Everyone, a philosopher might muse, undergoes the same risk of being born male or female. There is no reason to think that men in enlightened countries take deliberate steps to worsen the condition of women relative to what it would be otherwise in order to benefit themselves or that they tolerate culpable letting-happen knowing that it makes their lives more pleasant than they would otherwise be.

It is nevertheless far from clear that men as a group do not benefit unintentionally from actions and policies that reduce opportunities for female participation and that render women's lives more dangerous in some respects than men's are. While the effects of natural disasters, like the downstream effects of transactions not under the control of single agents, do not incriminate those who ignore them of profiteering, they can incriminate these persons of moral neglect. If gender inequality involves real deprivations, then, whether or not it

is, in some respects, the result of a natural misfortune, we have some obligation to remedy the deprivations at some cost to ourselves, provided the sacrifices are not, impersonally considered, overwhelming and considerable good is to be obtained.

This chapter is dedicated to the following three questions. First, what reason is there to think that statistical inequality between men and women is associated with the kind of inequality that is morally worrisome; namely, the kind that involves remediable objective deprivations? Second, what kind of misfortune is gender inequality? Is it best viewed as the result of a natural disaster, or as the unfortunate summary result of a series of individually unproblematic negotiations? Or, contrary to what was suggested above, does it involve advantage-taking on the part of a privileged class? Third, what are the costs of greater equality and should we be prepared to pay them?

8.2. Are Women Objectively Deprived?

The belief that the division of the cooperative surplus between men and women of the same social class, if not between nations or classes, is more or less just, and that women's lives are overall exactly as good as men's, is held in large measure by both sexes. As Beauvoir notes, in face-to-face encounters, the impression that equality prevails may be irresistible.³ In modern families, questions such as the division of labour, the relative importance of one spouse's career compared with the other's, the investment of resources in further education, and other issues that have a bearing on individuals' social standing and their prospects are discussed openly and often reasonably, allowing for their emotional character, and agreement is ordinarily secured. Legal rights with respect to ownership, liability, and testimony are symmetrical between men and women in most developed countries.

At the same time, objective evidence for women's enjoyment of lower levels of well-being relative to men of their own reference group is not hard to come by. According to most anthropologists, women in nearly all existing and past cultures of record have lower status than men do in two respects: They are excluded by custom and tradition

³ Simone de Beauvoir, *Le Deuxième Sexe*, i. 27.

from prestigious occupations, employments, and activities; and they perform most basic maintenance activities such as feeding, nursing, and cleaning. Men, whether they are unemployed, employed in unskilled positions, or occupy lucrative and visible ones, tend to assume a partial or full exemption from maintenance. E. O. Wilson notes that 'History records not a single culture in which women have controlled the political and economic lives of men. . . . Men have traditionally assumed the positions of chieftains, shamans, judges and warriors. Their modern technocratic counterparts rule the industrial states and head the corporations and churches.'⁴ These conditions justify us in referring to women as existing in a condition of subordination.

Educational and employment prospects, legal protection, and access to credit have improved for women in many parts of the world, but women's lower standing is pronounced even in the wealthiest and most rights-conscious nations. The following observations pertain to Western liberal democracies *c.*2000 CE.

- (1) Executive positions such as judge, professor, general, director, president, minister, doctor, mayor, and board member are overwhelmingly held by men.
- (2) Men initiate and carry through most significant financial transactions involving the investment and expenditure of public and corporate funds. They own approximately 90 per cent of the world's wealth.
- (3) Men constitute the greater portion of the clientele in fine restaurants, first-class sections of airplanes, and luxury hotels.
- (4) Men win more prizes, are awarded more badges and medals, and have more buildings, bridges, and roads named after them than women.
- (5) Men appear more frequently in non-entertainment magazines, and as subjects of news programmes. They are photographed more often performing tasks requiring specialized training and expertise.
- (6) Men come and go from the household with greater freedom than women. Recurrent or permanent defections from the household and the family by men are judged less severely than similar defections by women.

⁴ E. O. Wilson, *On Human Nature*, 128.

- (7) Men, with the exception of professional porters and caddies, carry fewer bags and sacks than women do, and push fewer prams, pushchairs, and shopping trolleys than women do.
- (8) Men come into contact with kitchen detritus, vomit, excrement, soiled clothing, and other taboo objects and substances less than women do.

Women are not numerous on the editorial boards of newspapers and magazines that collect and shape public opinion, or on the governing boards of research centres, universities, and regulatory agencies. Women, for the most part, do not decide where national boundaries are to be drawn, how cities are to be laid out, which drugs and surgical procedures to promote, how many airports or railways to build, what rights are to be enshrined in a constitution, and how many bombs and missiles of what kinds a country is to have and when it is to deploy them.

Women's lesser authority and status is not a function of their cognitive abilities. According to the textbook of Maccoby and Jacklin, '[B]eginning in early infancy, the two sexes show a remarkable degree of similarity in the basic intellectual processes of perception, learning, and memory.'⁵ There are subtle differences in colour vision, in the ability to recognize and remember faces and to ascertain the emotional states of others (female favouring), and in the ability to solve word problems and rotate imaginary figures (male favouring). Overall, however, women and men are not very different with respect to traits such as spatial ability, reasoning, divergent thinking, creativity, moral judgement, achievement striving, and task persistence.⁶ Though the feats of calculation, memory, and representation of the savant syndrome are more commonly exhibited by men by a factor of six to one, both sexes are liable to it. Only 5 per cent of the measurable variance in cognitive ability is attributable to sex, and women's lower maths scores are insufficient to explain their limited participation in science, mathematics, and engineering.⁷ The variance in other capabilities, such as eye-hand-coordination may be real and significant, but such differences can hardly explain women's lesser participation in politics, journalism, and the arts.

⁵ Eleanor Maccoby and Carol Jacklin, *The Psychology of Sex Differences*, 61.

⁶ *Ibid.* 75 ff.

⁷ Stephen Brush, 'Women in Science and Engineering', 412.

Nor is women's lower status explained by the debilitating effects of pregnancy and lactation. Some women seek to improve their personal socio-economic outcomes by making the difficult choice to delay or avoid childbearing. Not only is the decision medically questionable and a complication with respect to marriage, but childless women do not on the whole fare any better in professional hierarchies than women with children.⁸ Women perform more physical labour than men in many cultures, and their capacity for industry is taken for granted cross-culturally.⁹ Women with and without young children usually perform the agricultural work, exclusive of ploughing with draft animals, in traditional societies. Water carrying is almost universally a female activity, even though women's upper-body strength is 30 per cent less than men's. In industrial economies, women are well represented but poorly paid in factory work requiring manual dexterity, and in post-industrial economies, they are low-wage 'pink collar' clerks and secretaries. Women are compressed towards the low to middle end of the occupational spectrum: few are in jail, and few are at the tops of hierarchies.¹⁰

Women experience deficits with respect to each of the known and necessary components of well-being—consumption,¹¹ expression, affiliation, activity, participation, and respectful depiction. They are under-represented in the 'active world' of the collective imagination. A careful tally of listings in the local entertainment guide should convince any reader that modern cinema is overwhelmingly concerned with men, their thoughts and ideas, their conflicts and struggles. Even television and the comics, with their more domestic focus,

⁸ According to Virginia Valian, although academic women are less productive than academic men, women with children are no less productive and may even be more productive than women without. *Why So Slow?*, 269–70.

⁹ According to Esther Boesrup, in a Central African republic typical of many rural agrarian societies, 'women generally do the most exhausting and boring tasks, while the performance of the men is sometimes limited simply to being present in the fields to supervise the work of the women'. Anne Oakley, *Housewife*, 173 (cited from Boesrup, *Women's Role in Economic Development*). In Burundi, men and women both agreed that women were better suited for work than men. 'Everyone knows', a group of informants told Ethel Albert, 'that men are not suited *by nature* to heavy work. . . . Men drink too much and do not eat enough to keep up their strength; they are more tense and travel about too much to develop the habits or the muscles needed for sustained work on farms.' Oakley, *Housewife* 174 (cited from Albert, 'The Roles of Women').

¹⁰ Heather Joshi and Pieralla Paci, *Unequal Pay for Women and Men*, 18–19.

¹¹ Men spend more than women on personal items and intoxicants, particularly in poorer countries. See Dasgupta, 'Food, Care and Work: The Household as an Allocation Mechanism', ch. 11 of *An Inquiry into Well-Being and Deprivation*. Women shop more than men but make chiefly altruistic purchases.

represent the world as approximately two-thirds male and one-third female. Women are more frequently photographed in a state of undress. They are rarely depicted as absorbed in a task, oblivious to the gaze. These representational trends are more evident in the modern world under liberal democratic regimes than they were at earlier times when the power and charisma of hereditary aristocracies elevated some women to positions of prestige and even influence in the absence of meritocratic competition and social mobility. Finally, women are more liable to depression than men and lack social resiliency. Their personal reputations are more vulnerable; their position and status always more precariously maintained. It is estimated that eight times as many women as men are abandoned by their spouses when they develop a serious illness, and widowers are far more likely to remarry than widows are; the ends of women's longer lives are often spent in a state of physical frailty and loneliness. The financial consequences of the death or defection of a partner are more serious for women than for men.

In short, women are less likely at any stage of life to be found amongst the beneficiaries of the ancient system of exemption and privilege, enjoying the kind of life Aristotle understood as a good life for members of the moral community. Walzer, whose apparent tolerance for voluntary female subordination sits uneasily with it, observes that 'The real domination of women has less to do with their familial place than with their exclusion from all other places. They have been denied the freedom of the city, cut off from distributive processes and social goods outside the sphere of kinship and love.'¹² Whether or not Walzer overestimates women's access to the goods of kinship and love, he is certainly correct in his observation that they do not receive their fair share of social goods.

For the most part, women are the patient and willing workers of the world. They engage in activities that make lower cognitive, emotional, and aesthetic demands than specialized work, performing a diverse range of activities to an adequate standard without concentrating on one or two that bring social recognition and reward. While individual women may enter the specialized occupations and enjoy

¹² Walzer, *Spheres of Justice*, 240.

privileges and exemptions from ordinary work, they did not in ancient and do not in modern times, make up the proportion of governors and legislators, athletes, or investigators of other worlds—metaphysicians, priests, scientists, and explorers—that their numerosness and abilities would seem to predict. This is not to say that the lives of the best-off women are significantly worse than the lives of the best-off men, or that the lives of the worst-off women are significantly worse than the lives of the worst-off men. Nothing precludes the possibility of high degrees of well-being for women or utter misery for men. The initial generalization nevertheless holds despite the existence of numerous exceptions. The deprivations implied by women's lesser control over resources, political authority, and enjoyment of specialization opportunities in developed countries are trivial in comparison with the deprivations endured by women in countries in which they are denied basic legal rights and protection from physical abuse. However, in setting the anchor for the status of women at abjection, not at the other end of the spectrum of well-being, we unconsciously represent women as deserving less.

Faced with evidence that millions of interactions between individual parties, each of which may be in order from the microethical perspective, reliably sum to a situation of reduced female liberty, happiness, and authority, women and men have an array of conflict-resolving devices at their disposal to show why, despite the uncontroversial truth of the observations just cited, women are no worse off overall than men are. They may retail anecdotes of notable exceptions, as though these refuted the statistical evidence, stress the voluntary nature of the division of labour in male–female partnerships and the importance of personal choice, or draw attention to the personal satisfactions afforded human females by their service roles. They may present the merit-desert intuition, or advance the Aristotelian observation that what would be a deprivation for a man is not a deprivation for a woman. Statistical social equality, it might be said, is rapidly becoming a reality. Finally, it might be urged that, while women lack social status, they do not lack moral dignity, and only the latter is truly important. For, at least since Kant, most philosophers have insisted on a distinction between social prestige and intrinsic worth as a human being. Human dignity is not read off from a list of

public accomplishments. It is, as Kant maintained, posited, and it is assuredly posited in women as well as in men.

The suggestion that, as long as women have inner human dignity, their lacking wealth or direct legislative power, or even their liability to poverty and depression, is irrelevant to morality is unacceptable. The fact that it is possible for a morally dignified subject to live well in conditions of social deprivation cannot be the foundation of a theory of social justice. I can imagine a cow with the head of a goat. If I now go further and infer that there is no essential connection between being a cow and having a cow's head, I am on the wrong track. The ancient philosophers posited an intrinsic connection between living in a certain way, being regarded as excellent for living in that way, and being an excellent human being. Kant was offended by these elisions, which he sought to correct. The message he and subsequent Kantians have given out concerning the relationship between moral dignity and social dignity has been unclear as a consequence. And indeed it has to be. For the boundary between low socio-economic status and lesser moral dignity is not well defined. This is not to say that being-of-the-gender-that-rarely-has-bridges-named-after-one-of-them entails that you have less human dignity than being-of-the-gender-that-almost-always-has-bridges-named-after-one-of-them does, or that your dignity is injured or violated because there are few bridges named after persons of your sex. At the same time, if you are a member of a group, all of whose members it would be unthinkable or barely thinkable to name a bridge after, it is possible that others consider you as having less importance as persons. If you are a member of the sex that is more frequently aborted, or that dies more frequently of malnutrition, you are in a position to argue that your dignity has been injured—that something has been taken away—or that others are failing to acknowledge something in you. Between trivial and mostly unnoticed injuries to vanity and the serious violation of human rights we can interpolate cases indefinitely.

One who insists that statistical equality is on the way might well be asked for what future year his bets would reflect indifference to gender. Clearly we should prefer the bet that a randomly selected member of the category of, e.g., 'persons having signing authority in the year 2010 over sums of more than one million dollars', or 'persons

observed in the first-class section of an airplane in the year 2010' is a woman to the bet that a randomly selected member of the category 'persons having signing authority in the year 1980 over sums of more than one million dollars', or 'persons observed in the first-class section of an airplane in the year 1980', is a woman. No reasonable person, though, would be indifferent to the bet that a randomly selected member of the 2010 categories—and perhaps the 2050 categories—will be a man v. a women. Although the proportion of the wage gap in the full-time earnings of British men and women born in 1958 and later that can be explained by differences in experience and qualifications has diminished in past decades, the proportion attributable to discrimination has at least remained constant and may have risen.¹³

Perhaps the emotional rewards of femininity (of simply belonging to the childbearing sex) exactly balance the deprivations? On this view, there must be constituents of well-being with respect to which men are deprived to which women have access that even the score, leaving neither sex absolutely worse off. The deprivations allegedly endured by men that might be claimed to even the balance are several. Most men lack the opportunity to form a close bond with an infant and have relatively little control over household arrangements, and it is sometimes said that men are deprived of opportunities for forming intimate ties with same-sex age-mates. Yet history does not record any protests by literary and philosophical men with respect to their exclusion from places, activities, and experiences available to women. Literature and poetry are silent on the subject. Unless it can be shown that there are institutional obstacles that selectively restrict men's, but not women's, full employment of their intellectual, economic, and artistic abilities, the argument from compensating gratifications and frustrations fails.

Being born female is therefore a condition carrying excess risk with respect to the attainment of many of the components of well-being. In this respect it is like being born without an arm or being born into a poor family of goat-herders. These conditions do not preclude happiness and those who have the deficits may never really understand what they are missing. Being saddled with them is compatible

¹³ Joshi and Paci, *Unequal Pay*, 63–4.

with one's relatives and neighbours suffering their own set of unrelated deficits. Nevertheless, they impact on one's life chances.

Is this a moral problem? Everyone undergoes an equal risk of being born a woman, after all, and the role of a woman is far from being unacceptable. No reasonable person—unless conditions are worse than we think—would prefer non-existence to the role of a randomly chosen human female. It is not unjust, one might argue, but merely unfortunate, that, in a world in which A_1 and A_2 share all their independent properties except that A_2 is a paraplegic, outcomes are worse for A_2 . Nor is it unjust that in a world in which A_1 and A_2 share all their independent properties except that A_2 is a childbearer, outcomes are worse for A_2 .

This argument is unconvincing. First, it is impossible to view the childbearing characteristic as a disability whose sequelae are unfortunate but not unjust. The able-bodied do not significantly profit from the disabled and may even lose on their account, but non-childbearers do profit from reductions in well-being of the childbearers. Men have benefited and continue to benefit from the reduction in competition afforded by women's non-participation and from the surplus of leisure they obtain by their partial or even total exemption from responsibility for maintenance activities. If there were large numbers of congenital paraplegics in our world, if having a congenital paraplegic in one's household brought a generous subvention, and if congenital paraplegia were a treatable condition, it would be self-interested and unjust for the able-bodied to refuse to treat them on the grounds that their condition was brought on by bad luck.

Imagine an alternative world W' enduring over some ten thousand years, in which each past and present individual is a counterpart of some past or present individual in our world, with precisely his or her capabilities and predispositions—intelligence, speed at calculation, emotional warmth or coldness, love of money, indifference to political power, the tendency to wander, height, and even possession of secondary sexual characteristics, with the following difference: In W' —and this feature differentiates it radically from our world—none of the variance in childbearing ability and responsiveness to children is accounted for by sex. Half of the men and half of the women in W' are subject to pregnancy and experience the same caregiving drives.

If, in W' , the distribution of the components of well-being tracked gender, we would consider W' unjust. Why should the women of W' participate as little and have as little influence as they do in our world, and the men of W' enjoy exemptions from back-up and maintenance activities to the degree we observe here? But if W' is unjust, what makes our world just? It cannot merely be the fact that the percentage of women who bear children is greater than 50 per cent and the percentage of men who bear children is nil. For the women and men are assumed to be equivalent with respect to their competence and interests.

An apologist for the status quo who agrees that W' would be an unjust world might dismiss the thought experiment as irrelevant. Women and men in our world are not, he might insist, equivalent with respect to their competence and interests. The childbearing characteristic is not a feature that can be envisioned as snapped onto or off a human person, leaving all other characteristics unaffected. The presence or absence of the childbearing characteristic pervades the characters and determines the behaviour of women and men in a way 'left by nature', as Hume would say, 'to baffle all the pride of philosophy'. It is to this immanentist position, in its latest fashionable garb, that I now turn.

8.3. Some Favoured Explanations for Female Subordination

Sociobiology—or, as it is now termed, 'evolutionary psychology'—offers the latest and most credible approach to the question why women's status is lower than men's in most human societies. The general form of the answer is that women's and men's reproductive strategies not only are different but are especially conducive to social systems involving female subordination.

These strategies, it is alleged, have evolved through competition between members of the same sex to survive and reproduce. The winning strategies manifest themselves through emotions, aptitudes, and dispositions that are underpinned by differences in cerebral organization, in levels of circulating hormones, and in the presence or absence of receptors for these hormones throughout the body.

While each individual is physiologically unique, differences whose workings are perceptible in ordinary life emerge at the statistical level. According to the hypothesis of evolutionary psychology, women typically possess certain attributes incompatible with high social status and/or lack other attributes conducive to status. As a result, women have little motivation to strive to better their own status, and both men and women have little motivation to help them. The possession and lack of these attributes can be explained by conditions in the early adaptive environment and its pressure on the evolution of human physical and psychological traits.

Sociobiological explanations for the subordination of women, even considered as partial explanations, have not captured the interest of female theorists of sexual inequality for several reasons. First, in scientific discourse, women are often considered and spoken of as objects of knowledge in a manner experienced as insulting. Even in recent books on evolutionary psychology, there is likely to be an index entry for 'Women' but not one for 'Men', affording another example of the presumption that the active world of knowledge-seekers is composed of men who transcend the empirical limitations and determinations to which the more inert sex is subject. Second, the Theory of Women delivered by the tradition is famously unreliable. With regard to the nineteenth-century discourse on human sexuality, Michel Foucault remarks: 'When we compare these discourses . . . with what was known at the time about the physiology of plant and animal reproduction, we are struck by the incongruity. Their feeble content from the standpoint of elementary rationality, not to mention scientificity, earns them a place apart in the history of knowledge. They form a strangely muddled zone.'¹⁴ The texts of the medical moralists, he comments, exemplify at once the 'stubborn will to knowledge that has sustained scientific discourse in the West . . . [and the] stubborn will to nonknowledge. . . . It is as if a fundamental resistance blocked the development of a rationally formed discourse.'¹⁵ Much evolutionary psychology seems to its female readers to exhibit the same combination of weak methodology and projective fantasy.

¹⁴ Michel Foucault, *A History of Sexuality*, i, 54.

¹⁵ *Ibid.* 54–5.

Yet the virtual neglect of this literature is not adequately explained by worries about objectification or the scientific status of sociobiology. It is better explained by the images of reality that are reflected in it and their fit or lack of fit with readers' self-images. Male readers are happy enough to learn from this literature that they are by nature status-conscious and promiscuous, and that they control the distribution of resources to women. Females are puzzled to learn that they are by nature indifferent to rank, naturally monogamous, and have evolved to be nourished by men. Often evolutionary psychologists convey the impression that the variance in outcomes between the sexes is difficult, impossible, or dangerous to tamper with, in view of the great antiquity and serviceability of specialized sexual strategies. Historically, moralists who tolerated or favoured the subordination of women did so on the basis of differences they supposed to be given by nature. Those who argued against subordination did so on the basis of samenesses they supposed to have been overlooked and suppressed. In view of those samenesses, it was urged, the subordination of women was illogical or inconsistent.

Regardless of the historical precedents, the refusal to take seriously sociobiological explanations for female subordination is misguided. For evolutionary psychology has the virtue of shifting the discussion away from the (in theory) long-settled issue of female intellectual and artistic competence to consider other traits with respect to which differences between the sexes may actually be pronounced. The hypothesis that it is precisely those traits that hold the key to the explanation of subordination can explain very well the neglect of sameness, and the further neglect of the intellectual and practical inconsistency arising from this neglect. The accusation that the status of women reflects inconsistent beliefs or an inconsistency between belief and practice might have seemed, for a time, quite telling. Even for philosophers, though, this state of affairs does not constitute a crisis. People hold many conflicting views they do not strive to reconcile. In social and political theory, so long as there is agreement about what is interesting and important, the invitation to account for imputed inconsistencies can be put off indefinitely.

The biological disposition theory of modern sociobiology is not, in the hands of its most careful developers, either deterministic or objectionably essentialist where gender is concerned. It is not incon-

sistent with the observations that each person is a unique mosaic of traits, that individuals facilitate their own sorting by conforming to gender stereotypes and by encouraging or forcing one another to conform to these stereotypes, and that some individuals cannot be sorted easily into the category of male or female. The knowledge that someone belongs in the class of males or females does not enable us to make highly reliable predictions about that person's possession of most traits or capacities. Still, nearly 100 per cent of the variance in childbearing capacities is explained by sex, and this fact may be important. Further, there is reason to believe, not only that there is substantial variation with respect to an interest in and responsiveness to young children by adults, but also that some proportion of that variance can be accounted for by sex as well.¹⁶ It is reasonable to believe that some percentage of the variance, however small, observed in other morphological characteristics, competencies, deficits, and emotional responses can likely be explained by specialization for childbearing and childcare.

According to evolutionary psychologists, the different behavioural and emotional profiles of men and women influence the likelihood of various possible social patternings. Some logically possible patterns will never be manifested at all; others will occur frequently. Social systems in which females have lower status and are more burdened are, on this view, predictably common and assume numerous specific forms. These patterns are not plausibly explained as the results of conscious conspiracies, but they cannot be relegated to the category of mere historical accidents either; they represent various courses of least resistance for dimorphic organisms with certain endowments. On this view, just as our cognitive structures divide the class of logically possible experiences from the class of possible-for-us experiences, and just as the size of our teeth and nails, the structure of our digestive systems, and our metabolisms make some forms of nutrition, such as a diet of cellulose, virtually impossible for us and others possible but problematic, our behavioural and emotional tendencies make some possible institutions more difficult to instantiate than others, and render others impossible or virtually impossible. Evolutionary psychology has, however, no normative implications, since

¹⁶ Hrdy, *Mother Nature*, 212. Women hear infant cries more easily and are more moved to respond to them.

normative statements can only be understood as reality-constrained projections of ideal worlds. Persons whose initial moral commitment to gender equality is on the high-demand side should be able to agree on many facts with persons whose initial moral commitment is on the low-demand side.

With these caveats out of the way, it will be useful to consider the three most prominent explanations for female subordination in the literature. The first focuses on male aggression and competition, the second on the physical encumbrances of maternity, and the third on female vulnerability and altruism.

(a) *Aggression and competition*

Life is not easy for primate females. According to Wrangham and Peterson, writing in 1996:

Among [*pan troglodytes*] chimpanzees every adult male is dominant to every adult female, and he enjoys his dominance. She must move out of the way, acknowledge him with the appropriate call or gestures, bend to his whim—or risk punishment. The punishment by a bad tempered male can vary from a hit to a chase through trees and along the ground, until the female is caught and pulled and kicked and hit and dragged, screaming until her throat cramps, reminded to respect him the next time.¹⁷

Wrangham and Petersen infer that ‘Patriarchy is worldwide and history-wide, and its origins are detectable in the social lives of chimpanzees.’ The impression that male primates are hard on their own females is backed up by observation of other species. Barbara Smuts found that each female in a baboon troop was attacked by a male slightly more than once a week, and that, on average, each female could expect to receive a serious wound, one taking two or three months to heal, from a male once a year.¹⁸

The human platform is different from that of *troglodytes* and baboons. Not only is sexual dimorphism reduced in humans, indicating selection pressures against large and threatening males, the human brain is further specialized for the inhibition of impulses and for moral ideation. Yet it is undeniable that men’s differential ability to injure

¹⁷ Richard Wrangham and Dale Peterson, *Demonic Males: Apes and the Origin of Human Violence*, 205.

¹⁸ Smuts, *Sex and Friendship in Baboons*, 88.

and kill women and women's awareness of their relative physical weakness facilitates control by men of women's behaviour and movements. It creates the expectation in men that women will do what they want them to do, and resigns women to the idea that they ought to behave as men want them to behave.

A related notion is that men's intrinsically higher levels of energy, sexual curiosity, and tolerance for pain translate into accomplishments women cannot expect to match. Competition between men for 'access' to women who select them for their ability to provide 'resources' is seen as a motivating force that spurs the male sex to ever more daring, worthwhile, and lucrative achievements in science, literature, the arts, politics, drama, and sports. Women, with their 'larger gametes', are held to constitute a scarce resource, from which their non-competitive propensities follow.

(b) *Encumbrances of maternity*

In nature, female primates are often accompanied by their young, sometimes an infant and a juvenile, sometimes two infants. Males in many species play with, protect, and take care of juveniles, but the mother-infant bond is intense and universal, and primate mothers may take an interest in their offspring and vice versa for their entire lives. The newly parturient female is hormonally a distinctive creature, and the inclination to respond to a baby's cries, to carry it around, keep track of it, and to feed it seems to depend on the conjunction of the mother's temperament and hard-wiring, social learning by observation of others, and on the behaviour of the child.

Much excited writing pro and con has focused on the question whether maternal care in humans is innate or 'conditioned'. 'There is no such thing as the maternal instinct,' Anne Oakley stated confidently in *Housewife* in 1974. 'There is no biologically based drive which propels women into childbearing or forces them to become childrearers once the children are there.'¹⁹ But one might wonder exactly what is being disputed. Some women do not like children and are annoyed by their dependency; some are careless or vicious and manage to kill or hurt or malnourish them; some are overprotective

¹⁹ Oakley, *Housewife*, 199.

and inadvertently harm their children, but most women desire and welcome the arrival of offspring and are capable parents. Involuntary childlessness is perhaps the greatest source of anguish in women's lives. It is an eminently presentable deprivation.²⁰

Maternity seems to explain a good deal where women's subordination is concerned. Women can work, but maternity means that the work must be such that it is easily interrupted. Women's work must not be dangerous or too absorbing, since becoming motherless, or merely suffering maternal neglect, is a worse fate than becoming fatherless or suffering paternal neglect. It should not involve long journeys since young children have to be carried and cannot move at a comfortable adult speed on their own. All this results, according to the hypothesis, in specialization for detail work in or near the home on the part of women and in specialization for work involving travel, risk, and imagination to men.

Social inequality, on this view, follows from the fact that human males are not as interested in and committed to young children as human females are. While individual men may show high levels of interest and commitment and individual women low levels, the average differences between the sexes with respect to parental investment are thought to predict female social inferiority. By allocating more time to caring for and teaching children, women must invariably allocate less time to the production of other objects and states of affairs deemed valuable in a culture. Since childbearing is a salient feature of women, the disqualification of maternity attaches to the sex as a whole.

(c) *Vulnerability and altruism*

The characteristics that specialize women for childbearing and childcare, even if they do not bear on intellectual, artistic, and practical competence, seem to put them at a disadvantage in modern competitive institutions. According to an argument that is rarely propounded aloud, quite apart from the physical encumbrances of maternity, women's greater sensitivity to physical and emotional pain, greater altruism, and overall lack of toughness lead them down different life pathways.

²⁰ Rachel, in the Old Testament.

Women, it seems, do not attach the same importance to defeating and humiliating rivals as men do. It is known that drive levels tested in isolation are the same in males and females. Women have the same underlying desire to master tasks and to perform well as men do.²¹ Yet when they are placed in competitive situations, the drive to defeat rivals appears to diminish. Boys appear to respond to competition with greater output, girls with less.²² In comparison with men, women are observed to express more misgivings about their own abilities, to have lower expectations for themselves, and to be more tolerant of others' failures. They tend to refer their successes to extraneous factors.²³ To the frustration of their teachers, they often appear to be easily discouraged, and seem to require more coaching and personal attention to perform to a given level in professional life. According to Helen Fisher, females do not establish status ladders: 'They form cliques instead—laterally connected subgroups of individuals who care for one another's infants and protect and nurture each other in times of social chaos. Females are less aggressive, less dominance oriented, and this network can remain stable—and relatively egalitarian for years.'²⁴

The explanation for this diffidence is thought to be as follows: Women depend on the help of other women to raise their children. This leads them to treat female friends and relatives in kind and conciliatory ways. At the same time, they compete with one another for male attention, but not by signalling their cognitive superiority, fierceness, or athleticism. Rather, a shapely figure, maidenly demeanour, and maternal inclinations are believed to impress men. It might be predicted by evolutionary psychologists that, while male–male competitions for prestige will be eagerly studied by males and females alike, female–female skill competitions (unlike beauty pageants) will be thought relatively uninteresting. Cross-sex competition will be avoided by humans or considered not to matter.

Traits such as smaller size and lesser strength, childbearing, receptivity to infants, and greater patience, as well as a lower level of interest in status as opposed to the goods of affiliation are thus, for the evolutionary psychologist, the causes of women's subordination.

²¹ Maccoby and Jacklin, *Psychology of Sex Differences*, 134 ff.

²² *Ibid.* 149–50.

²³ *Ibid.* 154 ff.

²⁴ Helen Fisher, *The Anatomy of Love*, 222.

No malevolent intention on anyone's part need be hypothesized; hence no one can be blamed for women's social deficits.

Even if we accept the evolutionary psychologist's account of our underlying dispositions they do not predict and do not justify the current state of the world. A Martian ethologist given information concerning initial conditions in the early adaptative environment, and information about the competencies, drives, and tendencies of men and women would be unlikely to predict the present condition of the world, including male control of politics, economics, architecture, and culture, or the excess female risk of poverty and abandonment. The ethologist would assuredly not predict that men are more qualified as present-day politicians and artists. Nevertheless, *if* the Martian knew that one sex was going to gain supremacy in these areas, then, knowing *only* that men and women have the preferences and motivations just cited, it might well predict that the winning sex would be the male sex. If the Martian also knew that men are not more qualified than women to decide how things should go and that women do not prefer their condition of enhanced vulnerability, it could predict that Earth would become an unjust place.

The evolutionary psychologist may concede that our institutions have exaggerated the effects of our biology, but he is professionally unequipped to recognize the manner in which the history of our species has created a moral problem that was not present in the early adaptative environment of 75,000–300,000 years ago. Wilson refers to the accumulation of advantage in describing how 'a small evolutionary change in the behaviour patterns of individuals can be amplified into a major social effect by the expanding upward distribution of the effect into multiple facets of social life'.²⁵ This process of leveraging explains observed historical patterns in the relationship between men and women better than the supposition that they are a direct product of biological endowments that would have manifested themselves in any possible world.

Among the few remaining modern hunters and gatherers and nomadic herders whose lives are considered to resemble most closely those of our distant ancestors, female subordination is minimal, for the confinement and management of women is incompatible with

²⁵ E. O. Wilson, *On Human Nature*, 111.

survival. The emergence of permanent settlements and agriculture in the Neolithic altered the balance. Settlement created the need for intensive domestic work to keep dwelling places habitable, and women's larger caloric requirements in pregnancy and lactation and their lesser musculature encouraged sedentary habits that made them natural candidates for this job. Women became the chief tenders of gardens and fields, as they still are in small villages without draft animals or machinery, and the chief processors of grains and seeds. Arranged marriage and the ownership of women, perhaps suggested to the human mind by the ownership and controlled breeding of livestock, diverted men's energies from courtship and food-getting to cultural productions, displays, and politics.

With the advent of metalworking, the physically stronger males gained control of the manufacture and use of iron weapons and of plough agriculture, with its use of large draft animals, then cart transportation, then trade, and so money, writing, and administration.²⁶ Women remained generalists; men specialized. The work they perform varies from culture to culture, but in all human societies, women perform a greater number of separate tasks than men do.²⁷ The urbanization movements of 2000 BCE marked the beginning of women's claustration, exaggerating, to the point of extreme distortion, the female features of modesty and the tendency to energy conservation.

With women's tasks increasingly sequestered and unseen by men, militaristic societies worried about the contagiousness and debilitating influence of femininity. Young men were removed as early as possible from the company of their mothers and sisters, reducing their mutual understanding and their companionship value for one another. The route to high political office proceeded through a military or, by the Middle Ages, a church career, and the development of large administrative structures and political organizations that operated in secrecy tended to women's exclusion.²⁸ Many unisexual institutions—boarding schools, monasteries, and formerly universities—replicate some features of military discipline and camaraderie. It is easy to state conditions under which a system of sex and gender hierarchy is well entrenched and self-perpetuating. The pathways for men leading to positions of power and influence must be clearly

²⁶ As suggested by Marvin Harris, 'The Evolution of Gender Hierarchies: A Trial Formulation'.

²⁷ Boulding, *Underside of History*, 122.

²⁸ *Ibid.*

marked and well paved. The most prestigious educational institutions, with the best endowments and the finest facilities, and those requiring the highest parental investments, must be governed and staffed by males and restricted to them. Under such conditions, the actual competencies, dispositions, and preferences of women, though they cannot be entirely submerged, are thwarted.

Human institutions can be considered, like the nest of the bird, as entirely natural productions, springing from the needs and capacities of a species able, like nearly every species, to modify its own ambient environment. In this respect, we can agree that it is natural to create environments in which women occupy the crowded lower tiers of employment hierarchies and do not enjoy certain privileges. Nothing is added, in one sense, to the critique of institutions by insisting that exploitation and deprivation are merely cultural, for the admission that they are natural carries no implication as to the ease or difficulty of remedies. The cultural disability of lacking a high-school diploma is harder to compensate for than the natural disability of happening to lack a front tooth.

In another sense, however, the subordination of women is not natural, for it depends on contingencies that have nothing to do with sexual strategies of the early adaptive environment. If metal were not malleable at temperatures achievable with charcoal fires; if no wild animals had proved themselves amenable to domestication; if grain could not be stored for more than a week; and if no piles of stones over five feet high could stand, women would not be subordinated. This point is eloquently expressed by Sarah Hrdy:

Incontestably, weaker individuals are often victimized by stronger ones. This can certainly be documented throughout the primates, but never on the scale in which it occurs among people, and never exclusively against a particular sex. . . . Only in human societies are females as a class subject to the sort of treatment that among other species would be rather randomly accorded the more defenseless members of the group—the very young, the disabled, or the very old—regardless of sex. . . . Human ingenuity, and with it the ability to build walls, to count, and tell tales, to transport food and store it, and particularly to allocate labour (to control not just the reproductive but the productive capacities of other individuals), all of these eroded age-old female advantages.²⁹

²⁹ Hrdy, *The Woman that Never Evolved*, 185–7.

The same conditions, as observed earlier, that permit the most varied and spectacular expressions of human artistry and invention, permit at the same time the most varied and spectacular expressions of the will to dominate and control others.

8.4. The Argument from Heavy Costs

Women's past and present deprivations incriminate few identifiable individuals. History reveals for the most part a panorama of microethically unobjectionable choices and transactions, punctuated here and there by a noteworthy act of exceptional antifeminism. Nevertheless, the status of women presents us with a clear example of a moral dilemma. How much ought a well-off group to sacrifice in order to improve conditions for a less well-off group? To what extent does the vigorous and healthy self-interest of the former group, the fact that there is something that it is like to be a member of that group, and that the experiences of individuals within the group are profoundly influenced by membership in the group set limits on what can be asked of them? Moral concern mandates some transfer of advantage from better-off to worse-off, only the questions how much and how soon have no determinate answers. One can only defend the reasonableness of one's proposals in light of the reality constraint and the idealism characteristic.

Greater social equality for women implies some costs for men. Even if we can anticipate benefits to the collective from a reversal of women's fortunes, some individuals will be worse off, for social status and financial reward are limited resources and their allocation is virtually a zero-sum game. The familiar multipliers of costs are likely to weigh in at this stage of reflection. The prospect of greater social equality in Western countries rouses sentiments of relative deprivation; our women seem to lack the biddable qualities they had only a generation or two ago and still have in many places. Affirmative action is seen as externally imposed and the sacrifices implied by greater equality will not be temporary.

In a world in which women are rarely judges or professors, and rarely receive incomes of over \$100,000, my chances of enjoying these benefits are greatly increased if I am a man of the class that normally has access to these positions. In a society in which women

are always domestic cleaners and preparers of food, my chances of having to take on these burdensome tasks are virtually nil if I am a man of any class whatsoever. Self-interested men, therefore, have overwhelming reasons to favour a division of labour along customary lines. Moral philosophers do not share these reasons, for they are technically precluded from theorizing as self-interested men. Though the propensity to treat statistical inequality between the sexes as though it does not matter or to deem it other than a proper subtopic of general moral and political theory need not imply a disregard for the professional role, it can betray a certain forgetfulness.

Conversely, moral theorists who are female are obliged to evaluate betterness relations from an impersonal perspective, not with respect to the outcome they happen to prefer for their own sex. They cannot ignore the fact that greater social equality may impose heavy costs on some groups and persons, reducing, *prima facie*, the *prima facie* obligation to strive for it. Suppose, for example, that by sacrificing one randomly chosen person by lethal injection, statistical equality between the sexes could be assured. Would it be right to do this? Clearly not. However, the utilitarian sacrificing action is not forbidden because it is always forbidden to worsen the condition of even one agent in ways he could not agree to, no matter what benefit follows. It is simply the case that I judge a world in which gender equality has been produced by an execution to be worse, all things considered, than our existing unfair world and I expect competent judges to support me in this claim.

It would take a brilliant imagination to work out a convincing scenario in which the sacrifice of a single human being would produce statistical equality between the sexes. Even an ordinary imagination, though, can work out convincing scenarios in which some sacrifices by human beings contribute to greater equality. What level of equality can we reasonably aim for? How much would it cost us?

A number of respected philosophers appear to be pessimistic on this score. They advance versions of the familiar argument from heavy costs:

The Argument from Heavy Costs against Sexual Equality

The condition of women presents a moral problem that *prima facie*, we are obliged to solve, in so far as it contravenes not only *Q* but even the weaker moral proviso *L* that forbids agents to advance

their own interests when the costs to others are substantial. Although many women are accomplished and favourably treated, women have less of most human goods. But the status of women is nevertheless justifiable all-things-considered on the grounds that the difficulty of reducing the variance in the social status of the sexes reduces the obligation to strive for this goal. We would have to sacrifice many other human goods, in addition to the excess portion of male authority and enjoyment, to eliminate relative deprivation, and it is not worth it.

The pessimist's estimation of the reality constraint motivates his adherence to a low-demand position. He may acknowledge that, in fact, men's relationship to women, individually and in aggregate, often involves culpable letting-happen, as a by-product of self- or class-interest, not just the innocuous failure to prevent harm. For men and women know of one another's existence and know something of the conditions under which the other sex lives; the channels of communication are at least partially open. But, on his considered view, strenuous exertions to reduce the variances in agency and influence, and so to reduce advantage-taking between the sexes, are not all-things-considered morally required. The imposition of policies and procedures that transferred a large proportion of the holdings, liberties, and cultural and intellectual authority of men over to women would be disruptive of reasonably comfortable and efficient modes of life and traumatic for both sexes.

This version of the argument from heavy costs is commonly encountered both in the writings of reflective sociobiologists and in general moral philosophy when theorists have swept their gaze momentarily over the question of sexual equality. E. O. Wilson acknowledges that the aggressive use of quotas and remedial education would enable us to manipulate our society to bring about equality in the professions and cultural activities. 'Yet,' he says, 'the amount of regulation required would certainly place some personal freedoms in jeopardy, and at least a few individuals would not be allowed to reach their full potential.'³⁰ Williams, citing Nagel, presents the

³⁰ Unlike the majority of political philosophers, E. O. Wilson recognizes and states that 'equal opportunity' as currently understood is likely to leave men's higher socio-economic status intact. His preference appears to be for a society that 'condition[s] its members to exaggerate sexual differences in behaviour'. *On Human Nature*, 132-3. This is puzzling since the main thrust of the book is that counter-natural social engineering tends to be problematic.

argument with unusual frankness. 'It might just be', he ventures, 'that [equality] is too hard and will not work.'

That is, there is a *Spielraum* for human beings, an area in which it is possible for human beings individually—or even for a time societally—to do things of a certain kind, but it is so against the grain that some things are just, to use the phrase used by Tom Nagel, too much to ask. Someone will come along and say, 'Look it is possible to treat women just like men, at least almost just like men. But if we try to adopt this equality of treatment everywhere, there will be anxiety, disaster, collapse,—results which everybody knows are unacceptable to human society.' This is certainly a respectable form of claim.³¹

Nagel agrees that it would be a mistake to try to extend juridical equality so far that it produces statistical equality.³² The state, he theorizes, exists in order to prevent serious and remediable harms to persons and can guarantee their juridical status—and it has done so, ending racial and sexual discrimination—but it is not responsible for supervising social outcomes. Further, 'The impersonal desire for equality meets severe obstacles from individual motivation at every step; in regard to the basic institutions to which individuals are willing to give their allegiance, in the process of democratic politics, and in the operation of the economy.'³³

Williams's belief that it could be dangerous to throw open the doors for women by treating them equally everywhere reflects his conception of the social world as governed by *Sittlichkeit* even while the individual is ignobly fettered by ethical theory. His mood of gloomy foreboding might be supposed to arise from contemplation of the difficulties that would be involved in overcoming the disqualifications of maternity and in making women and institutions better adapted to each other than they currently are. Nevertheless, it seems to hint at something darker. Why does Williams think it is 'certainly a respectable form of claim' that a world in which women were treated just like men would spell disaster for human society? We need to

³¹ '[T]he strongest kind of sociobiological "cannot" would mean that the question never came up at all. As soon as [the sociobiologist] permits "can," then the philosopher says "is does not imply ought" and we have room for free choice. What the sociobiologists say here is, "Look, when we say 'can't' we do not mean 'absolutely can't.' What we mean is 'can't without terrific costs that any group of human beings will count as costs.'"' Williams 'Conclusions', in Stent (ed.), *Morality as a Biological Phenomenon*, 142.

³² Nagel, *Equality and Partiality*, 89–90.

³³ *Ibid.* 95.

know what forms of anxiety, disaster, and collapse are more likely to occur, given women's actual propensities and dispositions, when they are treated more equally. Only then can we decide whether the costs associated with a more moral world are really much heavier than we are willing to bear.

One way to interpret this foreboding is as follows: Since Rousseau first expressed it, there has been a presumption in force that women's greater compassion and lesser interest in personal acquisition would enable men to operate as they pleased without the world becoming uninhabitable. Women's altruism, their willingness to 'labour for love', would maintain the collective. Only half of mankind would be fuelled by *amour-propre* to engage in competitive displays and struggles, converting raw materials as fast as possible into artefacts and inventing new intellectual products. The other half would deal with the stresses and strains that are the inevitable by-product of such striving.

Imagine a world in which women are not only treated like men but behave like caricatures of men. A number of women become belligerent dictators, threatening other states. En masse, women adopt the specialist's mode of life, either displacing men or simply adding their labours and effort to double the amount of existing bridge-building, theorem-proving, and merging and acquiring. Further, women develop new predatory habits with respect to young men, and construct a commercial demi-monde parallel to the one we already have to relieve their newly manifest sexual boredom. They stop looking after men and attending to their needs. Social equality has been produced, but at the price of increased aggression and a faster rate of economic throughput, less devotion to the preservation of life. The world is now more dangerous, politically and psychologically, and less attractive than it used to be.

Another source of foreboding is the suspicion that a redistribution of social advantage from men to women might violate Rawlsian principles of distributive justice in worsening the position of the worst-off men. This might in turn have grave consequences for everyone. As a result of increased competition, men who were formerly able to find a social niche might be driven below the threshold of integration and esteem required to sustain productivity and goodwill. They might defect to a resentful and dangerous

underclass. On this view, larger personal incomes, disproportionate respect, and social liberties are preferentially awarded to men in recognition of the fact that the worst-off men are an exceptionally disadvantaged class whose defection can also be dangerous to the collective. Preferential treatment for the gender helps to boost morale in this vulnerable subpopulation.

A third source of foreboding is this: The full participation of young mothers in the important institutions of the modern world seems to require state-funded, board-certified, round-the clock daycare to enable them to meet the work, travel, and entertainment requirements of the modern corporation. This suggests that in any world in which women participate on an equal basis with men in all facets of culture and politics, the comforts of home will be diminished, the intimacies of marriage destroyed by exhaustion, and the acculturation of the young neglected. Overall good worlds, according to the argument from heavy costs, require socio-economic inequality between men and women. Greater equality is a moral luxury that is too expensive for us, not only in terms of money but in terms of our other values, including domestic comfort, responsibility to future generations, and freedom.

Should we be persuaded by these forecastings of doom? In considering this question, it is essential to keep in mind that it is unreasonable to reject all worse-for-someone states when some are morally better. It is often just, as well as relatively inexpensive, to worsen the condition of some people to improve that of others. The knowledge that humans are generally more averse to moving to lower levels of power and wealth than they are eager to advance to higher levels provides only one factual consideration relevant to the assessment of the overall costs of policy change.

The argument from heavy costs for a light, or hands-off, approach to sexual inequality cannot be dismissed out of hand. Its conclusion does not, however, approach the status of a confirmed theorem of the theory of distributive justice. The argument that the removal of the maternity disqualification and the adaptation of women to institutions and institutions to women will be so expensive and burdensome that it ought not be attempted is undermined by three considerations that are commonly recognized as defeaters of the argument from heavy costs. Male advantage reflects the enjoyment of ill-gotten

gains; implies the violation of an implicit contract to cooperate for mutual benefit; and is the product of increasingly culpable ignorance.

Men and women face one another as competitors for the same scarce resources—material goods, autonomy, and respect. At the same time, they are cooperators, dedicated to tasks of interest to both sexes. Women's cooperative role cannot reduce their entitlement to the same level of well-being as men enjoy, nor has nature failed to endow them with the cognitive and emotional resources and dispositions they need to earn their half-share of the cooperative surplus. As competitors, women are nevertheless handicapped by their specialization for one of their cooperative roles—childbearing and the nurture of the young. As cooperators, women are easily exploited in virtue of their seemingly deficient performance as competitors.

The two sexes did not stand in morally precarious relations in the early adaptative environment before the accumulation of male advantage. The suggestion that, by getting in touch with their evolutionary roots, women will come to appreciate the appropriateness of their status is thought-provoking but ultimately unconvincing. The task of the prescriptive moralist is to envision distributive and redistributive protocols that compensate for handicaps and that limit the facility with which women can be exploited to perform the least desirable tasks of the community.

8.5. Recursive Effects of Social Judgement

The extent to which existing meritocratic structures continue to filter men, to develop their talents, to demand that they meet certain performance standards, and to reward them for doing so, is still debated.³⁴ Many formal barriers have been removed, and it is unusual for any authority to express an intention to exclude, ignore, or decline to invest as heavily in women. But in careers in which learning is continuous, discrimination can produce non-measurable educational deficits that are then taken to justify lower rewards. If

³⁴ The Scholastic Aptitude Test, performance in which strongly influences college admissions in the United States, was revamped in the early 1970s in ways that reduced the advantage women had previously enjoyed on the verbal section; see Brush, 'Women in Science and Engineering', 408–9.

habits and customs furthering the transmission of helpful inside information, beneficial criticism, and the provision of opportunities in which to learn favour men, they can surpass women whose initial endowments were equivalent. Women's access to the doubtful and speculative components of well-being is thereby limited.

To understand why women's actual level of effort may have little effect on how they are evaluated, and why their incentives to defect to lower levels of striving and to represent defection as autonomous choice are powerful, it is necessary to look briefly at the theory of social judgement.

Men and women hold numerous beliefs about themselves and each other as members of their respective genders, including beliefs they rarely admit to and may not realize they hold. These beliefs offer a proximal explanation for why women's lesser status seems normal and unproblematic to many people, while facts about the hypothesized reproductive strategies of the sex with the larger and less numerous gametes offer a relatively distal explanation. Men, by contrast with women, tend to overestimate their own abilities and to be more optimistic about what their abilities will bring them. Further, they often fail to perceive how much of their success depends on circumstances and on the actions of others.³⁵ We mostly believe that it is somewhat normal for men to try to annoy, tease, harass, and discipline women, that, unlike a few celebrity exhibitionists on TV, the women one ordinarily meets are retiring by nature and do not seek the limelight, and that women are absorbed in and dedicated to the task of caring for young children, often to the exclusion of other goals and ambitions.

One might suppose accordingly that women's worse outcomes reflect improper stereotyping, but this claim is in many respects superficial.

First, stereotyping is a feature of our normal cognitive apparatus. It enables us to make decisions and predictions in conditions of incomplete information or under time constraints. While it is good advice in some contexts to try to think beyond stereotypes, it is counterproductive to insist that we should stop thinking with them. Second, if women who depart from the typical biological patterns for their sex

³⁵ Valian, *Why so Slow?* 154 ff.

receive favourable treatment while the rest do not, women's overall condition will be little improved. Third, we should be less concerned with representations of groups that smooth out individual differences and more concerned with the accuracy of those representations.

Beliefs about women's dislike of social competition, absorption in child-rearing, potential victim status with respect to men, and lack of competence in public affairs are all *somewhat* true. A sociologically minded visitor from Mars who read the newspapers, walked around in a large city, and watched television would form these beliefs about our women in a matter of days or weeks. And if the Martian were then to read a textbook of sociobiology, it might well go away feeling that it had achieved a deep understanding of the human social world. The Martian would not see much point in striving to better women's status. It will be hard, it would think, and it probably will not work.

Yet many beliefs about women are untrue. It is difficult to overestimate the disadvantage that accrues to women in situations in which they are in competition with men because they are smaller and lighter than men are and because their voices are higher and more childlike. People perceive women as smaller than they really are because they are smaller than men are.³⁶ And people perceive women as less competent in public affairs than they really are because they display less competence in public affairs than men do. Both men and women represent women to themselves as smaller and less capable than they really are because they are less of all these things in our world than men are. Misperception can reach startling levels. Married men, for example, believe not only that they are better financial providers and better informed, but that they are more intelligent and physically attractive than their wives are.³⁷ Because we know that male and female intelligence are equal, we can conclude that a significant number of men are unable to assess the intelligence, relative to their own, of women with whom they interact on a daily basis. It is unlikely that they do better when required to assess the intelligence of women whom they encounter on an occasional basis or in the course of a one-hour interview. Evaluators, it has been shown,

³⁶ Valian, *Why so Slow?*, 6. Height is a great advantage. Boulding reports research showing that bishops are taller than clergymen; university presidents than college presidents. *Underside*, 35.

³⁷ Louise H. Kidder, Michele A. Fagan, and Ellen S. Cohn, 'Giving and Receiving Social Justice in Close Relationships', 245.

persistently discount evidence regarding women's achievements. The same objective qualifications—the same degrees, publications, years of experience, and attainment of skill levels—redound to the credit of a woman who is being evaluated less than they do to the credit of a man.³⁸

Contrary to Fisher's confident assertion that women prefer conditions of equality, there has been little controlled observation to support the claim of female indifference to rank.³⁹ Yet women's behaviour in meritocratic institutions is often puzzling to managers. Adult women often seem less motivated than men to maintain party discipline, to accept the authority of others, to help other women, particularly those younger than themselves, and to expect help from older ones. Women do not, in this case, lack ambition; on the contrary, they tend to insist too much, by male standards, on preserving their independence.⁴⁰ They may fail to perceive the connection between ambitious striving and social rewards, either because rewards are withheld, or because they come unpredictably, or because they are observed to devolve upon the obviously wrong persons of their sex. To a female observer, the stable and relatively egalitarian system of 'laterally connected subgroups' is more visible amongst male professionals than amongst their female counterparts.

Automatic habits of appraisal and even of self-appraisal are nevertheless responsive to information and criticism. Levels of confidence, expectations, and self-attributions of merit are plausibly seen as reactive states of mind that are developed through social experience, that have a powerful causal effect on how things go, and that are easily influenced and changed by altering the customary habits of response. Affirmative action has effectively made available the reserve of previously unsuspected female competence by forcing those who control and distribute power to depart from their customary modes of directing their attention, reading signals, and awarding resources under their control. It is worth pointing out in this connection that investment in female talent before the introduction of affirmative action did not bring with it a reasonable pay-off for the investor. It

³⁸ Valian, *Why So Slow?*, 167 ff.

³⁹ Female-female competition and hierarchy-establishing behaviour is the basis of social organization in some primates. See Hrdy, *Woman that Never Evolved*, 128.

⁴⁰ Shepher and Tiger, 'Female Hierarchies in a Kibbutz Community', 232.

was once the case that only the most impartial and dedicated persons allocated their time and effort equally between the sexes. Affirmative action brings down the price of what was once an expensive moral luxury. Treating his or her students with equal consideration and attention is now within reach of the average professor.

Culpable ignorance in the Theory of Women is clearly on the wane. Yet not so long ago, many people seemed genuinely uncertain that females had the same cognitive abilities as males. Male savantism has made an overwhelming impression, registering with us as clear proof of overall male capability and female disqualification. Not so many years ago, a certain cultural anxiety hung in the air about whether women could perform complex tasks. Could they really deal with large numbers and make responsible decisions? Could they comprehend tangled legal cases, or would they become confused? Could they propound hypotheses as boldly and structure scholarly articles as elaborately as men? Looking around at the population of women, the intuitive statistician saw little evidence of an ability to do any of these things. He worked out the correlation coefficients in his head and concluded that there must be causal laws at work. There is no lack of objective evidence that women can do things they, for the most part, still do not do, as should be expected from the absence of significant differences in male and female perceptual and cognitive abilities and motivational structures. It is encouraging that hard-won genuine knowledge of women's underlying capabilities is gradually replacing the intuitive sociologist's deductions from experience. A letter of reference received recently for a female candidate reads as follows:

Beyond understanding causal inference models, this work requires familiarity with projections in Hilbert spaces, semiparametric efficiency theory involving tangent space calculations for infinite dimensional parameters, estimating function theory for the purpose of construction of locally efficient estimators. In addition, it involves understanding how to simulate data from a marginal structural model. Jennifer has also implemented a simulation study for the proposed estimators which has as its goal to determine the practical performance and practical challenge of estimation of causal parameters in longitudinal studies in which treatment at a given time is assigned as a function of the observed past.⁴¹

⁴¹ My thanks to the author of this letter, who shall remain nameless, for permission to reproduce this paragraph.

Evidently, Jennifer can access savant capabilities most women—and most men—cannot. Yet, for the letter-writer, she is an ordinary brilliant student, of the sort that appears now and then. Journalists nevertheless hint darkly that women's difficulties with spatial rotation and abstract thought reflect their non-adaptation to big-game hunting and explain why they do most of the housework.

Modern humans attach an importance to achieving and employing correct and effective representations. We believe that it is important to learn how to read competency signals correctly, not only to substantiate the claim that the professions are organized on meritocratic principles, but also to weed out the inefficiencies that result from preferring the less able. The costs of revising our ordinary beliefs about the nature of physical objects and their motions and discarding our intuitive cosmology, physics, and matter theory have been very high. Intellectual struggles by individuals, and substantial public and private funding, were required to get the reasonably accurate picture we think we possess. Yet no sociobiologists and few moral philosophers have expressed anxiety and pessimism about the diversion of resources needed to teach physics and chemistry, or the regimentation the learner must undergo, or the socially disruptive effects of new knowledge in those fields. It has never been considered an argument against teaching physics that the amount of regulation required places some personal freedoms in jeopardy, though millions of high school students can testify that it does so. It will be protested that the theory of gender similarities and differences and the theory of social judgement are far less certain than physics and chemistry, that much that currently passes under those headings is almost certainly wrong, and that to teach what might turn out to be mistaken is far more consequential in social theory than in chemistry. Very well, but all knowledge emerges from error and confusion and in the meantime there are disappointments, wasted efforts, and occasional unintended explosions.

8.6. Policies for Equality

The idealism characteristic of a theory of gender relations is still largely unformulated, perhaps because the reality constraint has

proved so problematic. Certainly, the paraworlds of the sociobiological literature do not encourage readers to see male–female relationships as potentially perfectly cooperative, in so far as they portray the sexes as exploitable resources for one another, which, from a scientific perspective, they are. The cooperative relationship of the modern household is built upon the pre-moral platform and the partners to it are likely to remain residuary opportunists, just as modern employers remain residuary slave-holders. Yet the superposition of moral ideation alters the terms of these primitive forms of getting along. Human ingenuity makes possible both intense exploitation and compensatory moral advantage-reduction.

Whenever a man or a woman joins an architectural firm, or goes to work for a hospital or educational institution, or in a law office, whether as a director or as a cleaner, he or she is entering into a cooperative relationship with members of the opposite sex for a broader purpose. For trust and non-exploitative cooperative behaviour to develop, individuals have to know each other well. Modes of existence in which the contributions of the two parties are distinctly visible to one another, even when they perform separate tasks, appear to be conducive to mutual respect and appreciation.⁴²

The possibility of a normalization of the workplace, such that the standards of reciprocity and mutual assistance, as well as friendly rivalry, are maintained in cross-gender interactions, is one of the most appealing aspects of the potential breakup of sex and gender hierarchies. The companionship value of men and women for one another is substantial wherever resentment has not taken hold. When one is given the opportunity to observe men, women, and children in informal surroundings—at picnics and festivals, for example, or while waiting around for something to happen, wherever men’s attention is not distracted by competition with other men or attempts to appease them—their easy sociability with children and their friendliness towards women, by contrast with the tension observed under other circumstances, is impressive.

Men stand under a moral obligation to divert some proportion of their energies from productive and directive activities to maintenance and uncompensated amateur activities. This general obligation is

⁴² Boulding, *Underside*, 10 ff., 35 ff., 292.

impersonal and objective. It derives not only from the distinct moral preferability of a world in which maintenance is not downloaded onto the sex whose lesser earning capacities are unrelated to the competence and interests of its members, but also from the practical failure of the division of labour with respect to sustainability. The traditional division into homemakers and resource extractors-and-converters has failed to project into a global system in which the maintenance of the habitat proceeds in parallel with and compensates for wealth production. The natural environment was for a long time seen as no one's responsibility, male or female, to look after, and the world is rapidly filling up with manufactured objects that no one knows how to clean, from bathroom taps and home juicers to burnt-out transformers and atomic power plants. It is further reasonable to divert some proportion of funds that would otherwise be spent on roadworks, street light upgrades, and subsidized development projects to childcare provisions for mothers in the active world. All such projects represent investments in infrastructure in which the public has an interest.

A glance at the magazine section of the airport news-stand with its neat division of the separate life-worlds of the two sexes—money, electronics, and muscle building for men; medical and diet news, home decor, and psychological manipulation for women—suggests that the calculated strengthening of gender identity is highly profitable. It is easy to induce people to forgo other consumption opportunities in order to purchase sex-specific esoteric knowledge, though the low news-stand price assures us that the information is neither very secret nor very useful. Yet Wilson's suggestion that 'In theory, at least, a carefully designed society with strong sexual divisions could be richer in spirit, more diversified, and even more productive than a unisex society'⁴³ seems correct in one respect. In order for women to gain in social dignity, liberty, and enjoyment, it ought not to be necessary for them to adopt the habits, postures, and interests of men. Beyond the permission to strive for sexual distinction with special gender-specific clothes, adornments, and mannerisms, accommodation to different genuine preferences with respect to childcare is reasonable. Insisting on a fifty-fifty division of baby-minding might be

⁴³ E. O. Wilson, *On Human Nature*, 132–3.

unacceptably burdensome for most men and unacceptably anxiety-provoking in most women.⁴⁴ Any system that demands that women sacrifice virtually all contact with children as a condition of their enjoying substantial participation and political influence and the use of their analytical and creative faculties is exceedingly costly in human terms.

At the same time, we know that many children see too much of their mothers and too little of their fathers. Exposure to the skills, interests, and energy levels of adult men is bracing and salutary for the young of both sexes, and exposure to the weaknesses and dependencies of the young is in turn beneficial for adult men. There are further reasons to be wary of the proposal that a society that strives to enhance rather than to efface sexual division will not only be more productive and diversified but also actually better than one that does not. Michael Chance surmises that the agonistic mode corresponds to a 'residual psychophysiological element' built into the constitution of men, and that social circumstances can elicit and strengthen hyper-vigilant behaviour. Early experience and training directed to what he terms an essentially rank-ordered adult life may strengthen the effects of hormonal biases 'that may, in later life, be at variance with a civilized life, create unsuspected biases of behaviour and perception, and thus lead men and women to become so divergent as not to get on with one another'.⁴⁵

According to legend, male cultural accomplishments are not a function of their exemption from domestic responsibilities and the social expectation that they will employ the resulting leisure profitably. Rather, men's intrinsically greater ambition produces the accomplishment that justifies the exemption and explains the expectation. There is no reason, however, to suppose that sexual curiosity, risk-friendliness, and impatience, to name the qualities most commonly supposed to distinguish the male sex, however well documented, and however hormonally mediated and thus resistant to acceptable tampering they may be, are the precursors of initiative, persistence, creativity, concentration, critical ability, and the ability to withstand criticism. There is no reason in other words

⁴⁴ For a contrary view, see Virginia Held, *Rights and Goods*, 205 as well as Okin, *Justice, Gender and Equality*, 175.

⁴⁵ Chance, 'Sex Differences in the Structure of Attention', 159.

to suppose a pre-institutional connection between competitive and dominance-seeking behaviour and cultural achievement and, indeed, there is good reason not to.

In his study of the focuses of attention in various primate species, Chance contrasts the 'hedonic' behaviour of the chimpanzees he observed with the 'agonistic' behaviour of baboons and macaques, arguing that the preoccupation of the latter with rank order 'has, in their evolutionary past, led to a limited expressive repertoire . . . ; it has imposed upon them a form of social attention that precludes the visual awareness essential for the development of tool use'.⁴⁶ To the extent that male chimpanzees experience a greater flexibility in their control of their attention and their emotional arousal, they are free to engage in playful and experimental activities. The ability to ignore status-establishing and maintaining activities and to refocus on other tasks is thought to be conspicuous in females, and it is perhaps not accidental that the Ancients ascribed the majority of human inventions—amongst them, writing, weaving, and numbers—to women.

As noted earlier, male humans are far more likely to exhibit savant syndrome, now considered to stem from damage to the left hemisphere and to higher-level memory circuitry that is compensated by the development of additional habit-memory and right-brain capacity.⁴⁷ Though psychologists consider savantism a dysfunction on account of the notable deficits in affect and conversational ability it entails, savant talents such as lightning calculation, mechanical ability, drawing, music, and map-memorizing are highly prized in normal people. It would not be surprising if normal men were slightly more able, or more likely to be able to access some savant capacities, though this is likely to occur—according to the hypothesis—at the expense of their ability to experience the depth and complexity of affect and to produce elaborate and meaningful verbal representations. Yet the differences between the sexes cannot be pronounced in this respect, for we have no doubt that men overall meet the human standard for feeling and language well enough to perform important social roles. To complement our Jennifer X, we have our occasional William S. and Leo T., manifesting a verbal fluency and mastery of emotional nuance remarkable in their sex.

⁴⁶ Chance, 'Sex Differences in the Structure of Attention', 144.

⁴⁷ Darold Treffert, 'The Idiot Savant: A Review of the Syndrome'.

If girls' presumed greater distractibility, sensitivity to social cues, and integrative abilities are not countered by an education that emphasizes concentration and abstraction, they may not develop valued savant capacities. If boys are not guided to the study of poetry and literature and if adult men do not continue to cultivate the arts of language and the shadings of affect in later life, Chance's vision of a depressingly polarized society in which the sexes do not understand each other and cannot get on well is likely to be realized. Traditional education was compensatory in forcing girls to do word problems and trigonometry and boys to ponder romantic novels; traditional notions of culture presupposed a common ground. The assumption was that some degree of repression could be justified as a necessary condition of the liberation of potentials and the enhancement of opportunities in later life. Though the educational process was effortful and not entirely successful, the thinking behind it was sound.

8.7. Love as a Morally Relevant Phenomenon

Nowhere is the discrepancy between first-person and third-person viewpoints as evident as in the experience of romantic love, which might be defined as the conjunction of delight in the speech and company of another, solicitude, and desire, together with a second-order demand for exclusivity or at least considerable privilege with respect to companionship and disclosure.

Love is at once an important subject of imperatives and a claimant on exemptions. It is characterized by, to paraphrase Plato, the involuntary assumption of a position of disadvantage, as well as self-serving resourcefulness. Myths and stories link love with sacrifice and labour, more often labour for than labour alongside, and wherever the labour is for another's benefit and involves self-sacrifice, it introduces moral significance into the relationship. Love is a great shredder of contracts, the severer as well as the forger of human bonds, stating its own demand for exemptions. Duties of presence are forgotten and their associated automatisms interrupted. Further, the vulnerabilities of the pair offer outlets for the controlling and even persecutory impulses of the socially dominant. Williams's 'Moral Luck' is perhaps really a *plaidoyer* for the privilege demanded by love, represented in the story

less by exotic Tahitian women than by 'Gauguin's' painting, to which he stands in a genuinely intimate relation, with which a powerful authority—what Williams terms 'the morality system'—threatens to interfere.

Sociobiologists may think they know what love is for, what its function or functions might be. But the notion that it produces the motives to parental care is not credible. Why should attachment to an unrelated man be a condition for a woman's caring for her own baby? The function of love may be simply to ensure conception with a preferred consort in a low-fertility species. Or its purpose may be to maintain fluidity in social systems, preventing them being dominated by autocratic individuals. A creature that is highly motivated to evade social control and select its own mates may have an advantage over a less emotional one. Deepening the mystery is the fact that homosexual pairs experience the same romantic emotions. In what follows I shall speak of men and women as offering the more visible and frequent examples of the phenomenon, but I do not mean to underestimate the intensity or worth of same-sex pairings.

Human ideation, in any case, makes all functional accounts inadequate. As E. O. Wilson remarks, the initial experience 'permanently alters the adolescent mind'.⁴⁸ It is a fountain of creativity that kicks away writer's block, sweeps out accumulated emotional debris, and dictates their own poetry to clods and pedants. The predisposition to the emotion is so great that it is capable of creating its own object. People fall in love with faces glimpsed in a window, film stars and characters in books, and with persons they have merely heard about. Even defeated love takes its time departing; our mental universes are populated by not only our relatives, living and departed, and assorted authority figures who have to be reckoned with, but the wronged or wronging ghosts of the past. Love did not need invention and both sexes are assuredly subject to it, but in periods of rough manners, it goes into cultural decline and needs periodic reinvention.

Where personal relations between the sexes are concerned, humans appear to operate with three principal systems, related in different ways to the phenomenon under discussion, the romantic or courtly, the labour-dominant, and the vice-driven. Individuals take

⁴⁸ E. O. Wilson, *On Human Nature*, 60.

part in—and also refuse to take part in—some or all of these systems simultaneously, in the courtly, because its modes of address and receptivity come naturally, wherever education or law has not suppressed them, in the labour-dominant, as creatures partaking in the struggle for existence. Vice—occasional relations that exclude mutual love and responsibility—is an accessory system. Moral problems arise from the intersection and interaction of the three systems, confusion or disagreement amongst participants as to which system of relationship is in play, and from the perils inherent in each.

Courtly relationships are based in elective affinity, in a preference for one person's company over the company of others. Their emotional temperature ranges from, at the low end, the weak but pleasing dependencies of acquaintances and colleagues who are less than perfectly indifferent to one another, to, at the high end, romantic obsession. They are characterized by mutual deference and attention. Their terms are not enforceable, they require trust between individuals that is not backed up by the approval of others or legal sanctions, and they may or may not be durable. Lacking external validation, romantic friendships, as Kant pointed out, tend to destroy themselves in quarrels and misunderstandings.

Any society that attaches importance to the emotions of individuals recognizes these preferential bondings and allows them some role in the organization of that society, through the permission rules governing the initial choice of a mate, the rights of divorce and remarriage, or socially tolerated infidelity by one or both sexes. Courtly relations may be precursors to fixed, labour-dominant relations, or noticed and fostered by third parties in order to found them. Or they may coexist as separate, paramarital systems. Such systems may be developed and formalized by non-labouring social elites—the society described by Lady Murasaki in fourteenth-century Japan, in which the lovers pass their days composing and hand-lettering poems for one another, or the historians of the *amours* of princes and princesses in eighteenth-century France. Even in small villages in which surveillance and gossip assist religious authorities in enforcing strict norms, loving attachments arise between men and women of the same age. The idiocy of rural life, as Marx saw it—or the idiocy of the life of the aristocratic Genji, as MacIntyre sees it—is perhaps mitigated by the way in which people tend to find particular

others interesting. They brave disgrace and even—in the case of women—execution or jealous murder, for reasons of preferential attachment.

Wilson describes the pair-bond as natural for humans and the intensity of these relations and their ubiquity in every period of history leave no doubt that such inventions as group marriage or Plato's community of wives can never be more than thought experiments or the shortest-lived of social experiments. But Wilson does not quite succeed in explaining which form of pair-bond should be considered natural.⁴⁹ In any case, what is natural is only one consideration, though an important one, with respect to the constitution of social rules. Courty modes of interaction between men and women, together with the vicious ones, can either stabilize or destabilize labour-dominant marriage, and films and stories, as well as workplace guidelines, correspond to different modalities in which societies discuss the threat of destabilization with themselves, and through which resolution of the conflict is hopefully, but inadequately, sought. Many novels affirm on their surface the precedence of *plein air* contract over the darker affinities, and the right of its upholders to apply force, while lodging a protest underneath. Institutional codes of conduct seek to prevent the confluence of political and emotional ends, as well as the appropriation of young love by aged vice, recognizing how tempting and at the same time how dangerous to the participants and to others such hybrid relationships can be. We are forced, in these matters as elsewhere, to elect positions that reflect our best estimations of reality, causal relations, and optimality, but no one can claim that their judgements in these matters are true or even well confirmed.

In the labour-dominant system, long-term partners select one another or are selected for one another—often, but not always through love and preceding courty relations—to serve as helpmeets and companions in the service of economic production and biological reproduction. This may involve joint labour in maintaining a household and raising children but also cooperative production, as occurs in peasant societies or family-run commercial enterprises. Importance is laid on the durability of these pair-bonds, and natural

⁴⁹ E. O. Wilson, *On Human Nature*, 139 ff.

sentiment, inner resolve, and external pressure contribute to their maintenance in varying proportions. They are the subject of contract and publicly witnessed, and there is legal recourse in the case of abandonment. Only a very high degree of mutual or one-sided disaffection, traceable to causes that can be enumerated and objectively verified, is regarded as justifying the violation of the agreement to work together, especially when children enter the picture.

Long-term monogamy is the least discriminatory of possible marriage systems in the following sense: It tends to even out the chances of reproductive success, at least amongst men and probably amongst women as well, by reducing the role of favour and preference in the distribution of resources. It furnishes the benefits of companionship and cooperation to more persons than other systems, including a system of transitory, hence highly competitive, bonding. In this respect, it offers *prima facie* moral advantages, preventing those with unusual allure from collecting too many mates and helpers and those with relatively little from being left out. It diminishes anxiety over the possible loss of a pleasing companion, gives rise to the satisfactions of narrative continuity and shared history, offers the opportunity to enjoy the company of one's own children and to influence their development, and provides assistance in daily living for the old. It is conducive to the preservation of wealth. Monogamy frees up energies for pursuits other than courtship in the young, and reduces ambient levels of jealousy. At the same time, long-term monogamy's moral and non-moral disadvantages—overfamiliarity, excessive dynastic-patterning, hidden violence, appropriation of labour, and marital condescension—have to be factored into the balance. To embark on it in a serious way is to renounce an advantage over and against a deceivable or dependent intimate. The exclusivity demanded is experienced as hard—a little hard for most people, very difficult for others, impossible for some. Where emotional security is concerned, marriage is thought to confer overwhelming benefits, yet abandonment has especially grave consequences in a culture that is socially and economically organized for formal monogamy. The anguish of the repudiated partner is perhaps no less in more fluid systems, or in unsanctioned unions. But to the insult—even if it leads hyperbolically to exile, the convent, or the grave—is not added the injury of financial shock and chaos in the domicile.

In vice-driven relationships, sensation, power, and revelatory knowledge are sought and are either acquired by misrepresentation of the quality and magnitude of the cooperative surplus in prospect, or simply paid for. Preference is one-sided or feigned for advantage. The term love may figure, as euphemism, or as a vehicle of deceit. For the sake of their own well-being, individuals are often called on to determine whether they are inadvertent participants. The boundary between courtly relations and vice is not always clear, for a single night can be the beginning of a great romance or mere exploitation. The boundary between labour-dominant marriage and vice is equally fuzzy; exactly what the trophy wife does for her maintenance may decide it.

The regulation of vice is the concern of public officials who deal with imagery and behaviour, and with their by-products, but vice impinges on private life as well, in so far as individuals must decide whether to participate in it or tolerate the participation of their intimates. Friedrich Engels's view that vice is the natural accompaniment of economic coercion and that it poisons and subverts all preference-based male–female relationships,⁵⁰ is shared by many persons whose repulsion from commercial sex is considered rather inarticulate and negligible, in so far as they often cannot name any specific harms that arise necessarily from the renting out of one's person to strangers or for the production of simulacra. The associated evils—disease, stress, drug-addiction, brutality, and the habit of mendacity—are held by libertarians to be contingent and unrelated to the essential character of the system, products of its partial repression, or inconsequential in the face of the rights to commercial and personal freedom. Were prostitutes to view their work and to have it viewed in the same light as, say, appliance repairmen or hairdressers, its sequelae, they think, would trouble us no more. Second-tier goods, one might suppose—and access to the opposite sex is a known and necessary component of well-being—should be supplied to all, regardless of the merit of the needy person. The devotion or good manners normally required to attract the favourable attention of women should not be preconditions of men's obtaining it. Vice, some believe, reduces the frequency of rape.

⁵⁰ Friedrich Engels, *The Origin of the Family, Private Property and the State*, 138.

It is not clear, however, that a world with a free market in women would have less brutality, stress, addiction, and disease than one without. On the grounds advanced by Williams, it would be unjustifiable to require *me* to become a prostitute to reduce the overall frequency of rape or grievous frustration in the population; is it right to assign some more tractable person this duty in a better world? Further, there are serious externalities to both prostitution and pornography. Disrespect transfers to all members of the-sex-that-lets-itself-be-prostituted, just as the maternal disqualification transfers to all members of the-sex-that-has-the-babies. Both prostitution and the easy availability of pornography reduce the leverage of wives and lovers by breaking their monopoly on the provision of gratifying experiences. One might feel that that particular monopoly ought to be broken, but female refusal provides needed balance in a world in which women are disadvantaged by their lesser earning power and have few means at their disposal to induce men to upgrade their behaviour. Access to the other sex is best considered a second-tier good, like parenthood or friendship. These affiliative goods are adequately supplied in more attractive worlds, but personal effort, not merely the exercise of purchasing power, is needed to obtain them.

Labour-dominant relations, meanwhile, are potentially as genuinely cooperative as any other relations. Two or more individuals engage in concerted or coordinated actions that enable them to accomplish work that neither could accomplish alone or accomplish as easily. The questions that can normally be asked about cooperative relationships can be asked about domestic partnerships, as well as all other mixed-sex partnerships. Is the work accomplished of importance to both parties? Is the cooperative surplus real, or deferred, or even imaginary? Is the cooperative surplus appropriately shared? Does it reflect the eligibility of the subtasks performed? Does it overcompensate for scarcity and accumulated advantage?

The formalities of marriage emphasize its cooperative aspects. At the same time, the romantic love that normally precedes and ushers in the prospect of labour-dominant relations often seems to make such questions irrelevant, or reduces the motivation to bargain, leaving many actual relationships at the basic end of the cooperative spectrum. Though game theorists emphasize that it is better to decline

exploitative offers than to improve one's condition relative to its present state, this policy is rational only under the assumption that the bargaining game with the same player will have further rounds. Many significant decisions are one-off; the same game will never be played again opposite the same partner. And relevant information regarding the propensities and dispositions, hence the likely future actions, of the other bargainer may be unavailable, or available but psychologically difficult to grasp. Rational, self-interested persons invited to play a single-transaction game—to decide to accept or issue an offer of marriage or not—often make a decision that improves their condition relative to what it would otherwise have been while creating a standing situation of injustice.

Societies that value the nuclear family and that consider the presence of affective ties between mothers and fathers of signal importance act semi-rationally from the economic point of view in making the institution of marriage easy to enter into and difficult to escape. Such incentives and restrictions are conducive to the fair distribution of the cooperative surplus, in so far as single parenthood is a liability for a woman in a wage-based economic system. Yet the incentives and restrictions do not always work as intended and contingency plans for their failure are lacking in our overconfident society. Nor is the emphasis on stability conducive in every way to human well-being. Again, the litany of universal human grievances bears witness. We know that people are made miserable, desperate, and even suicidal by laws, authorities, and relatives that stipulate or interfere with their companionship preferences. Individual happiness is fostered by greater trust of the pair-bonding instinct and of individual preferences, by deference towards these preferences from third parties, by easier exit in case of unilateral or bilateral disaffection, and, finally, by minimization of the unfortunate consequences of abandonment for adults and children, in short, by concessions to the biological platform that moderate exigent moralities. At the same time, we know that inaccessibility promotes ideation, that ideation can give rise to idealization, and that idealization produces what we think of as the greatest achievements in art, literature, and culture. Mild repression is consistent with the generally benign nature of a good moral system.

Human animals like to hang themselves with the weights of grave responsibility. Koehler's chimpanzees behaved in a way he at first found puzzling:

Almost daily, the animals can be seen walking about with a rope, a bit of rag, a blade of grass or a twig on their shoulders. If Tschego was given a metal chain, she would put it around her neck immediately . . . Tercera also has strings running down the back of her head and over her ears, so that they dangle down both sides of her face. . . . Chica, the sturdy, at one time took a fancy to carrying heavy stones about on her back; she began with four full pounds and soon reached a powerful block of lava weighing nine pounds.

These observations led Koehler to believe that the pleasure of adornment lies in 'the extraordinary heightened bodily consciousness of the animal. It is a feeling of stateliness and pride, indeed, which occurs in human beings when they decorate themselves with sashes or long tassels knocking against their legs . . . : when anything moves with our bodies we feel richer and more stately.'⁵¹ Pace Williams, the enhancement effect of the burdens and encumbrances of the morality system and its often heavily born responsibilities is not to be underestimated. Williams encodes happiness as flight and evasion. Though, from the objective standpoint, romance is a significant element in most human lives, this encoding cannot be taken as the fundamental insight of a theory of male–female relations, or of a morality for a social world.

⁵¹ Koehler, *Mentality of Apes*, 95.

Bibliography

- Abramovitch, Rona, and Strayer, Fred, 'Pre-school Social Organization', in Lester Kames, Patricia Pliner and Thomas Alloway (eds.), *Aggression, Dominance and Individual Spacing* (New York and London: Plenum, 1978), 107–28.
- Alexander, Richard D., *The Biology of Moral Systems* (Hawthorne, NY: de Gruyter, 1987).
- Annas, Julia, 'Personal Love and Kantian Ethics in *Effie Briest*', *Philosophy and Literature*, 8 (1984), 5–31.
- Aristotle, *Nicomachean Ethics*, tr. W. D. Ross, in *Basic Works*, ed. Richard McKeon (New York: Random House, 1941).
- Arneson, Richard J., 'Perfectionism and Politics', *Ethics*, 111 (2000), 37–63.
- Axelrod, Robert, *The Evolution of Cooperation* (New York: Basic Books, 1984).
- Baier, Annette, *Moral Prejudice: Essays on Ethics* (Cambridge, Mass.: Harvard University Press, 1994).
- Beauvoir, Simone de, *Le Deuxième Sexe*, 2 vols. (Paris: Gallimard, 1949).
- Bentham, Jeremy, *Introduction to the Principles of Morals and Legislation* (Buffalo: Prometheus, 1988).
- Blackburn, Simon, *Spreading the Word* (Oxford: Clarendon, 1984).
- *Ruling Passions* (London: Oxford University Press, 1998).
- Boulding, Elise, *The Underside of History* (Newbury Park: Sage, 1992).
- Braybrooke, David, 'The Concept of Needs, with a Heartwarming Offer to Utilitarianism,' in Brock (ed.), *Necessary Goods*, 57–72.
- *Meeting Needs* (Princeton: Princeton University Press, 1987).
- *Natural Law Modernized* (Toronto: University of Toronto Press, 2001).
- 'The Representation of Rules in Logic and their Definition', in Braybrooke (ed.), *Social Rules* (Boulder: Westview, 1996), 3–20.
- Brock, Gillian (ed.), *Necessary Goods* (Lanham, Md.: University Press of America, 1998).
- Brush, Stephen, 'Women in Science and Engineering', *American Scientist*, 79 (1991), 404–19.
- Burke, Edmund, *Reflection on the Revolution in France*, ed. J. G. A. Pocock (Indianapolis: Hackett, 1987).
- Chance, Michael R. A., 'Sex Differences in the Structure of Attention', in Tiger and Fowler (eds.), *Female Hierarchies*, 135–62.

- and Jolly, Clifford J., *Social Groups of Apes, Monkeys and Men* (London: Cape, 1970).
- Clutton-Brock, T. L., and Harvey, Paul, 'Evolutionary Rules and Primate Societies', in Clutton-Brock and Harvey (eds.), *Readings in Sociobiology* (San Francisco: Freeman, 1978), 300–3.
- Cohen, G. A., *If You're an Egalitarian, How Come You're so Rich?* (Cambridge, Mass.: Harvard University Press, 2000).
- 'Robert Nozick and Wilt Chamberlain: How Patterns Preserve Liberty', *Erkenntnis*, 11 (1977), 5–23.
- Copp, David, 'Moral Realism: Facts and Norms' (Review of David O. Brink, *Moral Realism*), *Ethics*, 101 (1991), 610–24.
- Cottingham, John, 'The Ethical Credentials of Partiality', *Proceedings of the Aristotelian Society*, 98 (1998), 1–21.
- 'Partiality, Favoritism, and Morality', *Philosophical Quarterly*, 36 (1986), 357–73.
- Daniels, Norman, 'Wide Reflective Equilibrium and Theory Acceptance in Ethics', *Journal of Philosophy*, 76 (1979), 256–82.
- Darwall, Stephen, *Philosophical Ethics* (Boulder: Westview, 1998).
- Dasgupta, Partha, *An Inquiry into Well-Being and Deprivation* (Oxford: Clarendon, 1993).
- Davidowicz, Lucy, *The War Against the Jews* (New York: Holt Rinehart and Winston, 1975).
- Davis, David Brion, *The Problem of Slavery in Western Culture* (Ithaca, NY: Cornell University Press, 1966).
- DePaul, Michael R., 'Two Conceptions of Coherence Methods in Ethics', *Mind*, 96 (1987), 463–81.
- Diana, Lewis, *The Prostitute and her Client* (Springfield: Charles C. Thomas, 1985).
- Douglas, Mary, *Purity and Danger* (London: Routledge and Kegan Paul, 1969).
- and Isherwood, Baron, *The World of Goods* (London: Allen Lane, 1978).
- Dover, Kenneth J., *Greek Popular Morality in the Time of Plato and Aristotle* (Oxford: Blackwell, 1974).
- Dworkin, Ronald, 'What is Equality: Part I: Equality of Welfare'; 'Part II: Equality of Resources', *Philosophy and Public Affairs*, 10 (1981), 185–246; 283–345.
- Edgerton, Robert B., *Rules, Exceptions and Social Order* (Berkeley and Los Angeles: University of California Press, 1985).
- Ehrenberg, Margaret, *Women in Prehistory* (London: British Museum Press, 1989).

- Eisenman, Stephen, *Gauguin's Skirt* (New York: Thames and Hudson, 1997).
- Engels, Friedrich, *The Origins of the Family, Private Property and the State*, tr. Alec West (New York: International Publishers, 1975).
- Eriksson, Björn, *Heavy Duty* (Stockholm: Almqvist and Wickell, 1994).
- Evans-Pritchard, E. E., 'The Position of Women in Primitive Societies and in our Own', in *The Position of Women in Primitive Societies* (London: Faber and Faber, 1965), 37–58.
- Fisher, Helen, *The Anatomy of Love: The Natural History of Monogamy, Adultery and Divorce* (London and New York: Simon and Schuster, 1992).
- Flack, Jessica, and de Waal, Franz B. M., 'Any Animal Whatever: Darwinian Building Blocks of Morality', in Leonard D. Katz (ed.), *Evolutionary Origins of Morality: Cross-Disciplinary Perspectives* (Bowling Green: Imprint Academic, 2000), 67–78.
- Folbre, Nancy, 'Roemer's Market Socialism: A Feminist Critique', in Erik Olin Wright (ed.), *Equal Shares: Making Market Socialism Work* (London and New York: Verso, 1996), 57–70.
- Foucault, Michel, *A History of Sexuality*, vol. i, tr. Robert Hurley (New York: Vintage, 1988).
- Frazer, J. G., article 'Taboo', *Encyclopedia Britannica*, 9th edn. (New York: H. G. Allen, 1888), vol. T–P, p. 13.
- Freud, Sigmund, 'Taboo and the Ambivalence of Emotions', in *Basic Writings*, tr. and ed. A. A. Brill (New York: Modern Library, 1938), 821–64.
- Furrow, Dwight, *Against Theory* (New York: Routledge, 1995).
- Gauthier, David, *Morals by Agreement* (Oxford: Clarendon, 1986).
- Geach, Peter, 'Ascriptivism', *Philosophical Review*, 69 (1960), 221–5.
 — 'Assertion', *Philosophical Review*, 74 (1965), 449–65.
- Gehlen, Arnold, *Moral und Hypermoral*, 2nd edn. (Frankfurt and Bonn: Athenäum, 1970).
- Gibbard, Allan, *Wise Choices, Apt Feelings: A Theory of Normative Judgment* (Cambridge, Mass.: Harvard University Press, 1990).
- Gilligan, Carol, *In a Different Voice* (Cambridge, Mass.: Harvard University Press, 1982).
- Goodin, Robert, *Protecting the Vulnerable: A Reanalysis of our Social Responsibilities* (Chicago: University of Chicago Press, 1985).
 — 'Vulnerabilities and Responsibilities: An Ethical Defense of the Welfare State', in Brock (ed.), *Necessary Goods*, 73–94.
- Goody, Jack, 'Literacy and Moral Rationality', in Stent (ed.), *Morality as a Biological Phenomenon*, 153–66.

- Grice, H. P., 'The Causal Theory of Perception', *Proceedings of the Aristotelian Society*, suppl. vol. 35 (1961), 121–52.
- Griffin, James, *Well-Being: Its Meaning, Measurement and Importance* (Oxford: Clarendon, 1986).
- Hare, R. M., *The Language of Morals* (New York: Oxford University Press, 1964).
- *Sorting out Ethics* (Oxford: Clarendon, 1997).
- Harman, Gilbert, *The Nature of Morality* (Oxford: Oxford University Press, 1977).
- Harris, Marvin, 'The Evolution of Gender Hierarchies: A Trial Formulation', in Miller (ed.), *Sex and Gender Hierarchies*, 57–80.
- Harvey, J., *Civilized Oppression* (Totowa: Rowman and Littlefield, 1999).
- 'Justice Theory and Oppression', in Catherine Wilson (ed.), *Civilization and Oppression, Canadian Journal of Philosophy*, suppl. vol. 25 (Calgary: University of Calgary Press, 1999), 171–90.
- Hayek, Friedrich, *Law, Legislation and Liberty*, 2 vols. (London: Routledge and Kegan Paul, 1976).
- *The Mirage of Social Justice* (London: Routledge and Kegan Paul, 1976).
- *The Road to Serfdom* (Chicago: University of Chicago, 1994).
- Held, Virginia, *Rights and Goods* (New York: Macmillan, 1984).
- Herodotus, *The Histories*, tr. Robin Waterfield (Oxford: Oxford University Press, 1998).
- Hobbes, Thomas, *Leviathan*, vols. 1 and 2 (Indianapolis: Library of Liberal Arts, 1958).
- Honderich, Ted, 'The Use of the Basic Proposition of a Theory of Justice', *Mind*, 84 (1975), 63–78.
- Hrdy, Sarah Blaffer, *Mother Nature: Maternal Instincts and How They Shape the Human Species* (New York: Ballantine, 1999).
- *The Woman that Never Evolved* (Cambridge, Mass.: Harvard University Press, 1981).
- Hume, David, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, 2nd edn. (Oxford: Clarendon, 1978).
- *An Enquiry Concerning the Principles of Morals*, ed. Tom Beauchamp (Oxford: Oxford University Press, 1998).
- Irani, K. D., 'The Idea of Social Justice in the Ancient World', in Irani and Silver (eds.), *Social Justice in the Ancient World*, 3–8.
- and Silver (eds.), Morris, *Social Justice in the Ancient World* (Westport and London: Greenwood Press, 1995).

- Joshi, Heather, and Paci, Pierella, *Unequal Pay for Women and Men: Evidence from the British Cohort Study* (Cambridge, Mass.: MIT, 1998).
- Kagan, Shelly, *The Limits of Morality* (Oxford: Oxford University Press, 1989).
- Kant, Immanuel, *Anthropology from a Pragmatic Point of View*, tr. Mary J. Gregor (The Hague: Nijhoff, 1974).
- *Lectures on Ethics*, tr. Louis Infield (Indianapolis: Hackett, 1963).
- *Foundations of the Metaphysics of Morals*, in *Ethical Philosophy*, tr. J. W. Ellington (Indianapolis and Cambridge: Hackett, 1983).
- ‘On a Supposed Right to Lie from Altruistic Motives’, in Lewis White Beck (tr. and ed.), *Critique of Practical Reason and Other Writings in Moral Philosophy* (Chicago: University of Chicago Press, 1949), 346–51.
- Kidder, Louise H., Fagan, Michele A., and Cohn, Ellen S., ‘Giving and Receiving Social Justice in Close Relationships’, in Melvin J. Lerner and Sally C. Lerner (eds.), *The Justice Motive in Social Behaviour: Adapting to Times of Scarcity and Change* (New York and London: Plenum, 1981), 235–59.
- Koehler, Wolfgang, *The Mentality of Apes*, tr. Ella Winter (New York: Harcourt Brace, 1925).
- Korsgaard, Christine, ‘Skepticism about Practical Reason’, *Journal of Philosophy*, 83 (1986), 5–25.
- *The Sources of Normativity* (Salt Lake: University of Utah Press, 1994).
- Kummer, Hans, ‘Analogues of Morality among Nonhuman Primates’, in Stent (ed.), *Morality as a Biological Phenomenon*, 31–47.
- Lane, Robert J. ‘“The Road Not Taken,” Friendship, Consumption, and Happiness’, in David A. Crocker and Toby Linden (eds.), *The Ethics of Consumption* (Lanham, Md.: Rowman and Littlefield, 1998), 218–48.
- Lerner, Melvin J., *The Belief in a Just World: A Fundamental Delusion* (New York and London: Plenum, 1980).
- Lewis, Matthew, *The Monk*, ed. Howard Anderson (Oxford: Oxford University Press, 1995).
- Locke, John, *Of Civil Government: Second Treatise* (Chicago, Henry Regnery, 1955).
- Lucretius, Titus Carus, *On the Nature of Things*, tr. W. H. D. Rouse, rev. Martin Ferguson Smith (Cambridge, Mass.: Harvard University Press, 1975).
- Lux, Kenneth, *Adam Smith’s Mistake* (Boston and London: Shambhala 1990).
- Maccoby, Eleanor, and Jacklin, Carol, *The Psychology of Sex Differences* (London: Oxford University Press, 1975).

- McDowell, John, 'Values and Secondary Qualities', in Ted Honderich (ed.), *Morality and Objectivity: A Tribute to J. L. Mackie* (London: Routledge and Kegan Paul, 1985), 110–29.
- MacIntyre, Alasdair, *After Virtue* (Notre Dame: Notre Dame University Press, 1984).
- Mackie, J. L., *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin, 1977).
- Macleod, Colin, *Liberalism, Justice and Markets: A Critique of Liberal Equality* (Oxford: Clarendon, 1998).
- Maine, Henry Sumner, *Ancient Law* (Tucson: University of Arizona, 1986).
- Malthus, *An Essay on the Principles of Population* (1798; Amherst, NY: Prometheus, 1998).
- Mannheim, Karl, *Ideology and Utopia*, tr. L. Worth and E. Shils (New York: Harcourt Brace, 1936).
- Marsden, William, *The History of Sumatra*, 3rd edn. (1783, Kuala Lumpur: Oxford University Press, 1966).
- Matthen, Mohan, and Ariew, André, 'Two Ways of Thinking about Fitness and Natural Selection', *Journal of Philosophy*, 99 (2002), 55–83.
- Maynard Smith, John, *Evolution and the Theory of Games* (Cambridge: Cambridge University Press, 1982).
- Merton, Robert K., 'The Matthew Effect in Science', *Science*, 159 (1968), 56–63.
- Mill, John Stuart, *The Collected Works of John Stuart Mill*, ed. John M. Robison, 33 vols. (Toronto: University of Toronto, 1962–91).
- Miller, Barbara Diane (ed.), *Sex and Gender Hierarchies* (Cambridge: Cambridge University Press, 1993).
- Morschauser, Scott N., 'The Ideological Basis for Social Justice/Responsibility in Ancient Egypt', in Irani and Silver (eds.), *Social Justice*, 101–14.
- Murphy, Liam, *Moral Demands in Nonideal Theory* (New York: Oxford University Press, 2000).
- Nagel, Thomas, 'Critical Notice of Rawls's A Theory of Justice', *Philosophical Review*, 82 (1973), 220–34.
- 'Equal Treatment and Compensatory Discrimination', *Philosophy and Public Affairs*, 2 (1973), 348–63.
- *Equality and Partiality* (New York: Oxford University Press, 1991).
- 'Libertarianism without Foundations', in J. Paul (ed.), *Reading Nozick* (Totowa: Rowman and Littlefield, 1981).
- *Mortal Questions* (Cambridge: Cambridge University Press, 1979).
- *The Possibility of Altruism* (New York: Oxford University Press, 1970).

- Nagel, Thomas, *The View from Nowhere* (New York: Oxford University Press, 1986).
- Nielsen, Kai, and Ware, Robert (eds.), *Exploitation* (Atlantic Highlands: Humanities Press, 1997).
- Nietzsche, Friedrich, *The Genealogy of Morals*, tr. Walter Kaufmann and R. J. Hollingdale (New York: Random House, 1969).
- Nisbett, Richard, and Ross, Lee, *Human Inference: Strategies and Shortcomings of Social Judgement* (Englewood Cliffs: Prentice-Hall, 1980).
- Nozick, Robert, *Anarchy, State and Utopia* (New York: Basic Books, 1974).
- Nussbaum, Martha, 'Human Functioning and Social Justice: In Defense of Aristotelian Essentialism', *Political Theory*, 20 (1992), 202–46.
- *Love's Knowledge* (London: Oxford University Press, 1990).
- and Sen, Amartya (eds.), *The Quality of Life* (London: Oxford University Press, 1993).
- Oakley, Anne, *Housewife* (London: Allen Lane, 1974).
- Okin, Susan Moller, *Justice, Gender and the Family* (HarperCollins: Basic Books, 1989).
- Omarck, Donald R., Strayer, F. F., and Freeman, David G. (eds.), *Dominance Relations: An Ethological View of Human Conflict and Social Interaction* (Hamden: Garland, 1980).
- O'Neill, Onora, 'Justice, Gender and International Boundaries', in Nussbaum and Sen (eds.), *Quality of Life*, 303–23.
- Pietroski, Paul M., 'Prima Facie Obligations, Ceteris Paribus Laws in Moral Theory', *Ethics*, 103 (1993), 489–515.
- Putnam, Hilary, 'Objectivity and the Science-Ethics Distinction', in Nussbaum and Sen (eds.), *Quality of Life*, 143–57.
- Radcliffe-Brown, A. R., *Taboo* (Cambridge: Cambridge University Press, 1939).
- Radin, Paul, *Primitive Man as Philosopher* (New York: Appleton, 1927).
- Railton, Peter, 'Moral Realism', *Philosophical Review*, 95 (1986), 163–207.
- Rawls, John, *The Law of Peoples* (Cambridge, Mass.: Harvard University Press, 1999).
- *Political Liberalism* (New York: Columbia University Press, 1993).
- *A Theory of Justice* (rev. edn., Cambridge, Mass.: Harvard University Press, 1971).
- Rheingold, H. L., and Hay, D. F., 'Prosocial Behaviour of the Very Young', in Stent (ed.), *Morality as a Biological Phenomenon*, 93–108.
- Richards, Robert, 'A Defense of Evolutionary Ethics', *Biology and Philosophy*, 1 (1986), 265–93.

- Roemer, John, *Egalitarian Perspectives* (Cambridge: Cambridge University Press, 1994).
- *Theories of Distributive Justice* (Cambridge, Mass.: Harvard University Press, 1996).
- Rorty, Richard, 'Why Can't a Man Be More Like a Woman? and Other Problems in Moral Philosophy' (Review of Annette Baier, *Moral Prejudice*), *London Review of Books*, 24 Feb. 1994, 3–6.
- Ross, W. D., *The Right and the Good* (Oxford: Clarendon, 1930).
- Runciman, W. G., *Relative Deprivation and Social Justice: A Study of Attitudes to Social Inequality in Twentieth Century England* (Berkeley and Los Angeles: University of California Press, 1966).
- Ruse, Michael, 'Evolutionary Ethics: Healthy Prospect or Last Infirmity?', in Mohan Matthen and Bernard Linsky (eds.), *Philosophy and Biology, Canadian Journal of Philosophy*, suppl. vol. 19 (Calgary: University of Calgary Press, 1988), 27–74.
- *Taking Darwin Seriously*, 2nd edn. (Amherst, NY: Prometheus, 1998).
- Santangelo, Antonio, *The Beginning and Meaning of Culture: The Cerebral Activity Underlying It* (Milan: La Pietra, 1993).
- Scanlon, T. M., 'Preference and Urgency', *Journal of Philosophy*, 72 (1975), 655–69.
- *What We Owe to Each Other* (Cambridge, Mass.: Harvard University Press, 1998).
- Scheffler, Samuel, *Boundaries and Allegiances: Problems of Justice and Responsibility in Liberal Thought* (Oxford: Oxford University Press, 2001).
- *Human Morality* (New York: Oxford University Press, 1992).
- *The Rejection of Consequentialism*, 2nd edn. (Oxford: Oxford University Press, 1994).
- Schein, Martin (ed.), *Social Hierarchy and Dominance* (Stroudsburg: Dowden, Hutchinson and Ross, 1975).
- Schumacher, E. F., *Small is Beautiful* (London: Abacus, 1974).
- Scott, John, *Poverty and Wealth: Citizenship, Deprivation and Privilege* (London and New York: Longman, 1994).
- Scruton, Roger, 'Contract, Consent and Exploitation: Kantian Themes', in Howard Williams (ed.), *Essays on Kant's Political Philosophy* (Cardiff: University of Wales Press, 1992), 213–27.
- Sen, Amartya, 'Equality of What?', in *Inequality Reexamined* (Cambridge, Mass.: Harvard University Press, 1992), 12–30.
- and Williams, Bernard (eds.), *Utilitarianism and Beyond* (Cambridge: Cambridge University Press, 1982).

- Shepher, Joseph, and Tiger, Lionel, 'Female Hierarchies in a Kibbutz Community', in Tiger and Fowler, (eds.), *Female Hierarchies*, 225-47.
- Sher, George, *Desert* (Princeton: Princeton University Press, 1987).
- 'Effort, Ability, and Personal Desert', *Philosophy and Public Affairs*, 8 (1979), 361-76.
- Silk, Joan B., 'Primatological Perspectives on Gender Hierarchies', in Barbara Diane Miller (ed.), *Sex and Gender Hierarchies*, 212-35.
- Singer, Peter, 'Famine, Affluence, and Morality', *Philosophy and Public Affairs*, 1 (1972), 229-43.
- Skyrms, Brian, *The Evolution of the Social Contract* (Cambridge: Cambridge University Press, 1996).
- Slote, Michael, *Common-Sense Morality and Consequentialism* (London: Routledge and Kegan Paul, 1985).
- Small, Meredith, *Female Choices* (Ithaca, NY: Cornell University Press, 1993).
- Smart, J. J. C., and Williams, Bernard (eds.), *Utilitarianism: For and Against* (Cambridge: Cambridge University Press, 1973).
- Smeeding, Timothy M, and Gottschalk, Peter, 'Cross-National Income Inequality: How Great Is It and What Can We Learn from It?', *Focus*, 19 (1998), 15-19.
- Smuts, Barbara, *Sex and Friendship in Baboons* (New York: Aldine, 1985).
- Sober, Elliott, and Wilson, David Sloan, *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Cambridge, Mass.: Harvard University Press, 1998).
- Stent, Gunther (ed.), *Morality as a Biological Phenomenon* (Berkeley and Los Angeles: University of California Press, 1980).
- Stokes, Eric, *The British Utilitarians and India* (Delhi: Oxford University Press, 1982).
- Strawson, P. F., 'Freedom and Resentment', in *Studies in the Philosophy of Thought and Action* (London: Oxford University Press, 1968), 71-96.
- *Individuals* (London: Methuen, 1959).
- Sturgeon, Nicholas, 'Moral Explanations', in David Copp and David Zimmerman (eds.), *Morality, Reason and Truth* (Totowa: Rowman and Allanheld, 1984), 49-78.
- Sweetman, David, *Paul Gauguin: A Complete Life* (London: Hodder and Stoughton, 1995).
- Tiger, Lionel, and Fowler, Heather T. (eds.), *Female Hierarchies* (Chicago: Beresford Books Service, 1978).
- Treffert, Darold, 'The Idiot Savant: A Review of the Syndrome', *American Journal of Psychiatry*, 145 (1988), 563-72.

- Trivers, Robert, 'The Evolution of Reciprocal Altruism', *Quarterly Review of Biology*, 46 (1971), 35–57.
- *Social Evolution* (Menlo Park: Benjamin Cummings, 1985).
- Unger, Peter *Living High and Letting Die: Our Illusion of Innocence* (Oxford: Oxford University Press, 1996).
- Valian, Virginia, *Why So Slow?* (Cambridge, Mass.: MIT Press, 1999).
- van Sommers, Peter, *The Biology of Behaviour* (Sydney, NY: Wiley, 1972).
- Walton, Kendall, 'How Remote are Fictional Worlds from the Real World?', *Journal of Aesthetics and Art Criticism*, 37 (1978), 11–23.
- Walzer, Michael, 'Objectivity and Social Meaning', in Nussbaum and Sen (eds.), *Quality of Life*, 165–77.
- *Spheres of Justice* (New York: Basic Books, 1983).
- Westermarck, Edward, *Ethical Relativity* (London: Kegan Paul, Trench and Trubner, 1932).
- Wiggins, David, 'A Sensible Subjectivism', in *Needs, Values, Truth*, 3rd edn. (Oxford: Clarendon Press, 1998).
- Williams, Bernard, 'A Critique of Utilitarianism', in Smart and Williams (eds.), *Utilitarianism*, 77–150.
- *Ethics and the Limits of Philosophy* (London: Fontana, 1985).
- 'Internal and External Reasons', in *Moral Luck*, 101–13.
- *Moral Luck: Philosophical Papers 1973–1980* (Cambridge: Cambridge University Press, 1981).
- *Problems of the Self: Collected Papers 1956–1971* (Cambridge: Cambridge University Press, 1973).
- Wilson, Catherine, 'Natural Dominance: A Reply to Michael Levin', *Philosophy*, 73 (1998), 572–92.
- 'On Some Alleged Limitations to Moral Endeavor', *Journal of Philosophy*, 90 (1993), 275–89.
- 'Prospects for Noncognitivism', *Inquiry*, 44 (2001), 1–24.
- Wilson, E. O., *On Human Nature* (Cambridge, Mass.: Harvard University Press, 1978).
- Wolf, Susan, 'Moral Saints', *Journal of Philosophy*, 79 (1982), 419–39.
- Wood, Allen W., 'Exploitation', in Nielsen and Ware (eds.), *Exploitation*, 2–26.
- Wrangham, Richard, and Peterson, Dale, *Demonic Males: Apes and the Origins of Human Violence* (London: Bloomsbury, 1996).
- Wright, Patricia Chapple, 'Variations in Male–Female Dominance and Offspring Care in Non-human Primates', in Miller (ed.), *Sex and Gender Hierarchies*, 127–45.

- Young, Robert, 'Egalitarianism and Personal Desert', *Ethics*, 102 (1992), 319–41.
- Zweig, Arnulf (ed.), *Kant: Philosophical Correspondence 1759–1799* (Chicago: University of Chicago Press, 1967).

Index

- advantage, accumulation of, 31,
72, 175–8, 188, 274–7, 210, 215,
283, 299; natural *v.* acquired, xi,
184; -reduction, morality and,
vii, 9–13, 21, 23–4, 29–31, 34,
72, 120–1, 124–5, 159, 166,
169–170, 172–3, 277–8, 283,
289–90; staggered policies of,
23–8
- affluence, 79, 84, 137, 144–6, 150,
193, 197, 246
- animals, proto-aesthetics and proto-
morality of, 1–4, 301. *See also* apes
- Annas, Julia, 57
- anonymity requirement, x, 124,
126–39, 143, 195, 278
- apes, 2–3, 20–1, 270–1, 292–3
- Aristotle, 29, 45, 117, 135, 196, 254,
261–2
- Axelrod, Robert, 201–2
- Beauvoir, Simone de, 11, 112, 254,
257
- Before the Revolution Principle,
88, 92, 277, 297
- belief, moral, 35–7, 41
- benevolence, 115–17, 142–7, 157–8,
185–7, 200, 281
- Bentham, Jeremy, 50, 75, 197
- Braybrooke, David, 9, 28
- Chance, Michael, R. A., 291–2
- Cohen, Gerald, 233
- confirmation, ix–x, 47–56, 60–2,
64–9, 91–2, 125
- consequentialism, 160, world level
v. theory level 67–8. *See also*
utilitarianism
- Contract Principle, 154–5, 158,
198, 209, 283, 296
- contractualism, 26, 138–42, 195–9
- co-operation, xi, 17, 21, 179–89,
209–10, 296–7, 299; basic *v.*
symmetrical 181–4, 194
- co-operative surplus, x, 21, 164,
171, 188, 257, 283, 298
- Counterweight Principle, 153;
applications of, x, 153–8. *See also*
Contract Principle Culpable
Ignorance Principle, Ill-Gotten
Gains Principle
- Crusoe₁ and Crusoe₂, 219–24
- Culpable Ignorance Principle, 154,
156, 159, 198, 209, 283, 287
- culpable letting happen, 169, 189,
224, 253, 279
- Dasgupta, Partha, 146, 192
- deprivation, xi, 164–5, 169–70,
189–90, 192–3, 200–2, 209,
242, 247, 251, 256–7, 260–1,
264, 300
- disqualification, 104, 113, 160,
254–5; thesis external, 105–11,
134; thesis internal, 111–13,
119, 203–4
- distributive justice, 166–79, 188,
202, 212–55
- distribution protocols, 144–5, 197,
212–34

- dominance, xi, 19–21, 123, 170,
184–5, 253, 270, 277, 286, 291, 293
duties of presence, 152–3, 293
- egalitarianism, vii, 93–4, 126, 163,
169, 189–91, 194, 209–11, 245–6.
See also equality
- Ehrenberg, Margaret, 165
- emotivism, 50, 61
- Engels, Friedrich, 298
- equality, x–xii, 20, 31, 166, 189–90,
194–5, 254, 256, 259, 262–4,
277–83, 285, 290–1
- Eriksson, Björn, 86–7, 138
- Evans-Pritchard, E. E., 108–10
- Evolution, 17–20, 24–5, 33, 184,
255, 283
- evolutionary psychology, 142,
266–74, 276, 288, 294
- exploitation 135, 186–8, 202, 214,
255, 283, 289, 298
- exigency, viii–x, 61, 70–7, 83–4, 90,
149, 152–3, 159, 163, 246, 278,
300. *See also* heavy costs,
argument from
- Externally-imposed Hardship
Principle, 88, 92, 199, 277
- first-person perspective, viii, 77–85,
93, 110, 124, 127–8, 130, 133, 143,
162–3, 167–9, 196, 277–8, 280
- Foucault, Michel, 267
- Freud, Sigmund, v, 32, 34, 49,
111–12
- “Gauguin,” moral fable of, 49, 82–3,
101, 123, 134, 156, 168, 294
- Gauthier, David, 20, 167, 172,
219–20, 223
- God, as equalizer, 190–1, 207
- Harman, Gilbert, 105–6
- Hayek, F. A., 212, 218, 220, 225,
236, 240
- heavy costs, argument from, x,
xii, 85–9, 90–1, 147–9, 152,
157, 166, 198, 206, 277–83,
288; Principle, 86, 93–4
- helping, *see* benevolence
- Hitler, Adolf, 54; and
disqualification, 106–8; and
moral realism, 41
- Hobbes, Thomas, 17–18, 21, 24,
30–1, 209
- Hobbes’ Theorem, 17, 33, 71
- Hrdy, Sarah, 276
- Hume, David, 4, 30, 45, 79,
95–7, 119–22, 241, 225–6,
266
- hypermorality, 28–9, 35, 71,
170–1, 169, 173
- idealism characteristic, 49, 53, 58,
218, 288
- ideology, 64, 113, 124–5; *v.* moral
theory, x, 126, 163
- Ill-Gotten Gains Principle,
154–5, 158–9, 198, 209,
282
- immanentism, 95, 97–104, 133,
163, 191, 202–8, 211, 266.
See also moral Pyrrhonism
- imperatives, 7–9, 38–40, 64, 107,
293. *See also* rules
- judgements, moral, vii, 7, 9, 35,
40, 51–2, 61, 107
- justice, procedural *v.* outcome
based theories of, xii, 166,
174–9, 189, 195, 210–11, 236.
See also distributive justice

- Kant, Immanuel, Kantianism, vii, viii, 4, 15, 34, 37, 44–8, 60, 72–6, 78–9, 92, 96, 124, 151–2, 203, 254, 262–3, 293
- Koehler, Wolfgang, 301
- Kummer, Hans, 22, 23–4
- Locke, John, 174–5, 219, 221, 227
- love, xii, 150, 200, 293–301
- luck, *see* risk
- Lucretius, Titus Carus, 16–17, 21
- MacIntyre, Alasdair, 57–8, 98, 295
- Mannheim, Karl 125
- marriage, 6, 101, 296–301
- Marx, Karl, 101, 295
- maternity, 260, 265–6, 269, 271, 291; disqualifications of 272, 283, 285, 299
- Maynard Smith, John, 18
- McLeod, Colin, 59
- merit, xi–xii, 207–9, 211, 213–42; detecting/ascribing, 232–3, 237–42, 244, 248–9, 252–3, 283–6
- merit-desert intuition, 216–17, 220–2, 224–8, 235, 237, 243–6
- Merit-priority, 212, 217–18, 221–3, 225, 228–30, 242
- Merit Principle, 213 ff
- Mill, John Stuart, 11–12
- moral theory, *see* theory, moral
- motivation, moral, 21–2, 74, 90, 123, 151, 222
- Nagel, Thomas, 73, 79, 85, 166–7, 173, 195–8, 206, 234, 244, 279–80
- needs, 146, 213–14. *See also* deprivation
- Need-priority, 212, 243–4, 252
- Need Principle, 220–2, 229, 242–3
- Nontheoretical Reasons Principle, 99, 122
- norms, 3, 5–9, 12–3, 28, 31–4, 56, 98, 213; authored *v.* unauthored, 38–9, 42–3, 62, 70, 89, 91–2, 127, 142, 151, 161, 163; bindingness of, 62, 70, 90, 125. *See also* rules
- Nozick, Robert, 172, 175, 177, 219, 220
- Nussbaum, Martha, 57
- objectivity, 73–4, 77, 79–80, 125. *See also* realism, moral
- opacity, 108, 110, 113–16, 124
- paraworlds, ix–x, 36, 48–53, 57, 60–1, 96, 101, 113–4, 128, 197, 289
- partiality, 79–82, 142–5
- Plato, 29, 101, 293
- prostitution, *see* vice
- Putnam, Hilary, 111–12, 118, 123
- Pyrrhonism, moral, 99–100, 119, 203, 207–8
- Rawls, John, 43, 135, 138–9, 172–3, 189, 214–15, 217, 234–5, 241, 246, 251, 281
- realism, moral, ix, 38–42, 55–6, 65–6
- reality constraint, 49, 53, 57–9, 68, 113, 117, 218, 288
- Reds and Greens, 135, 251–2
- redistribution protocols, 132–3, 197–202, 205, 208, 234, 248, 281–2

- Relative Deprivation Principle, 88, 92, 137, 158, 199, 277
- relativism, x, 131, 161; theory-level *v.* world-level, 62–4
- rights, 136–7, 257, 262, 298
- risk, 59–60, 102, 117, 164–5, 168, 175–6, 193, 248, 250–2, 256, 264
- Radcliffe-Brown, A. R., 32
- Rousseau, Jean-Jacques, 21, 184–5, 254
- rules, 6–10, 13–14, 16, 21–2, 31–3, 35, 37, 71, 96
- Runciman, W. D., 200
- Scanlon, Thomas, 139–40, 192, 196
- Scheffler, Samuel, 81, 131, 217, 228–9
- Scruton, Roger, 203–4
- Singer, Peter, 91, 158–9
- Sittlichkeit*, 101–2, 280
- Skyrms, Brian, 185
- slavery, 135–6, 108, 179, 199, 204–5, 208
- sociobiology, *see* evolution, evolutionary psychology
- Strawson, P. F., 80–1, 124
- subordination of women, xii, 58, 60, 113–19, 165, 254–61, 266–77
- suicide, 132, 148, 152
- taboo, vii, 8, 13, 15, 32, 124; substances, 200, 259
- theory, moral 36–8, 43–6, 52–6, 93, 96, 119, 123, 126–7; and practice, 147–51; as burdensome, viii–ix, 76–7, 92–4, 96–7, 159–60, 162; as intrusive, 104–5, 113, 133–4; as useless, 76–7, 98–100, 112, 137, 202; information and, 68–9, 139, 159–60, 162; local *v.* critical, 37–8; *v.* scientific theory, 44–8, 61, 122–3. *See also* confirmation, ideology
- Trivers, Robert, 19, 25.
- Unger, Peter, 158–9
- universalism, 22, 62. *See also* relativism
- utilitarianism, ix, 44, 75, 78, 80–1, 92, 128, 163, 199, 278. *See also* consequentialism
- vice, 227, 294–5, 298–9.
- virtue, virtue theory, 12–13, 95–6, 120–2, 126
- Walzer, Michael, 113–19, 203, 261
- well-being, xi, 154, 191–4, 201, 224, 243–6, 252, 254, 257, 260, 298. *See also* deprivation
- Westermarck, Edward, 22
- Williams, Bernard, ix, 75–7, 82–4, 97, 101–2, 111, 134, 151, 159, 160–1, 163, 279–80, 293–4, 301
- Wilson, E. O., 206, 258, 274, 279, 290, 294, 296
- Wolf, Susan, 84