RATIONAL DISAGREEMENT ABOUT SOCIAL JUSTICE

by

KURT PREINSPERG

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Department of Philosophy

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature)

Department of _Philosophy_

The University of British Columbia
Vancouver, Canada

Date _23rd September 1992_

ABSTRACT

Numerous conflicting theories about the just distribution of goods in society have emerged in response to John Rawls' 1971 treatise *A Theory of Justice*. Given that informed persons continue to disagee strongly about the demands of social justice, progress may come from better understanding the underlying reasons for disagreement about justice among philosophically informed people of moral goodwill.

In this dissertation I explore the idea, suggested by Larmore, Pogge, Ackerman and others, that *some* of the disagreement about justice among informed people of goodwill is "rational" disagreement. I identify, in the literature on social justice, recurrent conceptual, normative and empirical issues which we have reason to consider currently (or conceivably forever) irresolvable, at the first-order level of moral reasoning, by any information, arguments or methods accessible to us.

But claims concerning the possibility of rational disagreement about justice only differ non-trivially from skepticism about justice if plausible limits can be set to the scope of this disagreement. To characterize such limits, I seek to establish the following two wide-reflective-equilibrium-based presumptions. (1) A consequentialist metaethical framework is our most credible approach to moral justification (where the form of consequentialism defended is constructivist, non-foundational, value-pluralistic, and includes distribution-sensitivity among its ultimate values). (2) In moderately well-off, pluralistic societies, only those conceptions of justice fall within the scope of rational disagreement which propose broadly egalitarian-liberal, "directly responsive" principles (i.e. principles applicable to individual or group shares and not merely to basic social structure).

Some likely candidates for the status of rational disagreements about social justice are discussed: the criteria definition and inclusion problems; various balancing problems related to attempts to increase the comprehensiveness of principles (the priority problem, the aggregation-distribution problem and commensuration problems); the domain demarcation problem; and problems of imprecision associated with justifying claims about justice within a consequentialist framework.

An improved understanding of major sources of rational disagreement about social justice, as presented in this dissertation, helps define the normative weight of appeals to justice. This in turn clarifies the need to resolve many issues of social distribution otherwise than by relying on invocations of justice.

CONTENTS

LIST OF TABLES

ACKNOWLEDGMENTS

I feel deep gratitude toward my adopted country Canada for allowing my lifetime dream of studying philosophy to come true. I owe the University of British Columbia many years of joy and growth in an atmosphere of intellectual freedom.

I owe thanks to my Committee. Earl Winkler took on my supervision at a hectic stage in his own career and encouraged my fledgling thesis efforts up until his sabbatical year. Michael McDonald kindly agreed to take over as thesis supervisor, in spite of the crushing demands of his role as Director of UBC's new Centre for Applied Ethics. He gave me reams of valuable comments which challenged me to think harder and dig deeper.

Gary Wedeking and Peter Danielson invested hard work in my project, and I sincerely thank them for seeing it to completion. Thanks are also due to Jim Dybikowski and Bob Modrow for serving as my University Examiners, and to Wayne Sumner of the University of Toronto for serving as my External Examiner.

I also want to thank Bob Rowan, my M.A. thesis supervisor, whose passionate teaching of Plato, Hobbes and Rousseau originally got me excited about political philosophy. In his unforgettable words, "Yes, Plato can be shot down – but not with your guns!"

And I wish to acknowledge some special friends for their affirming support in sometimes difficult times: Dennis Bibby, Alice Gmuer, Marcia Monreal, Margaret Nicholson, Randy Reiffer, Antonia Rozario, Makiko Suzuki-Plimley.

INTRODUCTION

Inescapable questions confront human communities about how to divide benefits and burdens among their members. *Justice* is commonly invoked as a decisive constraint on social distributions of valued things. But there are fierce disagreements about justice which often provoke festering antagonisms and violent confrontations, even among philosophically informed people of moral goodwill.

There has been an upsurge of theorizing about justice thanks to the publication of John Rawls' *A Theory of Justice* in 1971. Rawls deserves praise for reopening the question of justice at a sophisticated level of argument. It is a striking fact, however, that Rawls' work has not so much stimulated convergence toward philosophical consensus about the just distribution of goods in society, as a proliferation of rival theories.[1]

In response to this fact, my dissertation seeks to develop the concept of "rational disagreement", in the hope of overcoming certain forms of conflict which are standardly couched in terms of justice. The concept of rational disagreement, though not yet prominent in justice theorizing, is not original with me. Some - for example Larmore - use it without, apparently, feeling much need to clarify it.[2] Others use more

---

[1] A very partial sampling of post-Rawlsian theories of social justice: Ackerman, Bruce A., *Social Justice in the Liberal State*, 1980; Ewin, R.E., *Liberty, Community, and Justice*, 1987; Galston, William A., *Justice and the Human Good*, 1980; Gauthier, David, *Morals by Agreement*, 1986; Heller, Agnes, *Beyond Justice*, 1987; Narveson, Jan, *The Libertarian Idea*, 1988; Nielsen, Kai, *Equality and Liberty: A Defense of Radical Egalitarianism*, 1985; Nozick, Robert, *Anarchy, State, and Utopia*,1974; Phillips, Derek L., *Toward a Just Social Order*, 1986; Reiman, Jeffrey, *Justice and Modern Moral Philosophy*, 1990; Soltan, Karol Edward, *The Causal Theory of Justice*, 1987; Sterba, James P., *The Demands of Justice*, 1980; Walzer, Michael, *Spheres of Justice: A Defense of Pluralism and Equality*, 1983; Winfield, Richard Dien, *Reason and Justice*, 1988.

[2] Larmore uses the concept in the following context: "If we have it in our power to satisfy the needs of others or to prevent their having physical pain, and if the good thereby effected is sufficiently great, we may well feel obligated to set aside temporarily the pursuit of our own projects. And even if we do not conclude that the greater good overall should prove decisive, we ought still to feel regret, or offer some explanation, or perhaps make some amends. I am disinclined to believe that there is any illuminating general rule to decide such cases (How great should the greater good overall be? How much does it matter how it is distributed among others?). Some cases may be decidable by judgment; others must be the *object of rational disagreement.*"(141-2; emphasis added).

or less closely related concepts. Pogge, for example, speaks of "reasonable disagreement";[3] Ackerman speaks of "good-faith disagreement";[4] Benjamin speaks of "the ineliminability of rationally irreconcilable moral conflict";[5] and Hardin speaks of "the limits of reason".[6] There is nothing very mysterious about the underlying idea of such talk. Every remotely plausible approach to moral decision-making will have to admit limits both to its precision and to its power to decide hard cases; and it is natural to speak of "room for disagreement" in such contexts. But in spite of its central importance, the idea of rational disagreement seems to me to have been left surprisingly unclear in recent philosophical works on justice.

I can attempt only part of this clarifying task here. My aims are essentially threefold: to give some sense of the complexities that arise from an attempt to apply the concept of rational disagreement to disputes about social justice; to identify some plausible sources of rational disagreement; and to reach a provisional verdict about some of the limits to rational disagreement.

This inquiry is organized into three main parts. Part I seeks to motivate talk about "rational disagreement" by sketching the need for such a concept in the context of contemporary justice theorizing and by offering a viable working conception of this

---

[3] Pogge's use of this concept is especially illuminating: "What is needed is the recognition that knowledgeable and intelligent persons of good will may reasonably disagree about the fundamental issues dividing the world today. For example, should the means of production be controlled by national governments, or locally by workers or by private owners? Is the best forum for democratic discussion and decision making afforded by a single-party, two-party, or multi-party system? Which is more important in the appraisal and reform of social institutions, the protection of civil and political liberties or the satisfaction of basic social and economic needs? If only we could understand our disagreements about such matters as *reasonable disagreements*, then we could jointly work toward a world in which alternative answers to these questions could coexist in a peaceful, friendly, and supportive international environment."(232; emphasis added)

[4] Ackerman says this: "And so we come to a first difficulty of second-best government. What decision rule should be used when liberal statesmen disagree about the best way to compromise the liberal ideal of undominated equality? Call this the problem of *good-faith disagreement*."(274; emphasis added)

[5] Benjamin thinks that "An appreciation of the ineliminability of rationally irreconcilable moral conflict will accord greater importance to the notion of compromise than one generally finds among ethical theories."(2)

[6] Limits of reason, according to Hardin, "are the usual cognitive and structural limits to good decision-making.... These include limits on mental ability, limits on time available for deciding, limits on information, and limits on relevant theory. Anyother major limit of moral reason that is not usually discussed in this context is the inability of a single individual to determine an outcome independently of the actions of others. This is the problem of strategic interaction..."(xvii).

concept. Part II and III advance and defend the following theses respectively:

1. Setting *non-trivial* limits to rational disagreement about justice will require metaethical commitments. (I will sketch a wide reflective equilibrium-based rationale for choosing a form of consequentialism over rival theories – but obviously without being able to give anything approaching a full defense here).

2. Once we accept a broadly consequentialist framework for moral thinking, then there arise a number of recurrent conceptual, normative and empirical issues which constitute plausible sources of rational disagreement about justice. (Many of these issues will, however, arise *mutatis mutandis* within certain non-consequentialist frameworks as well.)

I launch my search for plausible sources of rational disagreement about justice by examining two central problems arising from the question of what sorts of considerations are fundamentally relevant to determining justice. These problems are labelled the criteria definition problem and the criteria inclusion problem (Ch 3.1). Next I explore two closely connected sources of disagreement about principles of justice: disagreement about the appropriate level of comprehensiveness of such principles (Ch 3.2), and about the structure of their justification within a consequentialist framework (Ch 3.3). Four recurrent dilemmas facing theories of justice are briefly characterized: the priority problem, the aggregation-distribution problem, the commensuration problem, and the domain demarcation problem. Finally I try to assess the possibility of rational disagreement with respect to two of the most prominent disputes about justice: the dispute between liberal and non-liberal conceptions of justice (Ch 3.4), and the dispute between libertarian and egalitarian conceptions of liberalism (Ch 3.5).

I conclude that there is a presumption that only "directly responsive" principles of justice of a liberal-egalitarian type fall within the scope of rational disagreement (assuming a pluralist society in at least minimally favorable economic circumstances).[7]

---

[7] Principles of justice will be labelled *indirectly responsive* to the existing pattern of sentiment and

Disagreement about competing principles of this kind will generally have to be settled in the light of fairly context-specific consequentialist considerations rather than at the level of fundamental moral theory. But even context-specific considerations will frequently leave room for rational disagreement between mutually exclusive principles or decisions. Such disagreement must then be resolved at the second-order level of moral reasoning, the level of political compromise-seeking and convention-setting – ideally, *in clear recognition of the scope for rational first-order disagreement.*

Three cautionary remarks are in order. First, I cannot possibly defend the interconnected strands of my account all at once, and must often defer discussions to later chapters whose results are presupposed in earlier ones. Since a meaningful effort to think clearly about justice requires complex considerations to be kept in mind simultaneously, there is no ideal order of presentation which can resemble the linearity of a simple argument. Only a view of the whole can illuminate each part.

Second, I must presuppose acquaintance with contemporary theories of justice, most importantly John Rawls' and David Gauthier's work. There is simply no feasible way for me to restate their work sufficiently fully here to prepare the ground for my criticisms. I shall often refer to places in the literature where crucial arguments can be looked up whenever rehearsing them would lead me too far afield.

Third, my interest here is *not* in giving a concise summary and critique of the vast contemporary literature on justice, or even a history of key disputes that seem to go on without resolution. Parts of this immense task have been ably accomplished by a number of writers.[8] I shall assume that the empirical fact of deep philosophical disagreement about justice is not in doubt, and focus on my chosen subject of examining the possibility that at least some of this disagreement may be *rational* in the

---

expectations about justice in a society if they specify a just system of basic social institutions organized as a whole (e.g. Rawls' approach). By contrast, principles of justice will be labelled *directly responsive* to this pattern if they specify particular characteristics which individuals (or groups) must have for justly claiming a particular share of social resources. (See Sec 3.1.3 §3 below.)

[8] See e.g. Philip Pettit, *Judging Justice* (1980); Will Kymlicka, *Contemporary Political Philosophy* (1990); or Tom Campbell, *Justice* (1988) for admirable restatements and criticisms of prominent theories of justice.

relevant sense.

The dilemma which this sort of synoptic inquiry must face is that many issues are both centrally relevant to it and yet far too complex to pursue to a desirable degree of depth. I cannot follow many contemporary writers on justice into what I judge to be ultimately blind alleys, but must be highly selective with respect to the issues I engage. All I can hope to do here, on many issues, is to establish basic bearings. My project is motivated by a fundamental conviction, however, that finding our basic bearings in the welter of conflicting ideologies is more important than to erect blinkers around some little corner of it. The scope of this inquiry is broad; but when it comes to contemporary justice theorizing, I believe that a broad view is required to see any light at all.

PART 1: THE CONCEPT OF RATIONAL DISAGREEMENT

This part seeks to motivate the concept of rational disagreement in the context of contemporary justice theorizing; to develop a viable working conception of this concept; and to confront several serious objections to the project of identifying sources of rational disagreement about justice.

1. 1 THE PHILOSOPHICAL CONTEXT

§1 What sense can we make of the vast philosophical literature on justice? How much authoritative clarity about the demands of justice can decision makers gain from it? Are additions to the already crowded field of competing theories of justice likely to resolve conflicts about justice or more likely to exacerbate them? What sort of philosophical endeavor is now most likely to advance this field? These are my motivating concerns.

Two kinds of endeavor seem obviously worthwhile. The first would be to offer an adequate theory about the just distribution of goods in society. The second is to point out inadequacies in various proposed theories. What are the prospects for success in either of these endeavors?

It is a monumental challenge to offer an adequate theory of social justice: a theory, which no informed person of good will could reasonably reject, for a set of usefully determinate principles to decide the justice of major social distributions. Taking on *this* challenge is easily confused with a far less difficult one - that of proposing new principles of social distribution and defending them as *no less* permitted by reason than those we already have. My own initial desire was, in fact, to defend some principle as the following:

> *Distribute economic goods in such a way as to ensure the most modest level of material prosperity which is consistent with good health, and which has the psychological effect of minimizing preoccupation with material acquisition while at the same time maximizing opportunities for experiencing the beauty of unspoilt nature, the delights of art and scientific discovery and warm human relationships to the utmost compatible degree.*

I came to realize, however, that there was little hope of justifying my favored social vision as a principle of justice which any informed person of good will would have to accept. Since a wide range of conflicting visions of the ideal society already exist, it seems objectionable to advance any new vision of this kind as a theory of *justice* - unless one is prepared to accept the burden of arguing that it is the *best* social vision. And, as I hope to make clear, any such argument faces immense problems.

The second kind of endeavor, which makes up the bulk of current justice theorizing, consists in detailed critiques of proposed theories of justice. Such critiques come in two grades of severity: let me call them "fatal" and "non-fatal". Successful criticism of the *non-fatal* kind merely undermines a theory's claim to rational superiority and reduces its rational status to that of one competing option among others. Successful criticism of the *fatal* kind undermines even the modest claim of a theory to being one competing option among others. Stated differently, non-fatal criticism merely aims to show that *not all* informed persons of good will must accept the theory in question, while fatal criticism aims to show that *every* such person must reject the theory. The distinction between these two kinds of criticism, although rarely drawn explicitly, is crucial for my purposes, because non-fatal criticism places a theory squarely within the scope of rational disagreement. Many allegedly "devastating" criticisms of theories of justice are, in fact, charitably interpreted to be of the non-fatal rather than the fatal kind. They are devastating only to a theory's monopoly claim to rational superiority. While all theories of justice *qua* theories of *justice* must, I think, make precisely this monopoly claim, it does not follow that undermining any such claim "refutes" the theory, in *every* relevant sense of this ambiguous term. Criticizing a theory for not compelling universal rational assent is not equivalent to having raised a fatal objection to it, as long as the criteria for having justified a theory of justice as superior to all rival theories are themselves in dispute.

With practice one quickly gets proficient at spotting non-fatal defects in any theory of justice yet proposed, because every such theory can plausibly be interpreted as sacrificing something of great moral importance in order to gain something else. But those who do not share each other's fundamental normative perspective are rarely converted. Up to a point, the debate between defenders of various theories and their critics helps clarify the implications of each perspective, but I cannot discern much progress toward rational consensus in the predictable ritual which now unfolds every time someone proposes yet another theory of justice.

Has justice theorizing reached a dead end? This is hardly an unreasonable suspicion. Appeals to "justice" provoke flaming passions and festering conflicts throughout the contemporary world - yet the current normative disarray in the philosophy of justice must be disheartening for decision-makers seeking moral guidance. Still, if philosophers cannot illuminate the nature of justice, who can?

In the face of this seeming impasse, a third endeavor in the field of justice theorizing appears to me promising: to investigate the sources of normative indeterminacy in this field in a more systematic way, with the goal of conceptualizing the sources and limits of *rational* disagreement. How much disagreement about justice does reason permit? What are the conceptual and normative factors on the one hand, and the intractable empirical factors on the other, which so often lock informed people of good will into bitter disagreement about justice? Working toward an understanding of some important sources of rationally irremediable disagreement about justice – *this* is the line of inquiry I propose to pursue here.

§2 There is nothing strikingly new about a focus on the problem of sorting out the conceptual, normative and empirical factors that may allow a variety of

conflicting but equally rational beliefs about justice.[1] Serious writers on justice, and most impressively John Rawls, have explicitly acknowledged limits to rational decidability.[2] But efforts at determining the sources and limits of the rationally undecidable tend to be selective and inadequate. The all-important distinction between the rationally required and the merely permitted is not easy to draw, and often not clearly acknowledged by writers who must surely be aware that the weight of reasons favoring *their* particular views about justice is far short of compelling.

I want to group contemporary philosophical work on justice into three broad categories. There are developed normative theories which generally include critiques of major alternatives (e.g. Rawls, Nozick, Dworkin, Ackerman, Galston, Nielsen, Phillips, Walzer, Gauthier, Narveson, Reiman, Sterba and others). Then there are many detailed critiques and critical surveys of theories of justice, often strongly informed by normative commitments without, however, presenting them as a developed alternative to existing theories (cf. Barry, Sandel, Wolff, MacIntyre, Larmore, Raz, Fishkin, Pettit, Campbell, Kymlicka and others). And in addition there is a small but growing "counter-literature" to the almost obsessive emphasis on the construction and refutation of ambitious theories of justice in contemporary political philosophy, a literature that tries to evaluate the limits of theory

---

[1] Hume gives the following as an example of what reason alone cannot determine in a dispute about just property claims (although his example is admittedly fairly minor and has to do with the need for convention-setting):

> "Sometimes the interests of society may require a rule of justice in a particular case; but may not determine any particular rule, among several, which are all equally beneficial....Sometimes both *utility* and *analogy* fail and leave the laws of justice in total uncertainty. Thus, it is highly requisite, that prescription or long possession should convey property; but what number of days or months or years should be sufficient for that purpose, it is impossible for reason alone to determine."(*An Enquiry Concerning the Principles of Morals*, 29-30).

[2] Rawls' painstaking analyses of the "priority problem"(40f) and the "index problem"(93f) in *A Theory of Justice* remain guiding examples of the sort of project I think needs to be carried through more systematically, in the light of the criticisms Rawls' own theory has received. An article that is exceptionally sensitive to problems of achieving "political consensus in the face of fundamental moral disagreement" is Amy Gutman and Dennis Thompson's "Moral Conflict and Political Consensus" (1990). Such "second-order" agreement-seeking, however, presupposes *some* consensus on the limits to first-order agreement.

construction by reassessing fundamentals of rational justification in moral and political theory (I have in mind, for example, Mapel's *Justice Reconsidered*, Shklar's *The Faces of Injustice*, as well as recent work by Sumner, Hardin and Haslett).[3]

How do I understand my own project in relation to this literature? Clearly, it fits the third category best. My hope is to reach at least some tentative conclusions about the limits of justified belief about justice, in the light of existing theories of justice and fundamentals of moral justification. The larger enterprise to which I hope to contribute is to find ways of transcending the current impasse in justice theorizing. This is best attempted, in my view, by first seeking a clearer understanding of rational disagreement about justice. It is hard to see how such an understanding could evolve without renewed efforts to rethink justificatory fundamentals with a synoptic view of the shortcomings of existing theories of justice. But the issues involved in this enterprise, as I came to discover, are so relentlessly complex that it is perhaps impossible for any single mind to encompass them in their fullness.

§3 I must immediately address three misconceptions of my project. I shall argue that there is room for rational disagreement about justice - which is *not* the same as the skeptical claim that no usefully determinate principles of justice can possibly be justified as rationally required or the dogmatic claim that I can give a final verdict on what issues are rationally undecidable. It will be instructive to consider these misconceptions briefly.

First: I am not arguing the thesis that being "usefully determinate" and "rationally required" are mutually exclusive properties of principles of justice, but something much weaker: that on a plausible account of the nature of moral reasoning, far greater disagreement is rationally permitted about the demands of

---

[3] L.W. Sumner, *The Moral Foundations of Rights* (1987); R. Hardin, *Morality Within The Limits Of Reason* (1988); D.W. Haslett, *Equal Consideration: A Theory of Moral Justification* (1987).

justice than seems to have been contemplated by many contemporary theories of justice.

Even if all theories so far proposed can be shown to be inadequate, that does not imply the impossibility of an adequate theory. Obviously, the idea of such a theory is not self-contradictory, and only self-contradictory ideas are demonstrably unrealizable. My project does *not* imply that an adequate normative theory of justice is impossible. But by taking more seriously the possibility of rational disagreement about justice, we may clarify the limits of what anyone can expect to achieve in this field. It may be possible to develop concepts and arguments which enable us to repudiate, with great confidence and clarity, exaggerated claims made by many people about the rational status of their favorite conception of justice.

Second: I cannot offer a final verdict on what disagreements about justice are rationally undecidable - I can only offer a list of what I judge to be *plausible candidates* for the status of rational disagreements. The concept of rational disagreement permits rational disagreement about what is to count as rational disagreement. But this fact does not, I believe, lead to a vicious regress; it rather *confirms* how indispensable the concept of rational disagreement is.

Others will find shortcomings in my list of plausible candidates for rational disagreement which may lead me to revise this list. All value judgments arise necessarily from someone's perspective. But judging an issue explicitly to be *a matter of rational disagreement* invites a sharing of perspectives about the question of whether we sincerely believe this issue to be resolvable by information, arguments or methods currently at our disposal.

§4 We must, however, immediately draw an important distinction between "rational disagreement" and "disagreement among rational people". Disagreement among rational people is disagreement which informed people with impeccable rational credentials carry on in good faith by the means of rational persuasion and

which they judge to be resolvable, given currently available information, arguments and methods (otherwise they would find it pointless to carry on their dispute *by such means)*. In contrast, rational disagreement - in a sense to be further clarified below - is disagreement which (although we may not be conscious of it) we have reason to judge to be *irresolvable* at the present, because we have exhausted all relevant and epistemically accessible information, arguments and methods. Rational disagreement could be seen as a form of fruitless disagreement which is not appreciated as such.

That there exists disagreement among rational people about some issue is, therefore, neither a necessary nor a sufficient condition for this disagreement to be rational disagreement, in the sense of "rational" which interests me here. It is not necessary (i.e. not all rational disagreement is disagreement among rational people), because the fact that a disagreement is *not* carried on by informed people of good will through means of rational persuasion clearly does not affect the question whether this disagreement involves irresolvable conceptual, normative or empirical issues. And it is not sufficient (i.e. not all disagreement between rational people is rational disagreement), for at least three possible reasons: (1) Most people who freely invoke the concept of justice are surprisingly unsophisticated about the conceptual, normative and empirical sources of rational underdetermination in this field. In fact, many undoubtedly reasonable people seem not to have seriously contemplated the very possibility of *rational* disagreement about social justice, or at least they often do not realize that a particular issue is a matter of rational disagreement. (2) Many issues of social justice are both enormously complex and emotionally absorbing. Progress is often possible in settling disputes about justice by appeal to rational argument. But there is a hard-to-define point at which many debates cease to explore the limits of rational agreement in constructive ways and instead lapse into a sterile repetition of inconclusive considerations. Even people with impeccable rational credentials cannot always be expected to realize when this

point has been reached, especially given the primitive state of our vocabulary for articulating varieties of rational underdetermination. (3) There may in turn be room for rational disagreement about whether a particular issue is, in fact, a matter of rational disagreement. Whether *all* relevant and epistemically accessible information, arguments and methods have been brought to bear on an issue is not a question whose answer is always clear-cut. Concepts like "relevance" and "epistemic accessibility", although indispensable in many contexts, are notoriously difficult to clarify. Clearly, how inaccessible a particular piece of information is may depend on the effort we are prepared to make in getting it. If the cost of getting some particular information is prohibitive in relation to the value of goods whose just distribution is in dispute, then obviously such information is, for practical purposes, inaccessible. To the extent that the concepts of "relevance" and "epistemic accessibility" are not sharply bounded, the distinction between what is *inside* and *outside* the scope of rational disagreement will also be somewhat hazy and shifting. But from the fact that a distinction is not sharp it never follows that drawing it is unimportant.

Persistent disagreement among informed people of good will provides at best evidence, not proof, that some issue is a matter of rational disagreement, in the relevant sense of "rational". For someone to suggest, therefore, that a particular issue is *outside* the scope of rational disagreement implies neither that there is no longer any dispute among informed people of good will about this issue, nor that carrying on such disagreement is necessarily irrational in every sense of this ambiguous term (see below).

My claim is *not* that focussing on the question, "How much room is there for rational disagreement about social justice?" is a magic key to the ultimate truth about justice or that it will transport us to a vantage point beyond ideology. My claim is simply that conceptualizing certain recurrent disputes about justice as rational disagreements defines a constructive line of effort for philosophical dialectic - probably our best chance for progress. Those who are strongly committed

to their favorite normative perspective may well resist the possibility of *rational* disagreement. It is my goal to challenge the rationality of this kind of meta-position about rational alternatives without, however, embracing all-out skepticism.

1.2 A CONCEPTION OF RATIONAL DISAGREEMENT

§1 How can "rational disagreement" be defined? To understand what a concept includes it is often helpful to ask what it is meant to exclude. No doubt disputes about justice often result from the fact that the disputants are ignorant of relevant and accessible information, insincere in their professed concern for justice, overwhelmed by passion, misled by authority figures or otherwise clearly deficient in their reasoning powers. These and similar factors are common sources of disagreement about justice. But claiming that the intelligence, rationality or sincerity of some or all of the disputants is somehow blocked, that they must be stubborn, uninformed or misinformed, shortsighted, biased or otherwise mentally blighted cannot plausibly explain the extent of good-faith disagreement about justice among informed people with impeccable rational credentials.

The most satisfactory way to mark the distinction between "rational" and "irrational" forms of disagreement about justice would perhaps be to give an explicit, uncontroversial definition of rationality which, when applied to problems of justice, will neatly separate disagreements that are rational from those that are not. But to my knowledge, no such definition of rationality exists. An extended investigation of models of rationality, although relevant to my task, would quickly lead me too far afield.[4]

Nevertheless, I have in mind what seems to me a tolerably clear and viable concept of "rational disagreement". I want to risk an explicit definition, realizing

---

[4] Cf. e.g. H.I.Brown's interesting revisions of the traditional model of rationality in *Rationality*, London: Routledge, 1988.

that both my definition and the clarifying remarks which follow give rise to numerous questions:

> Rational disagreement about justice is either currently or fundamentally irremediable disagreement between (proponents of) *basically rational* conceptions of justice. It can be of a conceptual, normative or empirical kind. Rational disagreement can be regarded as *fundamentally* irremediable if we have reason to think (whether we are conscious of it or not) that there are basic conceptual or normative points at issue which no relevant information, clearer reasoning or conceivable method will ever settle, or if there are intractable empirical points at issue which would require epistemically inaccessible information. Rational disagreement can be regarded as *currently* irremediable if we have reason to think there are empirical or non-fundamental conceptual or normative points at issue which additional information, clearer reasoning or conceivable methods will settle eventually.

I shall assume that the distinction between the conceptual, the normative and the empirical as marking off three contrasting categories of disagreement is sufficiently settled in principle, without suggesting that it is always easy to draw in practice. But what is a *basically rational* conception of justice? Is rational disagreement completely relative to whatever conception of justice one adopts? How can we know which disagreements are remediable or irremediable in the senses defined? Why does my definition of rational disagreement include disagreements which are likely to be remediable eventually by the usual methods of rationally resolving such matters? I address each question in turn.

§2 *What is a basically rational conception of justice?* I must answer by anticipating later results of this inquiry. A basically rational conception of justice is (i) a set of *bona fide* principles of justice which (ii) are consistent with a rational morality. A *bona fide* principle of justice is a principle which specifies a reasonable compromise among competing justice-based claims of different individuals in some distributive domain. *Justice-based* claims are claims to shares of society's resources which individuals (or groups), in pursuit of their well-being, press as their due on the basis of just-making criteria such as desert, need, equality, rights or good-faith

expectations. A *distributive domain* is any distributive context (e.g. social roles, institutions, geographic areas, ethnic groupings) which, ideally, is sufficiently well-demarcated that outcomes or processes for the just distribution of valued things can be specified within it. A principle of justice embodies a *reasonable compromise* among competing justice-based claims in some distributive domain if it is "appropriately responsive" (in a sense to be explained) to patterns of informed expectations and sentiment about justice and can be approximated by some feasible distributive mechanism.

A basically rational conception of justice, I said, must *both* include *bona fide* principles of justice *and* be consistent with a rational morality. But talk of a "rational morality" will immediately arouse the deepest suspicion. If we must first take sides in intractable disputes in metaethics before applying the idea of rational disagreement to justice, have we not begged the most interesting questions? Informed people of good will are no more agreed on the ultimate nature of morality than they are on the demands of social justice. How then can we presume to judge what makes someone's understanding of the nature of morality rational or irrational?

While deferring my full answer to this important objection to Part II, I must issue at least a promissory note. We must provisionally introduce a moral framework to fix ideas and to define *non-trivial limits* to rational disagreement (otherwise this concept will simply become synoymous with skepticism about justice). Many disputes about justice arise because disputants have fundamentally confused views about morality. Rational disagreement about justice, in the sense I have in mind, is possible only between disputants who hold basically rational conceptions of justice - *or whose dispute does not turn on the fact that they don't.*

This last qualification is important for the feasibility of my project despite intractable disputes in metaethics. While I can define *limits* to rational disagreement about justice only at the cost of making certain metaethical

commitments, the broadly consequentialist framework I shall adopt does not beg all the interesting questions. For one, many theories of justice ruled out here as incompatible with this framework will contain plausible *sources* (i.e. plausible candidates for the status) of rational disagreements which are more or less analogous to those I shall discuss here.

Therefore, what counts as a matter of rational disagreement, in the sense I have in mind, is not strictly relative to whichever particular conception of morality or justice one adopts. If it were, then obviously no greater consensus could be expected about the scope for rational disagreement than about those conceptions themselves. But many of the fundamental problems afflicting theories of justice – such as the justification problem, the priority problem, the aggregation-distribution problem or various commensuration problems – can be appreciated as rich sources of rational disagreement from within many different conceptions of justice or morality. The prospects for achieving at least partial consensus on what issues are sensibly to be regarded as matters of *rational disagreement* about justice are far better, I think, than the prospects for achieving consensus on the best conception of justice (although achieving greater consensus of the first kind may ultimately promote greater consensus of the second kind).

It is important to realize that the idea of rational disagreement about justice, while presupposing commitment to a strong enough moral framework to define limits to rational disagreement, does *not* presuppose similar commitment in order to identify plausible sources of rational disagreement. Even theories of justice which are incompatible with the broadly consequentialist moral framework I shall adopt will, in most cases, contain sources of rational disagreement which are closely analogous to those I identify here.

§3 *How can we know which disagreements are remediable or irremediable in the senses defined?* My project is to find reasons for considering certain issues to be plausible candidates for the status of rational disagreements, without claiming epistemic finality for my findings. If a particular debate about what justice requires in some particular respect has divided informed people of good will for a long time, and if we can identify conceptual, normative or empirical issues about which, upon reflection, most of us are of divided minds within ourselves, then surely at some point we have reasonable support for the hypothesis that perhaps some of these issues cannot be rationally resolved, at least not currently, or not at the level of fundamental theory, or not by appeals to the concept of justice.[5] This is not to say that they must be irresolvable by any method whatever which we might broadly consider rational. Even if we came to realize that the demands of justice, or even of morality in general, are indeterminate in certain respects, it will generally still be rational to solve our disagreements by compromise rather than force. But a rational resolution to an issue, in this purely pragmatic sense of "rational", does not preclude *continued* rational disagreement about this issue, in the sense of "rational" of interest to me here.

§4 *Why does my definition include as "rational disagreements" not only irremediable disagreements, but also those which seem to us remediable eventually by the usual methods of clearer reasoning and better information?* I have defined the concept to reflect the purpose I have in mind for it. Surely we want to use

---

[5] Two important (and easily confused) distinctions are that between fundamental moral theory and contextual judging, and that between the first-order and second-order level of moral reasoning. Fundamental theory, unlike contextual judging, abstracts from empirical particulars of specific contexts and assumes only general information about the nature of reality, knowledge and human values. The first order level of moral reasoning, unlike the second-order level, is the degree to which moral questions can be settled by appeal to fundamental theory *or* contextual judging, but without resorting to bargaining or other forms of purely pragmatic consensus-seeking in the absence of principled guidance. So *not all* matters which cannot be resolved at the level of fundamental theory must necessarily be resolved at the second-order level of moral reasoning: often, principled ways of contextual judging are also possible. And of course, second-order reasoning will in turn be constrained, in some ways, by fundamental theory.

"rational disagreement" in a sense in which saying, "Perhaps there is room here for rational disagreement" serves to signal to informed people of good will that we think all pertinent considerations and currently accessible information have been brought to bear on an issue without settling it, and that the debate has reached a stalemate of rehearsing inconclusive arguments over and over again. It is meant to signal the need to shift the debate to another level of reasoning, the second-order level of pragmatic consensus-seeking.[6] If I restricted the concept of rational disagreement so as to include only *fundamentally* irremediable disagreement, the resulting implications of normative finality would make its use far more controversial and would tend to shift debates simply to whether the disagreement at issue must, in fact, be *forever* irremediable. And this debate is likely to be as inconclusive as the earlier debate about what justice requires because, even if it is *true* that no amount of relevant information or clearer reasoning will ever settle some disagreements, I believe that we can rarely know this truth with certainty. Our capacity for empirical discovery, and perhaps to a lesser extent our conceptual and moral frameworks, evolve in unpredictable ways and may change the terms in which we see our most perplexing dilemmas.

But what about disagreements which seem irremediable at some abstruse level of theorizing but lead to normative agreement in practice? Moral reasoning is meant to solve practical conflicts, and without a practical conflict, there is no

---

[6] I am assuming that civil war or other brute-force attempts at dealing with first-order rational disagreement will, in most contexts, be ruled out by a basically consequentialist moral framework (and generally even by the informed self-interest of all parties involved). My claim, therefore, is *not* that appeals to justice must be altogether suspended at the second-order level of moral theory and give way to the use of force or threat of force. My claim is only that questions about the justice of various strategies of negotiating the terms of mutual accommodation *in recognition of the scope of rational first-order disagreement* are different from questions about the fundamental constraints imposed by justice on distributive outcomes and processes.

But clearly, the first-order/second-order distinction between levels of moral theory depends itself on a prior understanding of the concept of rational disagreement. Its normative content will, therefore, become clear only in the course of our inquiry. In what follows I shall be concerned purely with sources and limits of rational disagreement at the first-order level of moral reasoning – although the problem of rational disagreement may, of course, reoccur in regard to the justice of various possible negotiation strategies for second-order accommodation in the face of rationally irremediable first-order disagreement about justice.

problem for moral reasoning to solve. Why concern ourselves at all with disagreements which dissolve at the applied level?[7]

It is a fortunate experience indeed that many metaethical and high-level normative differences do not, for a number of reasons, preclude agreement at the applied level. Different lines of moral reasoning often converge on the same decision. People with clashing metaethical convictions often share similar moral intuitions about some particular case. Supporters of different ideologies experience pressure to form coalitions with a common platform. And often a compromise on some second-best outcome will have clear advantages, from everyone's point of view, over escalating the level of conflict. But the way I propose to use this concept, "rational disagreement" refers to currently or fundamentally irremediable disagreement about conceptual, normative or empirical issues at *whatever* level of moral reasoning they occur. Moral differences at various abstract levels often converge on agreement at the applied level, but in many cases they do not, and they are deep sources of many festering conflicts dividing human communities. In any case, agreement at the applied level despite strong disagreements at a more abstract level (e.g. the agreement of many conservatives and feminists about the desirability of outlawing pornography) is bound to create at best temporary moral bedfellows.

But is perhaps the entire search for abstract principles of justice fundamentally misguided? Jonsen and Toulmin distinguish

> two very different accounts of ethics and morality: one that seeks eternal, invariable principles, the practical implications of which can be free of exceptions or qualifications, and another, which pays closest attention to the specific details of particular moral cases and circumstances.(2)[8]

As will become clear, I fundamentally agree with Jonsen and Toulmin in endorsing a version of the second approach to moral decision-making. But I think we must be

---

[7] Cf. Jonsen and Toulmin, *The Abuse of Casuistry* (1988).

[8] Numbers in brackets following quotes give relevant page numbers in an author's work, as listed in the *Bibliography*. Titles or year of publication are given when more than one work by an author is listed. All references to Rawls' work are to *A Theory of Justice* (1971), unless a year of publication is given.

careful not to draw a false dichotomy between case-based and principle-based ethics, as if our judgment in particular cases were dispensable in the search for principles or as if an appeal to principles were dispensable for our judgments in particular cases. The understanding of moral justification I shall defend, following Rawls, is the search for convergence among informed people of good will of their wide reflective equilibria between principles, judgments in particular cases, and background theories. On such a coherentist view of moral justification, both moral judgments and principles are rational according to the degree to which they mesh with a maximally extensive and coherent belief system (see Part II).

1.3 OBJECTIONS

Justice theorizing is a philosophical mine field. Almost every constructive attempt confronts paralyzing objections.[9] The following objections have been urged against my project, all of them serious enough that I can hardly proceed further without confronting them:

> Having decided that a disagreement looks irremediable, would that fact alone not be sufficient to make continued disagreement irrational? If so, the concept of rational disagreement would seem to be straightforwardly self-contradictory. Or if not self-contradictory, it would seem to lack sufficient unity for a meaningfully focussed project. In any case it does not seem to be interestingly different from skepticism. And at best, a project of identifying points of rational disagreement about social justice is a mere exercise in taxonomy or semantics.

§1 *Is the concept of rational disagreement self-contradictory?* Having decided that a disagreement cannot be settled by drawing on information, arguments or methods currently at our disposal, would that fact alone not suffice to make continued disagreement irrational? The very idea of "irremediability" would seem

---

[9] At least two prominent contemporary justice theorists - Rawls and Nozick - have found it necessary to recant substantial parts of their original theories (cf. Rawls, 1985, 1988; and Nozick, 1989, 286-296). Gauthier is, I think, also forced to make major concessions (cf. Vallentyne, 1991, 25; 323-330).

to limit the forms of disagreement that could be rational. It is surely irrational to persist in arguing once one realizes that one's own position is not backed up by better reasons than one's opponent's position. Suppose it could be shown of conflicting positions about justice that none is ultimately rationally superior. This fact would imply, not that ongoing disagreement between proponents of opposing positions is rational, but on the contrary, that persisting in such disagreement would be utterly irrational. Therefore, the project of identifying points of rational disagreement about social justice, if successful, would make *rational* disagreement impossible. Therefore, my project is self-defeating.

This objection depends on the fact that the term "rational" is, of course, multiply ambiguous. If there are arguments in a dispute, understood by all parties, that can settle the issue in favor of one side, then we would surely want to call continued disagreement over this issue "irrational", in *some* sense. But if there are no arguments in a dispute that can settle the issue in favor of one side, and all parties understand this fact, we would also want to call continued disagreement over this issue "irrational", though clearly in *another* (incompatible) sense.[10] We must, therefore, distinguish two levels of normative disagreement: "first-order" disagreement at the level of fundamental theory or contextual judging, and "second-order" disagreement at a meta-level which recognizes the scope and limits of first-order agreement. Disagreement which is irrational if judged from this meta-level may *not* be irrational if judged from the level of fundamental theory. The recognition of rational agreement at the level of fundamental theory should indeed

---

[10] In view of the multiple ambiguities of "rational", would it perhaps be clearer to speak of "moral undecidability", and to abandon talk of "rational disagreement" altogether? (I owe this suggestion to Peter Danielson.)

I am not convinced that such a terminological switch would more clearly articulate my central thesis: the claim that, at the root of many disagreements about justice among informed people with impeccable rational credentials, there often are unrecognized or as yet poorly understood conceptual, normative or empirical sources of rational underdetermination. The undecidability of certain moral issues is merely the *effect* of rational underdetermination - which means that, in our search for the underlying *causes* of certain disagreements about justice, there is ultimately no way around making claims about what rationality does and doesn't permit. (I shall argue in Part 2 that my choice of a fixed moral framework is itself constrained by rationally non-arbitrary considerations.)

produce agreement at the meta-level that an issue *must* be resolved by other means than purely an appeal to fundamental theory. It is precisely my hope that a vastly clearer philosophical (and eventually public) recognition of the rational undecidability of many issues of justice at the level of fundamental theory or even at the level of contextual judging will motivate agreement on the need for consensus-seeking at the second-order level, free of mutual accusations of injustice. If the search for consensus at the second-order level is not simply to degenerate into a power struggle, however, it must be informed by an understanding of the reasons for first-order disagreement and itself be constrained by procedural principles of justice. Ultimately the hope is that all rational disagreements about social justice will turn out to be "remediable" in at least this sense - that informed people of good will *can* recognize the ultimate reasons for a disagreement, recognize the impotence of fundamental moral theory or contextual judging to resolve this disagreement, and settle at the level of negotiated compromise. This does not mean that it cannot be rational for people to persist in their commitments to cherished personal values *despite* the realization that these commitments are irremediably underdetermined by reason.[11] It does mean, however, that insofar as one's cherished but rationally underdetermined commitments conflict with other people's equally cherished and underdetermined commitments, the recognition of this underdetermination should cause a readiness, in informed people of good will, to find a *modus vivendi* while constricting the legitimacy of appeals to justice.

This claim does not imply, of course, that the recognition of rational disagreement in a given case necessarily makes *acting* on one's preferred course of action irrational. On any plausible account, considerations of justice are generally neither necessary nor sufficient for determining what makes actions rational or

---

[11] As Larmore puts this point, "The fact that a conviction of mine about the meaning of life is controversial, rejected even by others whom I consider reasonable, may not offer me a sufficient reason to suspend belief in it, if it continues to make sense of my experience."(52)

irrational. I certainly do not make the sweeping claim, therefore, that recognizing the existence of rational disagreement about one's views about justice makes it generally irrational, from the perspective of some individual's (or group's) interests or desires, to act on these views. I only make the vastly more modest claim that such recognition makes it irrational to think that *justice requires* one's preferred course of action.

§2 *Does the concept of rational disagreement lack sufficient internal unity for a meaningfully focussed project?* I have proposed to include certain empirical as well as conceptual and normative disputes within the extension of "rational disagreement". But does such a concept not verbally unify what is, in substance, quite disparate? In the case of empirical disputes we disagree about cause and effect in complex social situations but agree on what truths would resolve them; whereas, in the case of fundamental conceptual and normative disputes, there is sometimes no truth to be found, or perhaps we disagree even on that.

Let me make this objection as clear as I can. Is an inquiry with "rational disagreement" as its unifying concept a little like a book with "banks" as its unifying concept which includes chapters about lights or other objects arranged in neat rows, financial institutions, and the edge of water bodies? If I thought so, I could limit the extension of my central concept so as to exclude empirical issues. But I am convinced that this would be a mistake because, when it comes to disagreements about justice, conceptual, normative and empirical questions are often inseparably intertwined. For example, is it purely an empirical question what will in practice satisfy the idea that each person can justly claim an equal right to the most *extensive* scheme of equal basic liberties compatible with a similar scheme for all? Is it purely an empirical question whether mass starvation in the world should be conceptualized as a problem of overpopulation or of food distribution? Even where the empirical questions can be separated from the conceptual or normative ones, the

information required to settle some of the fiercest disagreements about justice is epistemically so inaccessible to us that we may have to classify them as fundamentally irremediable. For example, what exact mix of public and private ownership of the means of production will lead to whatever we may plausibly specify as the most desirable consequences overall? Or what total system of social institutions will, over time, make the economically poorest ten percent of the population as well off in absolute terms as possible? Might informed people of good will not be in (first-order) disagreement about the best answer to these questions *forever*, simply because of the epistemic difficulties of assessing the different consequences of all the various options? Questions of social justice present themselves as a bundle of intertwined empirical, conceptual and normative threads, and trying to untangle the conceptual and normative threads in isolation from the empirical can't work. Ideal theorizing that is not at least roughly guided by feasibility considerations has led, I believe, to a great deal of remoteness from the real-world debates about justice and has made many theories of justice, in Mapel's terms, "unworkably complex" or altogether "senseless".[12] My inquiry will, therefore, not exclude disagreement about intractable empirical matters from the rubric of rational disagreement.

§3 *Is rational disagreement about justice simply skepticism by another name?* The possibility of rational disagreement would seem to imply that there can be conflicting principles of justice which are equally rational. Whether something is reasonably to be regarded as just would no longer seem to be the sort of question

---

[12] An illuminating critique of theories of justice whose proposed "metrics of equality" are either "senseless" or "unworkably complex" is developed by David Mapel in *Social Justice Reconsidered* (1989). According to Mapel, "A metric is senseless if it relies on imaginary calculations that are impossible or if it fails to identify any common currency by which the relative value of different sorts of resources might be commensurated. A metric is unworkable if it requires information that is too complex and counterfactual to lead to practical principles."(6) He identifies crucial parts of Dworkin's, Galston's, Ackerman's and Rawls' theories of justice as senseless or unworkable.

that always has a disjunctive yes or no answer. How, then, is recognizing rational disagreement interestingly different from skepticism?

Skepticism about justice can either mean that there are no true principles of justice to be discovered which exist in the nature of things independently of people's shifting understanding of what is likely to promote their well-being. Or it can mean that there is no room between mind-independent truth and mere subjectivist preference for constructing a shared understanding of justice in a way that will be motivationally potent for rational persons. I admit the first but deny the second. Recognizing rational disagreement implies only that *some* conflicting conceptions of justice are equally rational, but not that *all possible* conceptions are. Skepticism about justice is certainly an understandable reaction to the normative disarray in current justice theorizing as well as to the rather indiscriminate use of justice-talk in everyday life. It is part of my aim in this inquiry to keep such skepticism at bay by retrieving a relatively modest rational status for appeals to justice.

§4 *Is the project of identifying sources and limits of rational disagreement about justice a mere exercise in taxonomy or semantics?* Just as we have made progress in the field of practical reasoning by labelling certain recurring fallacies, we can perhaps make progress in the field of justice theorizing by clearly naming recurring reasons for seemingly interminable disagreement among informed people of good will. There is power in naming; and by more clearly conceptualizing such recurrent sources of disagreement about justice - such as, for example, the criteria definition and inclusion problems, the justification problem, the priority problem, the aggregregation-distribution problem, various other commensuration problems, or the description-relativity problem - we may be better able to discern the presence of these problems in specific normative disputes. Although initial attempts at mapping plausible sources and limits of rational disagreement will seem to yield

only a rather unruly collection of points, clearer ordering principles may emerge later on, as our insights deepen about the ultimate nature of such disagreements.

Attempting to sort out the ways in which informed people of good will can reasonably disagree about what justice requires is not, I believe, merely *semantically* different from those approaches to social justice which offer a closed set of principles spelling out what justice requires. As Pogge puts it,

> What is needed is the recognition that knowledgeable and intelligent persons of good will may reasonably disagree about the fundamental issues dividing the world today (232).

Despite the diversity of views about justice, the possibility of *rational* disagreement is often not conceded, and neither the sources nor limits of such disagreement seem to me to have been investigated in enough depth and detail. Progress in justice theorizing may need an expanded vocabulary - a vocabulary to characterize normative indeterminacies in a way that allows people to express their first-order differences about justice without seeming to draw each other's rationality or good faith into doubt. It would clarify the distinction between first-order and meta-level disagreement. It would not presuppose the adequacy of truth-valuational dichotomies for judgments about justice.

The concept of rational disagreement, I believe, deserves a central place in debates about justice. If patterns of rational underdetermination could be conceptualized as focal points in work on justice and widely taught - much like the roster of fallacies taught in practical reasoning courses - it might dislodge some of the dogmatism afflicting many disputes about justice. It might defuse conflicts by turning disagreements *between* people into uncertainty *within* each person. I see no more promising way in which many disagreements about social justice could be, if not resolved, then at least transcended.

PART 2: <u>THE CHOICE OF A CONSEQUENTIALIST MORAL FRAMEWORK</u>

This part gives the rationale for relativizing the concept of rational disagreement to a particular moral framework and sketches reasons for provisionally accepting a form of consequentialism.

2.1 DEFINING NON-TRIVIAL LIMITS TO RATIONAL DISAGREEMENT

§1 Making metaethical commitments to a particular moral theory, it might seem, is both *unnecessary* in the context of a thesis on rational disagreement about justice, and in any case bound to be *unsuccessful*. Why not define "rational disagreement" in a framework-invariant way and then show how the problem of rational disagreement arises within some moral theories of current interest?

To introduce a controversial commitment to consequentialism is not a step I have taken lightly. My original intent was to follow Rawls' example of sidestepping intractable disputes in moral theory altogether and to focus my inquiry on characterizing plausible candidates for the status of rational disagreements as they arise across the spectrum of contemporary theories of justice. The problem turned out to be twofold. First, I accumulated a large and chaotic array of different points of disagreement which, from some moral perspective or other, could plausibly be characterized as "rational". And second, the complete absence of any suggestion of principled *limits* to rational disagreement made my project seem to collapse into a defense of skepticism about justice.

There can hardly be interesting limits to rational disagreement if any dispute about justice becomes immediately a matter of rational disagreement, provided only that a moral theory held by *some* people leaves this dispute irresolvable or provided only that any of these theories conflict with any of the others about how to resolve it. If a theist appeals to the dictates of God, a relativist to beliefs cherished within her

particular reference group, a subjectivist to the strength of her personal convictions, a moral realist to his intuitions about some Platonic realm of moral entities or properties, and a skeptic to the utter absurdity of moral claims, it is not difficult to imagine that they will possess little by way of a shared rational framework to settle specific disputes about justice. If we could narrow the range of *prima facie* acceptable moral theories to those fashionable among competent analytic philosophers at the present time - or even if we narrowed the range further by shortlisting only disagreements *within* and *between* various forms of consequentialism, contractarianism, or rights theories as potential candidates for the status of rational disagreements - several vexing problems would still remain. First, if all moral theories held by minimally competent analytic philosophers were accorded equal status, the scope for rational disagreement would still come perilously close to skepticism. For example, a morally clear-cut issue for a consequentialist may not be at all so clear-cut for certain types of contractarians: I am thinking here of Gauthier's moral views about the mass extermination of cultural groups whose vastly inferior technology renders the expected benefits of cooperation less attractive than the expected benefits of outright extermination (1986: 231-2). Second, the task of deciding, with any precision, who to exclude from the ranks of minimally competent analytic philosophers would be fraught with difficulty, and there is in any case something embarassingly arbitrary about using, as one's ultimate criterion for a moral theory's rational status, simply current fashionableness among one's own fraternity.

§2 But why not simply pick some moral theory or other and examine sources of rational disagreement within it, without making any claims about its rational superiority compared to other moral theories? Would clarifying such conceptual interrelationships not count as doing perfectly respectable philosophy? Why not abandon the ambitious idea that we might be able to say, about a given issue, that it

involves matters of rational disagreement - and instead contend ourselves with saying that the issue involves matters of rational disagreement from the point of view of such-and-such a moral theory, but of course, if you happen to subscribe to some other equally acceptable moral theory, then of course it may not?

Saying that will indeed be my fallback option. But I don't want to admit the defeat of my original hopes for this project so easily. I would seem to face three options for interpreting the concept of "rational disagreement": (1) give up talk about rational disagreement *simpliciter* and admit that whether an issue is a matter for rational disagreement may well depend on which of several equally acceptable moral theories one adopts - and that in the end there are perhaps no interesting limits to rational disagreement about justice at all; (2) find plausible candidates for the status of rational disagreements which arise whichever moral theory one adopts; or (3) take the plunge into metaethics after all and defend my stand. Weighing the pros and cons of each option, I found myself attracted to the third on the following grounds:

(1) Talk of rational disagreement relative to a particular metaethical framework is at best uninteresting and at worst misleading *if* it is admitted that there are other, equally acceptable frameworks relative to which the same issue may *not* involve matters of rational disagreement. Such an admission would turn the concept of "rational disagreement" into something of an oxymoron - *unless* , of course, everything becomes a matter of rational disagreement simply because it is rationally irresolvable within at least one moral theory or it is resolved by different equally acceptable moral theories in conflicting ways. But by relaxing the conditions for something to qualify as a rational disagreement in these ways, we have failed to distinguish rational disagreement from skepticism in any interesting sense.

(2) Finding plausible sources of rational disagreement which arise relative to *every* moral theory professed in good faith by sane people - even if there were such disagreements - seems quite obviously uninteresting. If many different moral

theories - for example, the range of metaethical beliefs defended by analytic philosophers today (some of whom are moral skeptics) - are taken as equally acceptable from a rational point of view, then there exist no principled, non-trivial limits to rational disagreement about social justice. The fact that *some* identifiably analogous types of disagreement (e.g. the description-relativity problem) occur across different frameworks is simply an inconsequential curiosity.

(3) Under the pressure of these considerations I wondered if I could perhaps provisionally endorse some moral theory after all. But *which* theory? Obviously, the theory which would emerge as the theory least ravaged by devastating objections from the widest possible reflective equilibrium of available moral theories.

§3 I find myself forced, by the need to set non-trivial *limits* to rational disagreement, to relativize the concept of rational disagreement to a particular metaethical framework. And, in an attempt to avoid turning this concept immediately into an oxymoron, I find myself forced to embrace the chosen framework for constraining rational disagreements not simply as one attractive option among many equally reasonable competitors, but as the framework which, I claim, our widest possible reflective equilibrium is likely to identify as ultimately the most reasonable way to think about the basic nature of morality.

But my attempt to claim, however tentatively, some degree of rational superiority on behalf of a certain form of consequentialism will immediately encounter a seemingly paralyzing objection:

> You have taken the fact of persistent disagreement about justice among informed people of good will as evidence for the possibility of *rational* disagreement. But just as informed people of good will have not converged on one 'correct' conception of justice, so they have not converged on one 'correct' metaethical theory. Debates about most or all of the moral theories you dismiss as rationally unacceptable are still very much alive in philosophy, and in fact many contemporary disputes about justice are between defenders of fundamentally different moral theories. The tension between your insistence on rational disagreement about justice and your

dogmatic denial of equally rational disagreement about moral theory seems irreconcilable.

In philosophy we are quick to reject an enterprise when we detect an irreconcilable tension at its core. But while I admittedly face a dilemma, I think there may yet be a way to salvage my project.

What exactly is my dilemma? If I don't commit myself to any moral theory, I cannot draw interesting limits to rational disagreement about justice, and my project will be trivial. If I do commit myself to a moral theory, my defense of it will necessarily contain important gaps, and my claims about the limits of rational disagreement about justice will seem weak.

Is there any way out? Rawls' initially appealing suggestion that we could sidestep moral theory altogether in justifying a conception of justice turns out, in the end, to commit him to a deeply problematic pragmatic-communitarian metaethic (see below). Therefore, to set non-trivial limits to rational disagreement about justice, I see no way around taking sides in metaethical disputes. And I believe there are four lines of defense against the objection of being dogmatically blind to the possibility of rational disagreement about metaethics, four ways of weakening the force of this objection sufficiently to save the credibility of my project of identifying plausible candidates for rational disagreement about justice: (1) by outlining key elements of my reflective equilibrium-based rationale for choosing a consequentialist framework; (2) by conceding the tentativeness of my claims about the rational superiority of consequentialism and allowing a framework-relative interpretation of my findings as a fallback option; (3) by stressing that plausible candidates for rational disagreement which I identify within a consequentialist framework will probably emerge, in analogous form, in at least some prominent rival frameworks as well; and (4) by making a case that the consequentialist framework I defend offers greater promise for identifying "principles of accommodation between moralities" than most other moral theories, including

Rawls' own attempt to identify principles of justice by avowedly sidestepping metaethics altogether (Rawls: 1975, 540; cf. 1985). Let me say a few words about each of these lines of defense.

(1) While I cannot fill in the details of a wide reflective equilibrium, I can at least make clear what job would have to be done to make a rational choice among alternative moral frameworks and to establish some presumption that it can indeed be done.

(2) My inquiry is concerned with the scope for rational disagreement *within constructivist, non-foundational, distribution-sensitive consequentialism.*[1] I may admittedly be too optimistic about the prospect of an eventual convergence of the wide reflective equilibria of informed people of good will on the question of the most rational moral theory. Or perhaps some unexpected elaborations of the theories I reject *are* rationally superior to the consequentialism I adopt, from the perspective of a wide enough reflective equilibrium. Moreover, the fact that known meta-ethical options are flawed cannot rule out the possibility that additional options will be discovered. On the basis of the arguments I can offer I am certaintly not entitled to claim, nor do I claim, that those who reject consequentialism must inhabit a realm of superstition which places them beyond the reach of meaningful debate about justice. Those who endorse some alternative theory may interpret any claim p which I make about the sources and limits of rational disagreement as including an implicit proviso: "p if q", where "if q" stands for *"if CND consequentialism or something relevantly similar is indeed the most rational way to think about morality."* The sort of consequentialism I sketch is indisputably a major contender among moral theories.[2] Therefore, even though this proviso

---

[1] Abbreviated CND consequentialism. This form of consequentialism will be clarified below (Sec 2.3 and 2.4).

[2] A work that argues against contractarian and natural rights theories and clears the ground for consequentialist constructivism is L.W. Sumner's *The Moral Foundation of Rights* (1987). A work that attempts part of the actual constructivist normative task is D.W. Haslett's *Equal Consideration: A Theory of Moral Justification* (1987).

weakens the interest of my conclusions, it is still quite worthwhile to ask: "At what points can informed people of good will legitimately disagree about social justice, *assuming they share a broadly consequentialist framework?*"

(3) The *limits* to rational disagreement will strongly depend on which moral framework one adopts, but particular *plausible candidates* for the status of rational disagreements will not. For example, reconciling the tension between aggregative and distributive values (which I label the aggregation-distribution problem) is a rich source of rational disagreement within many theories; and so are the criteria definition and inclusion problems, the priority problem and several others. A number of candidates for rational disagreement which I identify within consequentialism will also emerge, *mutatis mutandis*, within rival moral frameworks. If so, my efforts at clarifying their nature will not have been in vain, even if some moral theory turns out to be rationally superior to the form consequentialism I adopt here.

(4) In assessing the charge that adopting a particular moral framework is dogmatic for purposes of identifying sources and limits of rational disagreement about social justice, one must appreciate a special pragmatic advantage of consequentialism: this framework has a great deal of context-dependent flexibility built in. It leaves a great many matters up to contextual judging which other theories would rule out at the level of fundamental theory. It can, therefore, accommodate politically (although of course not metaethically) a wide variety of rival moralities in a way in which such rival moralities could not accommodate each other. In complex, pluralist societies it can, I think, provide the "principles of accomodation between moralities" which Rawls speaks of by justifying a liberal type of social order (see Ch 3.4).

2.2 REFLECTIVE EQUILIBRIUM AS A METHOD OF MORAL JUSTIFICATION

§1 Before sketching my reflective equilibrium-based rationale for CND consequentialism, I must address this question: What exactly *is* the method of reflective equilibrium? Understanding this method is crucial for appreciating the limits to the certainty and precision we should expect in moral matters. We must see, according to Rawls,

> ...if the principles which would be chosen [in the original position] match our considered convictions of justice.....But presumably there will be discrepancies. In this case we have a choice. We can either modify the account of the initial situation or we can revise our existing judgments, for even the judgments we take provisionally as fixed points are liable to revision. By going back and forth, sometimes altering the conditions of the contractual circumstances, at others withdrawing our judgments and conforming them to principle, I assume that eventually we shall find a description of the initial situation that both expresses reasonable conditions and yields principles which match our considered judgments duly pruned and adjusted. This state of affairs I refer to as reflective equilibrium. It is an equilibrium because at last our principles and judgments coincide; and it is reflective since we know to what principles our judgments conform and the premises of their derivation. (19-20)

On a superficial reading, Rawls' idea of establishing a reflective equilibrium seems very simple - closely related, in fact, to the dialectic method of testing principles against intuitively plausible counterexamples which moral theorists have employed all along. But Rawls is *not* arguing that simple coherence between principles and considered judgments is sufficient to justify acceptance of such principles. The method of reflective equilibrium differs from the traditional dialectic method in the way Rawls seeks to establish *simultaneous* coherence between our considered judgments about justice and a set of principles *and* between relevant moral as well as non-moral background theories and a description of a fair contractarian choice situation in which more or less the same principles of justice could plausibly be accepted. Relevant background theories sketched by Rawls include alternative moral theories, a theory of moral personhood, of procedural fairness, of the role of

morality in society as well as relevant findings of the social sciences, especially moral psychology. It might be thought that Rawls is trying to do even more than establish simultaneous coherence, namely simultaneous *implication* or *derivability*. He has sometimes been interpreted as arguing that his principles of justice are derivable from the contractarian choice situation *independently* of someone's considered moral judgments about specific cases.[3] But this interpretation of Rawls' argument, along with many criticisms based on it, are deeply misguided. Although Rawls misdescribes his theory repeatedly as "the contract doctrine"(e.g. 329), he does not, and could not, seriously claim that his original position has *by itself* the power to yield very determinate principles. And it would add a further layer of misunderstanding to construe Rawls' original position as a fundamentally contractarian attempt to get moral conclusions *from morally neutral premises,* although here again Rawls originally misdescribes his theory as part of the theory of rational choice (cf. 16, 583). But he also clearly admits, "We want to define the original position so that we get the desired solution"(141).[4]

Such statements have often earned Rawls the charge of circular reasoning (cf. e.g. Hare, 1973). But all Rawls claims, and can plausibly claim, is that our considered judgments about what constitutes a fair contractarian choice situation can be based on various background theories and kept independent, to *some* extent at least, of our considered judgments about what is just or unjust in particular cases; and that, therefore, trying to find a match or overlap between principles suggested by *either* set of considered judgements is not simply narrowly circular reasoning but establishes a degree of coherence among our moral beliefs which can be construed as

---

[3] Ackerman, for example, seems unclear about the limited role of the original position in Rawls' argument: "When the Rawlsian Zero confronts the Infinite Choice Set, it is impossible to choose *any* principles of justice until he is endowed by his creator with *some* set of preferences to guide his judgment."(339)

[4] As Rawls clarifies the role of the original position in his 1985 paper, "the original position is simply a device of representation: it describes the parties, each of whom are responsible for the essential interests of a free and equal person, as fairly situated and as reaching an agreement subject to appropriate restrictions on what are to count as good reasons."(237)

the best sort of support which any such beliefs can ultimately have. Although Rawls' choice of words occasionally opens him up to misunderstandings, it is a serious mistake to criticize him as if he considered principles of justice to be derivable from the original position, *independently* of whatever principles are antecedently suggested to us by our considered judgements of what is just or unjust in particular cases.

I think the way Rawls employs the method of reflective equilibrium must be understood roughly as follows. First he asks what could constitute a fair contractarian choice situation, based on our considered judgments about procedural fairness. Second he asks what set of principles of justice would be supported by our considered moral judgments about what is just or unjust in particular cases. Third he asks what *subset* of such principles would also be supported by a fair contractarian choice situation. Fourth he gives strong arguments about why the principles so identified must apply, not to the question of a given individual's distributive share, but to the question of what overall distributive outcomes a just society's basic structure must achieve. That most reflective people who followed Rawls' reasoning would in fact also converge on much the same set of principles was perhaps Rawls' hope and, briefly, his promise. (I shall discuss the shortcomings of Rawls' approach from the point of view of identifying sources and limits of rational disagreement about justice in Ch 3.2.)

§2 Daniels helpfully clarifies reflective equilibrium methodology as a generally applicable method of moral justification. Following Rawls, he distinguishes *narrow* reflective equilibrium (the traditional dialectic method of testing principles against intuitively plausible judgments) from *wide* reflective equilibrium, which he characterizes as follows:

> The method of wide reflective equilibrium is an attempt to produce coherence in an ordered triple of sets of beliefs held by a particular person,

namely, (a) a set of considered moral judgments, (b) a set of moral principles, and (c) a set of relevant background theories....We do *not* simply settle for the best fit of principles with judgments, however, which would give us only a *narrow* equilibrium. Instead, we advance philosophical arguments intended to bring out the relative strengths and weaknesses of the alternative sets of principles (or competing moral conceptions). These arguments can be construed as inferences from some set of relevant background theories....We can imagine the agent working back and forth, making adjustments to his considered judgments, his moral principles, and his background theories. In this way he arrives at an equilibrium point that consists of the ordered triple (a), (b), (c). (258)

The gist of the method of wide reflective equilibrium seems to me this: we must establish a *first* coherence between considered judgments about specific cases of a certain type and a moral principle which covers cases of this type, and a *second* coherence between these principles and relevant moral and non-moral background theories, *with an important further proviso*: the considered moral judgments which constrain our (moral and non-moral) background theories "must be to a significant extent disjoint" from the considered moral judgments that directly constrain our principle (Daniels, 260). "Coherence", however, is a vague concept that may denote logical relationships of widely varying supportive strength, ranging from mere *consistency* of thematically unrelated beliefs, via increasing *probability* or *derivability* in some loose sense, all the way to *deductive implication*. Obviously, when we give coherence arguments in support of some controversial conclusion, we want in every case to make the connection of logical support between beliefs as strong as we can possibly make it.

As I understand it, applying wide reflective equilibrium methodology (WREM) to the problem of justifying principles of justice essentially means constructing, on the basis of one's considered judgments about justice in specific cases, general principles which imply these judgments and which can then be used to make new judgments, *provided that* (1) these principles are consistent with well-established background beliefs such as, for example, the scientific world view, the findings of the various sciences, or well-founded beliefs about the hopelessness of

expecting a resolution to certain meta-ethical disputes; and *provided that* (2) the reasons that led one to espouse these background beliefs are, to some degree, independent of one's considered judgments about justice in specific cases. This method, as Rawls himself notes (49), is somewhat analogous to the way scientists construct scientific hypotheses on the basis of empirical observations - hypotheses which imply these observations and which can then be used to predict new observations (where such hypotheses are also independently constrained by relevant background beliefs, such as the mass of already well-confirmed scientific theories).[5] Just as a sufficient number of false predictions derived from some scientific hypothesis will eventually lead scientists to discard the hypothesis, so a sufficient number of intuitively unacceptable judgments derived from a moral principle will eventually lead proponents of WREM to discard a principle. Just as scientific method aims to establish the greatest possible coherence between observations, a particular hypothesis and all other relevant and already well-confirmed scientific theories, so WREM aims to establish the greatest possible coherence between considered moral judgments, a particular set of moral principles and well-founded and relevant background theories. But that is as far as Rawls allows the analogy to go, at least in the case of principles of justice (49). Whereas (most) scientists presumably regard hypotheses which are maximally consistent with both observations and already well-confirmed scientific theories as approximating some mind-independent physical reality, Rawls does *not* regard principles of justice which are maximally consistent with considered judgments and relevant background theories as approximating any mind-independent moral

---

[5] Daniels rejects this analogy between "considered moral judgments and observation reports" because accepting it seems to him to make reflective equilibrium methodology vulnerable to the "no credibility" criticism (271-3). That criticism is that the causal stories we can tell about how we came by our observations gives them credibility, whereas the causal stories we can tell about how we came by our considered moral judgments does not give them similar credibility. Daniels succeeds in pointing out plausible disanalogies between observation reports and moral judgments; but I do not see how that changes the point of the analogy, namely to illuminate the basic functioning of the method of reflective equilibrium in a helpful way.

reality. He leaves no doubt about his uncompromising anti-realism about principles of social justice. His two principles of justice, he says,

> are not regarded as a workable approximation to the moral facts: there are no moral facts to which the principles adopted could approximate. (1980, 564).

§3 What is the value of WREM as a method of justification in ethics, and in particular for principles of social justice? Is it our best method, is it one plausible method among several, is it completely worthless, or is it downright pernicious? There are arguments for all four views.

There certainly is no guarantee, using WREM, that even thinkers who start out with roughly the same "considered moral judgments" will accept the same principles in the end. As Rawls himself states (50), it is not at all a foregone conclusion that different thinkers, seeking maximum coherence between their considered judgments about justice, principles and relevant background theories, must converge on pretty much the *same* principles. There is an important difference between moral theorists working back and forth between moral judgments, principles, and relevant background theories, on the one hand, and scientists working back and forth between observations, hypotheses and already well-confirmed scientific theories. In this "working-back-and-forth" process, observations are constrained by empirical input in a way in which moral judgments are simply not, or at least *seem* not to be to many thoughtful people. For a contractarian, for example, a particular way of specifying the choice situation and the principles derived from it may be in reflective equilibrium with that person's considered moral judgments and background beliefs. For a religious person, the story of the divine origin of moral commandments, along with his particular religion's set of commandments, may be in reflective equilibrium with that person's considered judgments and background beliefs. For an intuitionist realist, the belief in the mind-independent existence of moral properties and in a moral

faculty of intuition by which they can be known, along with these intuitions themselves, may be in reflective equilibrium with his considered judgments and background beliefs. At best, someone may have examined absolutely *every* possible option of thinking about morality, but the moral principles and the account of their origin which she eventually accepts will only constitute yet another very wide reflective equilibrium with her considered judgment and background beliefs.

A proposed method of justification for principles of justice would seem to be of dubious value, in fact downright pernicious, if it allows different thinkers, all of them using this method in good faith, to emerge with strongly conflicting ideologies while legitimizing the claim of each to have "justice" on his side. But there also seems to be a strong case for considering WREM to be *trivial*. It seems trivially true that, if *any* method can supply consensus on principles of justice, wide reflective equilibrium method can, because in the final analysis there would seem to be no reasonable alternative to it. The idea of converging wide reflective equilibria is simply a description of the dialectic process at its most general, carried through to a point of maximum coherence of all beliefs of all people about all things. If the idea of converging wide reflective equilibria is construed in this way, as I think it must be, then it is of course true that it is our only hope of reaching rational consensus on moral principles. The best we can *ever* do is to achieve one sort of reflective equilibrium or another, and the more aware a person is of all the coherentist pros and cons of the various alternative ways of thinking about the nature of morality, the more sophisticated his particular reflective equilibrium will be. In this sense, WREM *subsumes* all other methods of moral argument within its scope.

We would seem, then, to have reached the paradoxical conclusion that WREM is both pernicious and trivial. On the one hand, using WREM has led different moral theorists to conflicting conceptions of justice. But on the other, there would seem to be no genuine alternative to WREM as our ultimate method of moral justification. How can we resolve this paradox?

The lesson to draw from the recent history of justice theorizing is not, I think, that WREM should be discarded as either pernicious or trivial. Instead we must recognize that the intellectual demands which this method makes on a moral theorist are so enormous and perhaps overwhelming that we can only have limited confidence in the claim of any one thinker to have established a *conclusive* wide reflective equilibrium for a particular set of principles. A plausible explanation for why using WREM to produce convergence in philosopher's beliefs about justice has been strikingly less successful than using scientific method in producing convergence in scientists' beliefs about the structure of the natural world would seem to me to be the following. Justice theorizing must draw on background theories in meta-ethics, metaphysics, epistemology, history and the social sciences, all of them fields of inquiry which involve such unsettled and elusive conceptual, normative and empirical matters that formulating readily testable hypotheses is rarely possible, and even informed thinkers acting in good faith are likely to emerge with substantially different conclusions. The unsettled state of knowledge in these fields does not, of course, imply that progress is impossible and could not lead to a progressive convergence of the reflective equilibria of different justice theorists. But justice theorists must, I believe, cultivate a fuller appreciation of the manifold possibilities for rational disagreement regarding the conclusions they have reached employing the method of wide reflective equilibrium.

2.3 CHOOSING CONSEQUENTIALISM IN REFLECTIVE EQUILIBRIUM

§1 I now want to sketch my understanding of how the method of wide reflective equilibrium applies to the problem of choosing a theory about the basic nature of morality. Justifying a metaethical theory by using WREM requires us, ideally, to examine all options and choosing that option which coheres best with *both* our considered moral judgments *and* all relevant background theories which

we have found reason to accept. I shall mention briefly the eight metaethical options that struck me as *prima facie* plausible at some point, only to put aside the first five without further argument.

The first five options, which I shall group together as "non-constructivist" theories, include theistic moral theories, social relativism, subjectivism, all forms of moral realism, and radical skepticism. The next three options, which I shall call "constructivist" theories, employ rights-based, contractarian, or consequentialist methodologies, even though they all arise from the same basic understanding of the main job which morality as a social institution is to accomplish. The difference I want to mark by using the distinction between "constructivist" and "non-constructivist" is suggested by Rawls when he explains his own Kantian constructivism:

> ...the parties in the original position do not agree on what the moral facts are, as if there already were such facts. It is not that, being situated impartially, they have a clear and undistorted view of a prior and independent moral order. Rather (for constructivism), there is no such order, and therefore no such facts apart from the procedure of construction as a whole; the facts are identified by the principles that result. (1980: 568)

On my understanding (which seems compatible with Rawls'), what makes a moral theory "constructivist" is that it regards morality fundamentally as a peculiar product of human reason, a *sui generis* type of rational social construct whose central function is to promote the satisfaction of human interests by suggesting suitable ways of restraining and coordinating people's behavior. Part of the idea of moral constructivism is the possibility that a variety of strongly conflicting systems of constraints on people's behavior will be equally serviceable for doing morality's job, although there will generally be a strong presumption against upsetting too many of people's ingrained customs and habits all at once. By contrast, what makes a moral theory "non-constructivist" is that this theory does not accord reason a fundamentally creative role in constructing morality so as to promote mutual trust

and reconcile human interests, but (at most) a purely epistemic role in discovering morality's pre-existing content.[6]

An overview of five non-constructivist options in their simplest terms will be helpful at this point to establish our bearings. Theists generally think that a divine being's commands are either the *direct source* of what is morally right and good, or our *best epistemic criterion*. Social relativists identify the moral right and the good with the laws or customs of some preferred social reference group. Subjectivists, the limiting case of relativists, consider the moral right and good to be matters of personal opinion. Moral realists hold that the right and the good involve natural or non-natural moral properties or entities existing independently of what anyone thinks about them. It is common, although certainly not necessary, for realists to be intuitionists, and to combine the idea that the right and the good are mind-independent realities with a distinctive doctrine about how we come to know this reality, called "intuitionism". Radical skeptics, on the other hand, think that moral talk captures neither a mind-independent nor a mind-constructed rational order, but is altogether irrational in every sense of that word.

What makes the above five theories non-constructivist? Theistic moralities are non-constructivist because they regard the content of morality as either metaphysically or epistemically given by God's commands, rather than constructed by human reason in the light of an at least partial convergence of our evolving understandings of what sorts of behaviors will best promote collective and individual well-being. Social relativism (as I understand this term here) is non-constructivist because it regards the content of morality as given by either the laws or customs of a social reference group, no matter whether these laws or customs pass reason's test of being plausible means to well-being. Subjectivism (as I defined

---

[6] Note that there is no suggestion in my definition of a constructivist moral theory that it *necessarily* involves the use of game theory or decision theory to derive moral principles from a contractarian choice situation. This approach, although sometimes regarded as the paradigm of moral constructivism (cf. Gibbard, 265), is simply one of several *prima facie* plausible constructivist approaches to morality.

this term) is non-constructivist, because it simply regards the content of morality as determined by whatever happens to be the content of a person's opinion about morality. Moral realism is non-constructivist insofar as it regards morality as somehow part of the basic ontological fabric of the universe. And radical skepticism is non-constructivist in the (utterly trivial) sense that it denies that morality, even if understood as an artificial construct of reason for the sake of promoting human goals, has rational status. For reasons of space I shall reject all these theories about the ultimate nature of morality out of hand here.

But two questions arise: Could there be constructivist versions of the above theories? And do I mean to reject them simply because they are non-constructivist? The answer to the first question is no. There could, of course, be constructivist moralities whose normative content is identical with the normative content of one or the other of these theories, but the meta-ethical interpretation of this content would have shifted in such fundamental ways that it would only confuse matters to regard them as the same theories or to use the same labels for them.

The answer to the second question is also no. I mean to reject the above theories, *not* because I claim to have an *a priori* argument to show that all non-constructivist moral theories must be false, but because I think that each of these theories has its own characteristic shortcomings.

What all non-constructivist options (aside from skepticism) have in common is that they accord human reason at best an epistemic role in discovering morality's content, rather than a fundamentally creative role in constructing and improving morality to serve human purposes and interests. Before discussing *prima facie* plausible constructivist methodologies, I want to ask if I have perhaps left out important non-constructivist options. I might seem to have ignored Kantian ethics or non-constructivist forms of consequentialism, such as Aristotelian or neo-Aristotelian teleology. But I think Kant's moral principles - the idea of moral consistency and the idea of respect for persons - are best given a purely constructivist

interpretation. Non-constructivist forms of consequentialism (let me call them "teleologies") face the question of how we could ultimately determine the good, apart from reflecting on people's diverse interests, values and purposes and then using WREM to set certain broad limits to the range of minimally plausible forms of the good life. But if a teleological theory is highly pluralistic in the goods it wants to promote, regards morality as a mind-constructed device for overcoming certain obstacles to social cooperation, and accepts WREM as the method for constructing limits to what can plausibly be considered to constitute a good human life, then it differs at most in degree from the understanding of morality I am about to develop.[7]

§2 Having set aside forms of non-constructivism, I will briefly consider three constructivist methologies which I shall call "rights-based", "contractarian" and "consequentialist". The methodology I want to adopt, in the end, is meant to be selected on the basis of a judgment, reached in the widest possible reflective equilibrium, about which of all *prima facie* plausible constructivist options is least afflicted with serious shortcomings. My efforts to sketch such a reflective equilibrium here will, however, necessarily be very truncated.

First, let me clarify further what I take all constructivist options to have in common. For a moral constructivist, morality is a system of behavioral rules and attitudes, and social pressures to uphold them, which evolved in human history in a variety of ways and due to a variety of stimuli, but whose continued existence and normative force and content can be rationally justified - and justified purely in terms of its necessary function in the promotion of people's interest in well-being.

But why is morality useful or even necessary for the satisfaction of people's interests? Haslett summarizes lucidly what we ultimately need a morality for:

---

[7] It is important to be clear about the distinction between WREM and constructivism. Constructivism is a basic conception of the nature of morality. WREM is the ultimate method for deciding whether a conception of morality, or any particular moral belief, meshes sufficiently well with the rest of one's total belief system to be rationally acceptable.

> Why should a society have a code of morality anyway? At least part of the answer is that a decent code of morality enables there to be the degree of security, and of trust in others, that is necessary in order for there to be mutually beneficial interaction among people. It makes this security and trust possible, in large part, by making possible a much greater degree of uniformity, and thus predictability, in people's behavior than there would otherwise be. In a society with a decent code of morality people will, because of this code, pay their debts, keep their promises, refrain from violence, tell the truth, and so on. Not all of the time, but most of the time. Therefore, to this extent, there will be a certain uniformity, and thus predictability, in people's behavior, the kind of uniformity and predictability upon which security and trust depend. Making this uniformity, and thus predictability, possible is the job, or one of the main jobs, of a code of morality. (77)

There can, of course, be rational disagreement about how sophisticated or well-adapted or farsighted any particular moral code is in doing its main job, but it seems clear what this job is. It is to ensure enough uniformity and thus predictability in people's behavior so as to create the necessary sense of security and trust in order to promote almost universally beneficial interaction.

Another clear formulation that encapsulates the rational basis for morality is the following by Peter Railton. Think of morality, he says,

> as a set of behavioral restraints and recognized permissions and obligations that function within a group to promote co-ordination in those circumstances in which co-ordination would be mutually beneficial. *Nothing here about representing an independently-ordered reality*. Still, we can see how the need to discourage special pleading and free riders, to promote consistent expectations and co-ordination in the projection of existing norms into new cases, to have teachable principles and associated habits of conduct, to have publicly-ascertainable procedures for application, and so on, would create quite practical pressure on behalf of consistency, simplicity, generality, the avoidance of singularities, and the appeal to abstraction to mitigate conflict and to accommodate approximation....A practice of discouraging singularities [special pleading by individuals] within the group becomes a closer approximation to a demand of universalizability as the group itself, or the reference group with respect to which it seeks to defend itself, becomes a closer approximation of all humankind. (187-8. Emphasis added.)

This description of morality's function seems to me to provide the essentials for *explaining* positive morality as well as for *justifying* changes to it.[8] It suggests how moralities have evolved under practical pressure as norms of co-ordination. And to the extent that rational foresight can help us figure out which logically possible norms of co-ordination are practically superior to others in terms of what is to our "mutual benefit", to that extent we will also be justified in choosing one particular content for these norms of co-ordination over others. It is true that Railton considers this account of morality "rather starkly practical - and perhaps therefore incomplete"(187); but it is not clear what *more* we are justified in believing there to be to morality.

§3 That rational morality does not describe a given order independent of what we conceive to be to our benefit, but can only be constructed by fallible human beings out of their shifting understanding of what sorts of behavioral restraints will effectively serve their mutual interests, is one of the fixed points of my own moral understanding, supported by my widest and most stable reflective equilibrium. How, then, might we go about restructuring positive morality in a more rational way?

1. The rights-based approach: Although rights-talk is notoriously associated with theistic or realist moral theories, there seems to be initially no reason why rights could not, in principle, become fundamental building blocks of a constructivist moral vocabulary. In fact, the core idea of this approach seems temptingly simple. The core idea of a rights-based approach is to look to some salient human capacities, vital needs, widespread interests or other common

---

[8] It is obviously important not to confuse the moral beliefs people actually have (the *positive* morality) with the moral beliefs that would be rational for them to have (the *rational* morality). Sometimes the term "ideal morality" is used for what I call "rational morality". This term, however, gives rise to confusion because any rational morality will have to be further subdivided into the moral beliefs which we should rationally follow under ideal circumstances and the moral beliefs which we should rationally follow under real-life circumstances.

features of human nature and simply to postulate rights of a certain content, scope and strength for possessors of such features. Constructivist rights theorists (e.g. Mackie; the early Nozick) may claim, for example, that the fact that individuals are capable of having their own ideas of the good should be construed as sufficient to give them a right of a certain strength and scope to seek their own good in their own way.

One problem with taking rights as morally basic by linking them up with human interests, needs or capacities is that there seems to be no clear rationale for deciding which of potentially infinitely many interests, needs or capacities should be regarded as giving rise to rights of what sort of content, scope and strength. As Sumner asks, after reciting a familiar litany of currently popular, but strongly conflicting rights-claims:

> ...which aspects of our nature are the relevant ones? We are beings capable of choice - do we therefore have liberty-rights? We are also beings capable of being injured by others - do we therefore have claim-rights not to be harmed? We are also beings who need the support and assistance of others - do we therefore have claim-rights to be given assistance? If we lack any of these rights, why do we lack them? If we have them all then how can our nature determine which is to take precedence when they conflict? How, in general, can we distinguish between the relevant and the irrelevant aspects of our nature without presupposing a particular outcome for the argument?...[N]ature, even our nature, underdetermines selection of a set of basic rights and thus provides no effective control over the proliferation of basic rights principles. (125-6).

Linking features of human nature with rights-claims without any additional checks, it turns out, is far too simplistic a way to go about constructing a viable morality. Such a rights-postulating process might work in limited contexts if employed with restraint, but it couldn't accomplish morality's main job in a large-scale pluralistic society, because it will immediately lead to an unprincipled proliferation of conflicting rights-claims. As Sumner shows, there seem to be insuperable obstacles to building ways of avoiding such conflicts into the strength and scope of each right itself (1987; 92-126).

In addition to *pure* rights-based theories in which only rights are basic, there would seem to be hybrid theories for which both rights and goals are basic (e.g. Nozick, Dworkin). In these approaches, rights place absolute or at least very strong side constraints on the pursuit of goals, or goals are invoked to settle conflicts among rights. But hybrid theories face a dilemma. Either the rights postulated by such theories constrain the pursuit of consequentialist goals only in ways which seem ultimately more effective in promoting consequentialist goals - in which case these rights are justifiable in terms of goals and therefore non-basic after all; or rights are basic and therefore not justifiable in terms of goals - in which case they must postulate rights of a kind which will effectively hamper the pursuit of consequentialist goals without any compensating beneficial consequences. The first type of hybrid theory obviously collapses into indirect consequentialism, while the second needs a very strong rationale for recognizing goals while yet postulating anti-consequentialist rights (and this rationale would then make rights non-basic after all). And for a moral constructivist, there is only one *prima facie* promising place where a rationale for anti-consequentialist rights could come from: contractarian arguments.

Since the practice of generating substantive rights directly by an appeal to features of human nature does not seem to admit of any principled limit, since conflicts between such rights can only be sorted out by other considerations, since hybrid theories would inevitably seem to collapse into other approaches and since consequentialism can itself provide a clear rationale for rights-talk (see Ch 3.1), I can see no good reason why moral constructivists would want to admit rights as basic.

But doesn't this rejection of rights as basic building blocks of a constructivist morality overlook the way in which WREM is ultimately also rights-based – insofar as WREM seems to presuppose everyone's *right to be convinced by reason* of moral principles? (Peter Danielson, pers. comm.)

In reply I want to deny that WREM presupposes a "right" to be convinced by reason, in the sense in which "rights" are at issue here: as morally basic constructs which allow individuals or groups to make claims to certain means to meet/satisfy/develop certain human human needs/interests/capacities, where the mere existence of some need/interest/capacity is regarded as *sufficient* justification for the corresponding rights-claim. It is true, the philosophical demand for justification of moral claims can itself be grammatically reformulated as a rights-claim, but this reformulation would, I believe, be misleading. The demand for justification is so fundamentally tied up with the very enterprise of making sense of the social phenomenon called morality that I would regard "the right to be convinced by reason" as of a clearly different order than all other rights. My claim about rights is simply that the prospects seem poor that constructing a morality purely by spelling out a list of basic rights will be able to do the job for which we need a morality in the first place. In fact, I would claim that no practical sense attaches to the idea of a "right to be convinced by reason", because requiring a full understanding of the complex WREM-based rationale behind many moral principles and judgments would surely overtax most people's rational capacities. I do not take the employment of WREM by moral philosophers to entail such a requirement.

2. *The contractarian approach:* The core idea of the contractarian approach is to conceive of constructing the rules of morality as analogous to negotiating the clauses of a multi-lateral contract. This core idea can be elaborated in various ways (e.g. Hobbes, Locke, Rousseau, Rawls, Scanlon, Gauthier). I find it useful to distinguish simple contractarianism, contractualism and mutual-benefit individualism.

(1) *Simple contractarianism* wants to derive moral constraints on individual behavior or social policy simply from the idea that these constraints must be such

that individuals could voluntarily agree to them for their mutual benefit. But can this idea *alone* generate a morality? What sorts of principles would turn out to be accepted by contractarian bargainers is highly sensitive to how the conditions are specified under which the agreement is imagined to have been reached. That poses a dilemma for the simple contractarian. Either the conditions under which agreement on principles is reached are specified to include moral constraints or they are not. If the conditions under which principles are chosen *are* specified to include moral constraints (as in the case of Rawls' "original position"), then obviously those constraints are not themselves based on contractarian reasoning and presuppose some other basis. If just any arbitrary specification of the contractarian choice situation is picked, then the particular principles chosen under those conditions will have no moral force, since different principles would have been chosen given a different arbitrary specification of the choice situation. Contractarian constructivists who choose the original position which will give them their own favorite principles obviously beg the question (unless they give consequentialist justifications for their choices - but then they are basically consequentialists, not contractarians). Therefore, it would seem, simple contractarian approaches to constructing morality will fail (cf. Sumner, 1987a: 159-60; Ripstein, 115-137).

Is this criticism, pressed strongly by Sumner, entirely fair? Is he not applying a double standard - rejecting the contractarian approach because of the underdetermination of the choice situation, while construing the major indeterminacies of consequentialism benignly as leaving room for rational disagreement? It would seem, however, that the simple contractarian's dilemma must be understood, not as a problem of underdetermination, but rather as a blatant case of question-begging. Contractarian bargainers cannot even *begin* to bring the logic of bargaining to bear on the problem of getting morality off the ground without first making morally loaded assumptions about features of the choice situation. Lomasky makes this point eloquently when he says,

> Contract may be a useful heuristic for fixing attention on the agent-relativity of value to which political prescriptions must conform themselves, but neither inside nor outside of an original position is it able in a non-question-begging way to establish generally binding principles of justice. (153-4)

This is not to deny that the simple contractarian idea is methodologically useful, but it is best understood, in the manner of Rawls, as an expository device within reflective-equilibrium-based consequentialism, *after* various plausible moral assumptions have been made about the fairness of the choice situation.

(2) *Contractualism* conceives of morality as those social arrangements which can be publicly defended as *reasonable*, in the sense that persons who are motivated to seek informed, unforced general agreement on the terms of social cooperation will eventually converge on these arrangements as creating the greatest possible social space for their more or less widely differing ideas of the good. Contractualists (e.g. Scanlon, Larmore, Stout) use the idea of the social contract simply as a way to make vivid the point that morality is socially constructed out of both aggregative and distributive concerns of individuals. The idea of contractarian bargaining can model this construction in wide or narrow contexts in which the need for moral principles arises and, given certain moral presuppositions, can serve as a useful criterion for testing existing moral beliefs. These presuppositions - that the agreement must be informed, unforced, general, and essentially liberal - can be plausibly derived from consequentialist assumptions. In fact, the normative implications of contractualism and WREM-based consequentialism seem to me essentially indistinguishable.

But if so, which is morally basic: consequentialism or contractualism? What, ultimately, matters morally? Is it the promotion of some interpersonal goal which can plausibly be regarded as maximizing well-being, or the achievement of unforced general agreement among informed, autonomous persons?

I see no need for moral constructivists to be reductionist here. It would seem plausible that, in most contexts, the contractualist criterion can be given a

consequentialist rationale, and the consequentialist criterion a contractualist rationale. We could, therefore, plausibly aim to satisfy the consequentialist and contractualist criterion simultaneously and construct interpersonal goals which informed persons who are motivated to seek unforced general agreement on the basic terms of social cooperation would choose as affording the greatest possible social space for realizing different but equally acceptable conceptions of personal well-being.

(3) *Mutual-benefit individualism*, however, is a different story. As I understand it, the core idea of this form of contractarianism - associated most prominently with David Gauthier - is to eschew any attempt at aggregating different individuals' interests in order to define interpersonal goals (of either a global or contextual nature), on the grounds that a rational person has absolutely no reason to pursue *inter*personal or collective goals, except insofar as they happen to coincide with her personal goals. Rational persons are those who care only about maximal fulfillment of their own ("considered") preferences, whatever these may happen to be. The project of mutual-benefit individualists is to reconstruct the social order in such a way that it embodies only those moral constraints which every person will almost always find it rational (in his *actual* self-interest) to comply with. It is easy enough, of course, to lower one's moral standards to the point where they coincide with people's *perceived* self-interest; but what is interesting about Gauthier's theory is the claim that redesigning the social order along libertarian lines would lead to maximum informed preference satisfaction of all individuals (without, presumably, stunting their preferences in any objectionable way). Obviously, if mutual-benefit individualists have a blueprint for social reform so as to ensure maximal informed preference fulfillment - while avoiding any messy aggregation of the preferences of different individuals in the manner of standard cost-benefit analysis - then it would seem that Gauthier's mutual-benefit individualism ought to be happily embraced by a consequentialist constructivist as well.

My main project in this inquiry - identifying sources and limits of rational disagreement about social justice - will obviously take a very different shape, depending on the success or failure of Gauthier's contractarian justification of libertarianism. I face a difficult decision. I clearly cannot proceed in my main project without judging the merits of Gauthier's theory. To discuss the merits of Gauthier's work in a few paragraphs would simply seem to be attacking a strawman. But to make space here for the extended discussion it deserves would threaten to overwhelm other stage-setting concerns for my main project.

It seems to me best, therefore, to anticipate my conclusion but postpone my discussion of Gauthier's theory to a later chapter where I assess the possibility for rational disagreement between egalitarian and libertarian liberalism (Ch 3.5). My conclusion will be that mutual-benefit individualism ultimately fails to justify libertarianism and offers only a very partial substitute for moral decision-making in terms of promoting interpersonal goals.

3. The consequentialist approach: The core ideas of a consequentialist approach, as I understand this term here, are concisely expressed by Sumner:

> I begin with some assumptions about the likely shape of a consequentialist goal. The raw materials for that goal will be an inventory of ultimate goods. Whatever other goods this inventory might include, it seems reasonable to suppose that in one way or another it will acknowledge the value of those states which are the standard sources or components of individual well-being: life, health, liberty, autonomy, sociality, the development and exercise of powers and abilities, and so on. These goods must then be collated into some global value. Whatever other considerations this global value might include, it seems reasonable to suppose that it will be aggregation-sensitive, thus that it will acknowledge the force of increasing the overall extent to which individuals enjoy these central ingredients of their well-being....It might well consist of more than this [maximizing the sum of individual welfare], if it admits other particular goods as well or adopts a combinatory rule which is also sensitive to factors other than their aggregation. (201-2)

Sumner's account exemplifies the undogmatic, non-foundationalist way in which consequentialist constructivists must approach their task.[11] Consequentialism, as I understand it, is that form of moral constructivism which says that we should give our morality whatever features are likely to lead to the most *desirable* combination of satisfied interests of individuals in their well-being. I shall call this goal the "ideal consequentialist goal" and clarify below what "desirable" here can plausibly mean. I shall also use the phrase "X promotes well-being" for the more cumbersome phrase "X promotes the ideal consequentialist goal" (where X can be anything whatever - generally a principle, a right, an act, a judgment, a criterion, a decision, a policy, a state of affairs, a social institution, an attitude or something of that sort). The consequentialism I want to defend is both constructivist and non-foundational; constructivist, because it denies that moral properties would exist even if human beings had not constructed them to suit their purposes; and non-foundational, because it makes our understanding of what counts as "promoting well-being" dependent on a coherentist process of seeking a wide reflective equilibrium of relevant considerations (see below). Moreover, I want to defend a consequentialism which, in Sumner's words, "adopts a combinatory rule which is also sensitive to factors other than...aggregation"(202).

Three major questions arise: Why, of all imaginable consequentialisms, should we choose this particular form? Is its goal ultimately coherent? And how can we decide, *in practice*, what would count as having achieved or approximated this goal?

---

[11] "Consequentialism" has, of course, been defined in many different ways. Traditional textbooks often define it, unhelpfully, as the theory which evaluates actions purely by their consequences rather than their intrinsic nature. Others define it as doing whatever will produce the most good or the best state of affairs overall (cf. Scheffler). Some define it as the moral theory which uses agent-neutral as opposed to agent-relative reasons (cf. Nagel). Larmore defines it as that moral theory which demands that we must hold ourselves responsible, not just for what we alone do, but also (to the extent that this is in our power) for what will follow because of what everyone else will do as a result of our acting this way (cf. 147-8). Raz lists seven theses as "historically associated with consequentialism"(268). To clarify the relations between all these definitions of consequentialism, and to weigh the pros and cons of each, would be a book in itself.

2.4 THE STRUCTURE OF CND CONSEQUENTIALISM

§1 Why pick the standard of promoting the most desirable combination of satisfied interests of individuals in their own well-being as our ultimate consequentialist goal? The answer is that given *what we want a morality for* in the first place, it makes sense to construct it out of individuals' interests in their well-being. It is a necessary (though not sufficient) condition for acceptable moral arguments that they must be practical and engage motivations that people have or could acquire through moral education. But why, of all the imaginable ways to construct a morality out of individuals' interests and a minimally objective conception of well-being, should we pick exactly *this* way? Again my answer must ultimately appeal to a wide reflective equilibrium. This type of consequentialism seems to me best able to meet all serious objections to standard rule utilitarianism without giving rise to the objections afflicting perfectionist teleologies. In particular, it meets the objection that consequentialist aggregation fails to respect the "distinction between persons" or "individuals' separate existences", and therefore does not treat them with equal consideration.[12] The ideal consequentialist goal (to be clarified below) seems to me to embody the equal consideration principle in the deepest possible way. It is not easy to see what more it could possibly require to respect individuals' separate existences than to use, as our guiding ideal, this consequentialist goal.

In reply to the familiar objections to utilitarianism based on aggregation, Haslett's lucid defense seems to me worth quoting at some length:

---

[12] These two expressions come from Rawls and Nozick, respectively. For the debate about the merits of utilitarianism, cf. e.g. J.J.C. Smart and Bernard Williams, *Utilitarianism: For and Against* (1973) and Amartya Sen and Bernard Williams, eds., *Utilitarianism and beyond* (1982). I was persuaded of the merits of CND consequentialism by D.W. Haslett's *Equal Consideration: A Theory of Moral Justification (1987)*, which defends a version of it at length against anti-utilitarian objections.

...a social pressure system, including the moral norms inherent in it, must be evaluated in terms of people's *interests*....[But] in terms of *whose* interests should it be evaluated? The answer proposed here is: In terms of everyone's, giving each individual's interests *equal* consideration. If, indeed, a social-pressure system is to be evaluated in terms of people's interests, it is hard to see how, merely by insisting that everyone's interests be considered *equally*, the utilitarian could thereby be accused of not taking the separation of persons seriously.

But the critics might press matters still further. *Why*, they might ask, should we consider everyone's interests equally, especially in that sense of "equal consideration" being defended here?....we should consider everyone's interests equally because doing so is in everyone's best interests. It is in everyone's best interests in that, from a standpoint of self-interest, each of us would *ourselves* choose that we do so if fully informed and not subject to irrelevant influences – if, in other words, our choice were made under ideal conditions, the very conditions, I submit, that would have to be met in order for our choice to have been made altogether freely. This is why equal consideration is justified; this is the connection between self-interest and morality; and this is the most profound reply to any who would claim that the utilitarian theory proposed here does not take the separation of persons seriously. For if evaluating social-pressure systems by considering everyone's interests equally is in the best interests of each individual – each *separate, distinct* individual – then how could this utilitarian theory be said to treat anyone as a mere means? (221-2)

Thoroughgoing individualists of the libertarian kind will, of course, remain unmoved by Haslett's argument. They may see transaction-by-transaction calculations of mutual advantage, rather than the quest for impartial justifications of interpersonal goals, as the *only* promising motivation or rationale for social cooperation. I shall defer the discussion of libertarianism to Ch 3.5.

The question I must address now is whether the proposed consequentialist goal is fundamentally intelligible, self-consistent or coherent. What exactly is "the most *desirable* combination of satisfied interests of individuals in their own well-being"?

The following account of the ultimate consequentialist goal seems to me to be the most plausible way to spell out the idea of achieving the most desirable combination of satisfied interests of individuals in their own well-being. The idea behind this goal is not very different from the idea underlying the Golden Rule or Ideal Observer theories of ethics, implemented from a perspective which is (in

principle) enriched by a full understanding of all other perspectives. I find a thought experiment suggested by Haslett useful (29-39), although I want to clarify its details and limitations. You can judge for yourself, in principle, what would be the required course of action from the standpoint of achieving the most desirable combination of satisfied interests of individuals in their own well-being. Imagine you yourself had each person's life to live, and so experienced the interest of each person in her own well-being - including, of course, the very strong interests of most people in not being seriously disadvantaged for the sake of other people's greater well-being, as well as their interest in having pleasant hedonic states brought about by actual experiences rather than hallucinations. Suppose you had perfect knowledge of the extent to which each possible course of action would satisfy the interest of each individual in her own well-being. The interest I am talking about here is neither some paternalistic "objective" interest nor a completely unreflective "subjective" interest, but an idea of interest that reflects each person's subjective interest *constrained by the very wide limits of rationally permissible interpretations of human well-being*. Then, if you were guided by the consequentialist standard, you would choose the course of action which you would choose if you yourself had each separate person's life to live (and were not, of course, permitted any inconsistencies across person-stages). Then, given the assumptions about perfect knowledge and the absence of other influences except concern for your own well-being at each person-stage, you would choose a course of action within a certain range of equally acceptable courses of action.

Other perfect knowers, going through the same thought experiment, will probably not settle on exactly the same course of action. The courses of action they would choose would, however, differ only to the extent that they would modify their unreflective subjective interest at each person-stage differently from you in the light of the very wide limits of rationally permissible interpretations of human well-being (cf. Sec 3.4.2). To the extent to which different people who strive for this

goal in an ideal way would settle on different courses of action, to that extent the ideal consequentialist goal may simply be fundamentally indeterminate.

There is no reason I can see why the ideal consequentialist goal should not contain a degree of fundamental indeterminacy, in the sense that it permits a range of options. Because of this fundamental, as well as the vast epistemic, indeterminacy of this ideal goal, practical decision-making procedures will have to be adopted which are plausible contextual approximations to it.

§2 Let me explain the rationale behind three key components of my formulation of the ideal consequentialist goal. First: What all consequentialists must want, by definition, is to promote well-being in some plausible sense. Well-being in this sense must be somehow related to people's interests, but this relation isn't simple. It is not plausible that simply allowing everyone to pursue their raw *subjective* interests will promote well-being, because such interests may be self-undermining or socially destructive, and can completely miss the objective content of the concept of human well-being (see Sec. 3.4.2). But neither will forcing a person to pursue someone else's idea of what his or her objective interests are (no matter how far these diverge from subjective interests) plausibly promote well-being because a person's interests are, after all, constitutive of this person's identity, and the free pursuit of one's interests is perhaps the deepest source of human fulfillment. Any plausible empirical index of what makes people happy just does not bear out the claim that people's happiness is a matter of everyone's realizing the same determinate set of values. If neither the (distribution-sensitive) aggregation of raw subjective interests nor the satisfaction of rigidly objective interests plausibly captures what is involved in promoting well-being, then obviously consequentialists need some sort of compromise between the two. That is why I think we must, in defining the ideal consequentialist goal, temper the distribution-sensitive aggregation of subjective interests with a theory of human well-being

which limits the range of interests whose satisfaction can count toward promoting well-being. The concept of well-being must, therefore, be regarded as having "objective" content which imposes broad limits on the range of interests whose satisfaction CND consequentialists think morally important (see Sec 3.4.2).

Second: The ideal consequentialist goal requires you, first, to put yourself into the shoes or psyche of each person and to assess, *from that person's perspective,* what is in his or her interest before trying to find a way that will best harmonize all people's interests. Some such "perspective-switch" must obviously be incorporated into any consequentialist moral standard which is to do justice to the fact of our separate existences. You are to imagine that you yourself would actually be the person suffering every single one of the consequences of whatever course of action you settle on, with no compensation in a future life. But having done your imagining, you will actually have to decide on one course of action, because only one course of action can be realized in the actual world - and that means that some people's interests will unfortunately have to be frustrated and their well-being compromised. If you yourself suffered every single consequence without compensation, you would want to frustrate as few vital interests as possible, and certainly not sacrifice all of some people's well-being for the sake of marginal increases in the well-being of the majority. So plausible second-best criteria to approximate the ideal consequentialist goal, as well as extremely difficult judgments, will be necessary. But no self-consistent standard - consequentialist or otherwise - can escape these judgments or avoid ever sacrificing some people's vital interests or even their entire well-being. There are genuine tragic choices where, no matter what we do, some people will have to suffer so that a greater number of others won't have to suffer even more. So our proposed consequentialist standard permits the inevitable, for example, sacrificing some so that others may survive. But built into this ideal standard is, it seems to me, the most plausible *conceptual* procedure for capturing what giving equal consideration to everyone's interests

would mean. How to make this conceptual procedure even remotely practical is another question, the problem of second-best approximation, which depends on first getting as clear as we possibly can about our ideal aim.

Third: It may seem inconsistent of this standard to require you to find a resolution among conflicting interests from the perspective of having to live everyone's life. The deep problem here is not that it is actually impossible for you to *know* what everyone's interests are. There clearly is a fact of the matter here, and there are in any case various practical ways to discover what people's approximate interests are. But when it comes to harmonizing all the conflicting interests of individuals, there exists no neutral perspective, *in addition* to all the separate perspectives of individuals. To be sure, each individual can in principle harmonize conflicting interests from his perspective, but if all perspectives of all individuals become yours, which of the resulting perspectives should you choose to settle on the course of action that necessarily sacrifices some people's interests for the sake of satisfying those of others? Whose perspective is the privileged one in whose interest the conflict will get resolved?

I think this objection, while troubling, is not fatal to distribution-sensitive consequentialism. Take first a two-person universe in which the interests of A and B are in conflict. What would it mean to resolve the conflict by choosing the course of action which each would choose if each had both lives to live? If the conflict is resolved from A's perspective, A's interests would seem likely to fare better than B's; if from B's perspective, B's interests would seem likely to fare better than A's. A's and B's perspectives are the only ones there are; there is no third, neutral perspective from which conflicts could be resolved. But this is an oversimplified description of the situation. It is true that the conflict must be resolved from either one perspective or the other; it is not true that those two perspectives, after being subjected to the process described in our thought experiment, are necessarily very far apart. The very point of the thought experiment is that you imagine being the

subject of other persons' experience. If each of two people came to know what the other person's experience is like, then a psychological merging of perspectives could take place which would not be identical to the original perspective of either. It is this *enriched* perspective - which, in the real world, we can of course only approximate in the crudest way - from which the consequentialist standard asks you to resolve conflicts between different people's interests. If, in the two-person case, your judgment is informed by the impartial sympathy for *both* perspectives required by the Golden Rule, you would choose that course of action which ensures or promotes the well-being of both individuals together, wherever that is possible. You would not routinely maximize (total or average) interest satisfaction between A and B, because in that case, B might forever end up making sacrifices, and that is not the choice you would make if you actually had each life to live and experienced B's interests, including his interest in not being seriously disadvantaged for the sake of maximizing interest satisfaction overall. In situations where sacrifices on the part of either A or B become inevitable, and the current level of well-being of both is roughly equal, both will be called upon to sacrifice some amount. In some tragic situations where, for example, one person's life has to be sacrificed for the other's survival, there simply may not be a morally preferred answer to the question of whose life should be sacrificed (although some procedural ways of settling this question may be morally preferable to others).

It may still be objected, however, that it is far from clear how a definite decision results. Will at some point a commitment to maximization take over, or are there some distributive constraints built into the imaginative projection of myself as having everyone's life to live? As McDonald has put this objection,

> What you picture [in the above thought experiment] is that when A looks at both his own and B's life and B does the same their judgment will converge. The process of convergence seems to be causal. There is no perspectival criterion advanced, such as maximization. But then it may well matter how

> A and B feel about the separateness of individuals. What if A is in his guts a maximizer and B a separator (valuing the separateness of persons more)? Moreover, what if A and B are not equally prudent in their enriched judgments - what if A's enriched sense of their interests leads him to approve of risking admittedly vital interests for trivial ones, just to add zest to life? To be sure A and B each picture what is best on the basis of their enriched perspectives. Neither leaves out of account something the other has included. But each values or weighs matters differently. Or to use another analogy, each colours the same picture with a different hue. (Personal comm.)

In principle we could, of course, handle the sort of problems which McDonald raises here by requiring *iterations* of the 'perspective-switch' thought experiment until convergence is achieved. We can imagine the first attempt at achieving an enriched perspective to remain unenriched by such factors as people's differing in risk aversion or being "maximizers" as opposed to being "separators" at heart. During the second run of our thought experiment, once the results of the first run are in, we can factor into our enriched perspectives a mutual appreciation of the extent to which risk aversion and maximizer-separator preferences are differentially distributed. Even given such mutual appreciation of differences, however, risk-friendly maximizers or risk-friendly separators may still weigh outcomes differently from risk-averse maximizers or risk-averse separators. Perhaps we need to imagine additional iterations of our 'enriched perspective' thought experiment to achieve sufficient convergence in our judgments about what constitutes the most desirable combination of satisfied interests of individuals in their well-being.

But this suggestion strikes me as rather gimmicky. There is no denying that both problems McDonald raises – the problem of differences in risk aversion and the problem of reconciling maximizing and distributive considerations – constitute rich sources of rational disagreement about justice. There will be many occasions, then, when consequentialists will, at the first-order level of moral reasoning, converge on similar decisions only in a very limited sense. While the consequentialist well-being test will rule out some distributive options as unacceptable, the choice among a wide range of remaining options will have to be

made at the second-order level of moral reasoning, by seeking conventions of mutual accommodation through political bargaining.

I cannot discover any way to spell out, in the abstract, just at what point we could allow maximization to take over without opening ourselves up to the charge of failing to respect the separateness of persons. We must admit, I believe, that *at level of fundamental theory divorced from specific contexts,* imagining myself to have everyone's interests can generally yield only very vague answers to the problem of resolving different people's conflicting interests. We can say, for example, that it is generally better to satisfy a greater aggregate of interests of a larger number of people than a smaller aggregate of a smaller number, provided only that the vital interests of the few are not sacrificed for the less vital interests of the many. *At the applied level,* however, once we know what the actual interests of people affected by a decision are, we will usually be able to make more or less confident qualitative judgments which strike a credible balance between aggregative and distributive considerations.

I conclude that the proposed consequentialist goal is self-consistent, albeit afflicted with both fundamental and epistemic indeterminacies. How we are to judge, in practice, what best approximates this goal will be discussed in greater detail below (Ch 3.3).

§3 My project is to identify plausible sources and limits of rational disagreement about justice. I approach this project by (1) tentatively accepting a consequentialist moral framework as our most promising metaethical option; and (2) examining how this framework constrains principles of justice. But my approach faces the following objection:

> You insist that morality is constructed by ourselves as an instrument for the promotion of human well-being *and* that it constrains ideas about justice. But how can ideas about justice be constrained by your artificially constructed moral standards unless you yourself have built these constraints into them?

And how can you build these constraints into them, unless you already *know* what the right ideas about justice are? You reject the contractarian approach as question-begging. But isn't your attempt to find rational constraints on permissible ideas about justice by deriving such constraints from your consequentialist goal hopelessly question-begging as well?

This seems like a devastating objection. But it rests on a misunderstanding. It interprets the constructivist project of contractarians and that of consequentialists in analogous ways: as a foundationalist enterprise. But while moral foundationalism is indeed characteristic of contractarians (at least of the mutual-benefit individualists among them), the form of consequentialism I endorse is coherentist rather than foundationalist. We cannot avoid talking metaphorically about our most central epistemic assumptions. What does the justifying of moral claims, on my account, is a web of cohering considerations whose basic geometry is governed by a clear understanding of what we want a morality for, and whose points of contact with the empirical world are both universal and contextual interests, a fairly open-ended conception of human well-being, social meanings and conventions, widespread and deeply ingrained patterns of expectations and sentiments, and the best available information about the state of the world and the causal workings of things. The ideal consequentialist goal, and plausible derivatives of it, are simply nodes in the web of our examined moral and non-moral beliefs, "Archimedean points" within a reflective equilibrium upon which many strands of reasoning converge and from which many strands of reasoning radiate outward (cf. Rawls, 260). What I have called the ideal consequentialist goal is meant to sum up the general spirit and guiding direction of the consequentialist enterprise, and not somehow to stand alone as a self-evident, deductive foundation for moral decision-making which is known prior to and independently of everything implied by it.

My project of rationally reconstructing morality is, therefore, not linear in the sense of starting from independently justified foundations which then allow us to deduce principles of justice without circularity. We must construct all parts of our morality together in a coherent way, and cannot first produce an account of moral

foundations, untainted by any assumptions about justice, and then somehow derive standards to constrain ideas about justice. If our ideas about justice had to be consistent with moral standards which are untainted by assumptions about justice, then obviously justice-based considerations would be routinely overridden by the non-justice considerations contained in these moral standards. And that, as many critics of utilitarianism point out, would fundamentally conflict with our considered moral judgments. It would also conflict with our background theory of morality's main job and would, therefore, be doubly ruled out by the method of wide reflective equilibrium. It is clear that morality's job can only be accomplished by taking people's justice-based claims seriously as an important source of information about their well-being and by avoiding as far as possible the destructive unleashing of their sense of injustice.

Let me consider the feature of the basic consequentialist standard which seems perhaps most obviously question-begging. Why should consequentialist constructivism embody a general commitment, not simply to maximizing interest satisfaction, but also to considering different people's interests equally? The rationale for including, at the core of the basic consequentialist standard, a commitment to equal consideration has to do with elementary facts about human psychology which Brian Barry, among others, has eloquently spelled out:

> Someone who was engaged in impartially choosing principles to govern his life with others would not endorse a principle on the basis of its favoring himself, his friends and relations, or those with whom he felt some kind of affinity. But in just the same way someone seriously engaged in the search for principles that could not reasonably be rejected by others engaged in the same search would surely recognize that it would be a waste of time (as well as a breach of good faith) to put forward a principle whose only merit was that it would be favorable to himself, his friends and relations, or those with whom he felt some kind of affinity. For it is obvious that those who were put at a relative disadvantage by such a principle would wish to reject it, and it could not possibly be said that they were being unreasonable in so doing. (1989: 290)

In constructing the details of our basic moral standards, we had better incorporate from the start, as integral building blocks, some explicit reference to a sort of

impartiality which plausibly captures, as traditional utilitarian theories did not, the distributive concerns of individuals making claims *qua* individuals. The reasons are two-fold. First, the greatest well-being overall varies in direct proportion to the well-being of individuals, and an individual's justice-based claims to social resources are a useful, though far from infallible, source of information about what is likely to promote that individual's well-being. Second, if an individual's expectations to the shares of social resources which he sincerely regards as his due get disappointed in a way that cannot be explained to that individual in terms of the necessity of balancing competing claims of comparable urgency, there is a great likelihood of anger, humiliation, resentment and similar sentiments. These are, from a consequentialist point of view, undesirable in themselves and also have potential for being unleashed in socially destructive ways. That is why any plausible form of consequentialism must take the difficult job of balancing the competing claims of individuals to a share of social resources in a principled, publicly defensible way extremely seriously.

§4 It may be doubted whether distributive *constraints* on the aggregation of goods can themselves, consistently, become one of the *goods* to be factored into this aggregation. But "goods", in the relevant sense, are simply whatever human beings care about within objectively plausible limits of what can constitute personal well-being. "Constraints" on the aggregation of goods are whatever measures are necessary or useful to prevent individuals from sabotaging each other's pursuit of personal well-being. Where do we get such constraints from? Where else, but from asking ourselves what basic types of goods we care about sufficiently to want to build them into the goals spelled out by our ultimate consequentialist goal. We sufficiently care about having the distinction between persons recognized in distributional matters to make the recognition of this distinction a justifiable consequentialist constraint. So by the definition of what makes something a good

and a constraint, respect for the separateness of persons is both a good and a constraint. In fact, whether one says that respect for the separateness of persons and happiness are both goods to be weighed against each other in pursuit of an acceptable mix, or that the one functions as a constraint on the pursuit of the other, is a purely semantic matter for a moral constructivist.

Including the good of equal consideration, along with the good of happiness and possibly other values, into our ultimate consequentialist goal makes this standard more complicated to apply as well as fundamentally indeterminate in many cases. But no form of consequentialism could avoid this complication without becoming immediately vulnerable to persuasive counterexamples. One cannot consistently criticize a fundamental moral standard *both* for not taking seriously the separateness of persons *and* for becoming more complex and less determinate as a result of doing so. Of course, the ideal consequentialist goal will require plausible second-best approximations to become normatively useful in specific contexts of application (see Ch 3.3).

As we construct, or more accurately reconstruct and modify, the social pressure system called morality in a rationally transparent way, we can't all at once toss everything out that exists either in the philosophical tradition or in the cultural world around us. Although nothing is beyond question, and every component of our morality must ultimately find its functional rationale in the whole, we must immediately distinguish two intertwined components of any practicable morality we are likely to end up with: components for which a *universal* rationale can be given in terms of the reasonable preconditions for any plausible type of social living; and a much larger proportion of components whose rationale is *culturally or otherwise contextually relative* to the way of life of specific groups and their self-understanding as members of that group. To expect a non-contextual rationale for every component of our rationally reconstructed morality, or even to ask that the two basic types of components be neatly separable, is not a reasonable expectation of

the enterprise of constructivist, non-foundational, distribution-sensitive consequentialism.

2.5 RAWLS' AVOIDANCE OF MORAL THEORY

§1 If one of the greatest moral philosophers of our time thinks he can bypass most metaethical issues in justifying principles of justice, we cannot simply dismiss this idea. The case for explicit metaethical commitments (and for a broadly consequentialist commitment in particular) will be strengthened, I believe, by briefly considering Rawls' rationale for wanting to avoid moral theory at almost any cost.

In his later writings, Rawls still maintains that, ultimately, the wide reflective equilibria of different people will converge essentially on his two principles of justice. But he now gives these principles an explicitly pragmatic-communitarian interpretation (cf. 1985, 1988). In "The Priority of Right and Ideas of the Good", for example, he says

> ...the are several ways in which I would now revise the presentation of goodness as rationality; perhaps the most important would be to make sure that it is understood as part of a political conception of justice viewed as a form of political liberalism, and not as part of a comprehensive moral doctrine. The distinction between a comprehensive doctrine and a political conception is absent from *Theory*, and while I believe nearly all the structure and substantive content of justice as fairness (including goodness as rationality) is unchanged when it is seen as a political conception, the understanding of the view as a whole is very significantly shifted. Charles Larmore in his *Patterns of Moral Complexity*...is quite correct in vigorously criticizing the ambiguity of *Theory* on this fundamental matter. (1988: 254 footnote)

If WREM applied to the problem of social justice is followed through beyond the conflict of ideologies - Rawls now seems to think - the ultimate equilibrium to be reached does not yield a rationally superior moral theory, but merely "fundamental intuitive ideas viewed as latent in the public political culture"(1988, 252). These ideas alone can serve as principles of political accommodation among groups with

irreconcilably diverse moral beliefs. A theory of justice has a very limited role: defining whatever terms of political accommodation are most likely to promote peace and cooperation in a particular society. It would be a mistake to base one's theory of justice on any controversial beliefs about the nature of morality, because only by not taking sides in the intractable debates of moral theory can a theory of justice discharge its social role. The terms of mutual accommodation among groups with diverse ideas of the good may be very different for different societies. In advanced pluralist democracies, Rawls claims, his own principles of justice define such a society's best chance for achieving a reasonably peaceful and stable social order.[13]

The hope that WREM allows the political philosopher to altogether avoid the complex debates in moral theory - a hope which Rawls perplexingly labels "the method of avoidance"(1988, 240) - certainly has its initial attraction. But is this hope justified? If wide reflective equilibrium methodology subsumes all other moral arguments within it - as I suggested above - then it seems doubtful that bypassing all questions about the nature of morality is compatible with WREM. On the other hand, if no agreement on the nature of morality is realistically to be hoped for, then this verdict itself will, of course, become one of the well-established background beliefs to be factored into our reflective equilibrium. But it is far from clear that the fact of irresolvable disagreement about morality among informed people of good will - rather than implying moral skepticism - would imply that the task of justice theorizing consists purely in making explicit whatever fundamental intuitive ideas latent in a society's public political culture may best lend themselves to the purpose of mutual political accommodation among different groups.

---

[13] The idea of justifying liberal principles of justice purely in terms of their ability to provide a *modus vivendi* is eloquently defended by Charles E. Larmore, *Patterns of Moral Complexity* (1987), p. 74-6 and 123-7.

A number of troubling questions arise in the face of Rawls' pragmatic-communitarian shift. Is the fact that some set of principles fulfills the *social role* of justice (i.e. defining terms of political accommodation) sufficient to accept them as principles of *justice* - rather than, say, as a concession to power politics? Can principles actually fulfill the social role of justice unless they are first recognized or justified *on independent grounds* - that is, on the basis of an understanding of justice as more than simply those widespread intuitions which will best promote political accommodation? Can reflective equilibrium methodology actually justify a suitable (i.e. sufficiently determinate) set of principles convincingly as principles of *justice* to allow these principles to fulfill the social role of justice without involving any controversial metaethical commitments whatever? These issues are obviously important for an inquiry into the sources and limits of rational disagreement about justice. I want to approach them in two steps: First, why exactly did Rawls find it necessary to shift to such a pragmatic-communitarian justification of his theory? Second, is Rawls' new interpretation of his theory be convincingly derivable by applying WREM to the problem of social justice?

§2 What motivated Rawls' pragmatic-communitarian shift? In "Justice as Fairness: Political not Metaphysical", Rawls tells us that understanding his theory as a political conception of justice is intended to counter the criticism that

> this conception depends on philosophical claims I should like to avoid, for example, claims to universal truth, or claims about the essential nature and identity of persons. (1985: 223).

The claims to universal truth Rawls wants to avoid are claims about which general conception of the nature of morality is rationally superior. "The essential point," he says,

> is this: as a practical political matter no general moral conception can provide a publicly recognized basis for a conception of justice in a modern democratic state....[S]uch a conception must allow for a diversity of doctrines and the

plurality of conflicting, and indeed incommensurable, conceptions of the good affirmed by the members of existing democratic societies. (1985: 225).

The first and obvious reason for Rawls' shift in the interpretation of his theory of justice is his desire to find principles of accommodation between different moralities without taking sides in the intractable disputes of moral theory. But there is another important reason for this shift. By interpreting his two principles of justice, not as spelling out the demands of universal justice *sub specie aeternitatis*, but far more modestly as a basis for an overlapping consensus among citizens of advanced pluralist democracies, Rawls hopes to meet the communitarian objection about the nature of the self presupposed by his theory. Rawls' description of the identity of the bargainers in the original position, stripped of most of their individuating characteristics, seemed to postulate a fictitious "disencumbered" self which can make choices in splendid isolation from the particulars of its social environment. Rawls had even claimed (although in a different context) that "the self is prior to the ends which are affirmed by it"(560).

Communitarians such as MacIntyre, Sandel and Taylor were quick to argue that this Kantian idea of the self as choosing its goals according the universal dictates of reason is untenable: the self is embedded in a particular community and culture, it is constituted by its choices, and these choices are fundamentally dependent on its social roles and relationships.[14]

Rawls answers this criticism in two steps. *First*, he stresses that the description of the bargainers in the original position was always intended, not to capture the essential nature and identity of persons for whom his theory defines just arrangements, but purely as an expository device designed to make vivid a

---

[14] Cf. MacIntyre, *After Virtue* (1981) and *Whose Justice? Which Rationality?* (1988); Sandel, *Liberalism and the Limits of Justice* (1982); and Taylor, *Hegel and Modern Society* (1979). The Kantian idea of the self attacked by communitarians assumes, as Kymlicka puts it,

> "that the self is prior to its socially given roles and relationships and is free only if it is capable of holding these features of its social situation at a distance and judging them according to the dictates of reason."(1990)

conception of what constitutes reasonable conditions of impartiality for choosing principles of justice (1985, 237). That Rawls' reply to his communitarian critics represents more than a minor clarification can be gleaned from his admission that

> ...it was an error in *Theory* (and a very misleading one) to describe a theory of justice as part of the theory of rational choice, as on pp. 16 and 583. What I should have said is that the conception of justice as fairness uses an account of rational choice subject to reasonable conditions to characterize the deliberations of the parties as representives [sic] of free and equal persons; and all of this within a political conception of justice, which is, of course, a moral conception. There is no thought of trying to derive the content of justice within a framework that uses an idea of the rational as the sole normative idea. (237, footnote)

Second, Rawls now tries to make his theory of justice compatible with the communitarian idea of the self as embedded in communal practice simply by adopting a radically communitarian interpretation of his theory. He makes clear that his theory is not a universal theory of justice for human beings in a wide variety of cultural settings, but on the contrary a highly restricted theory applying only to persons whose identity has already been shaped by conditions prevailing in a particular type of community - namely advanced industrialized pluralist democracies. Rawls stresses that his theory "starts from within a certain political tradition", and it

> "tries to draw solely upon basic intuitive ideas that are embedded in the political institutions of a constitutional democratic regime and the public traditions of their interpretation"(1985, 225).

People whose self has been shaped by *such* a cultural environment, Rawls seems to argue, will generally have the sort of self which values roughly his list of primary goods as all-purpose means to the satisfaction of their goals.

Rawls' pragmatic shift attempts to undercut the case of his communitarian critics by reinterpreting his theory as a version of communitarianism. But Rawls' bid to join the communitarian camp seems to have made him few friends among communitarians. The debate seems simply to have shifted from the question as to

whether Rawls' morally justified liberalism depends on an implausible theory of the self as detachable from its ends to the question as to whether Rawls' pragmatically justified liberalism can take seriously the value of community as a fundamental ingredient of the good life for human beings.

§3 Is Rawls' pragmatic-communitarian interpretation of his theory of justice consistent with regarding reflective equilibrium methodology as the ultimate method of justification? If the two were inconsistent, it is clear that Rawls would have to give up his pragmatic justification rather than WREM. Obviously, the idea that a pragmatically justified principle (i.e. one which promises to promote political accommodation) could override a principle which is justified by a wide reflective equilibrium (the greatest possible coherence of a moral theorist's beliefs) is incoherent. Therefore, Rawls' pragmatism is sustainable only if it is consistent with WREM.

Given the central importance of showing the pragmatic justification of principles of justice to be fully consistent with and, in fact, implied by WREM, it is strange that, in Rawls' 1985 paper, the only reference to his earlier much-acclaimed method of moral justification occurs as a parenthetical remark in the following context:

> We look, then, to our public political culture itself, including its main institutions and the historical traditions of their interpretation, as the shared fund of implicitly recognized basic ideas and principles. The hope is that these ideas and principles can be formulated clearly enough to be combined into a conception of political justice congenial to our most firmly held convictions. We express this by saying that a political conception of justice, to be acceptable, must be in accordance with our considered convictions, at all levels of generality, on due reflection (or in what I have called "reflective equilibrium"). (1985: 228)

I admit to being puzzled by how this remark, or any of the other things Rawls says in his 1985 paper, could possibly be construed as sufficient to make the logical compatibility between WREM and the pragmatic thesis clear and convincing.

Although Rawls has weighty reasons for his pragmatic-communitarian shift, there are even weightier considerations for thinking that Rawls' "method of avoidance" is inadequate for justice theorizing.

(1) *The epistemic worry:* Rawls' argument for bypassing moral theory relies heavily on the claim that no sufficient measure of agreement is to be hoped for in this area. But how does someone know this to be true, and with what degree of certainty? From the fact that universal agreement has not happened in the past, it does not follow that expecting a sufficiently widespread measure of agreement in the future is completely unrealistic. Can we categorically say that none of our metaethical options shows any promise whatever, if articulated with clarity, to win the adherence of at least a clear majority of rational people? In any case, such a rather pessimistic conclusion can obviously not be accepted by all justice theorists simply on the strength of our faith in Rawls' judgment, but must be reached in wide reflective equilibrium, as the outcome of examining the pros and cons of our various metaethical options in the light of sociological realities. The pessimistic conclusion that the disputes of metaethics are intractable presupposes intimate acquaintance with these disputes and a deep understanding of the potential of each possible moral theory. In what sense, then, is it correct to say that justice theorizing can bypass moral theory? And how plausible is it for Rawls to imply that all justice theorists employing WREM must emerge with the pessimistic conclusion that no moral theory shows sufficient promise to become widely accepted as our most rational understanding of morality? Certainly not all justice theorists, applying the method of reflective equilibrium the way Rawls' instructed them to do in *A Theory of Justice*, have ended up with Rawls' skepticism about the prospects of finding some measure of agreement about the best understanding of morality by ruling out some currently popular metaethical positions as rationally unacceptable.

(2) *The pragmatic paradox:* But suppose Rawls is right, and WREM forces the pessimistic conclusion on us that we cannot hope to resolve the conflict between

rival metaethical position or to win public acceptance for any such resolution. Would it follow that we must then adopt a purely pragmatic justification of principles of justice and accept whatever principles promise to work best, relative to a given society, in producing an "overlapping consensus"? It seems to me that the attempt to give a purely pragmatic justification for principles of justice confronts a somewhat analogous paradox to that which haunts William James' attempt to give a pragmatic justification of religious belief. J. L. Christian explains the pragmatic paradox as follows:

> A human condition in which an idea must be believed to be true in terms of correspondence - that is, one must be convinced that some object/event exists as a real entity - before the idea can be considered to be true on the pragmatic test. For example, one must believe that immortality exists as a real event...before the belief in immortality can produce positive results in his life. (596)

Let me make clear the analogy I see between the paradox faced by James and Rawls. Just as James' pragmatic justification of religious belief will have little or no success in instilling such belief in the absence of support from antecedently held religious convictions, so Rawls' pragmatic justification of principles of justice will have little or no success in motivating agreement on these particular principles in the presence of conflicting moral convictions. A pragmatic justification of principles of justice, if understood as *merely* pragmatic, will not have sufficient normative force in the minds of people to override their antecedently held beliefs about justice. People's beliefs about justice are generally nourished by deep moral and often religious commitments. It would seem fairly obvious that even the most pragmatic conception of social justice *must* claim moral authority to overrule the dictates of *those* moral and religious commitments which conflict with any such conception; and to that extent (which may be considerable) it cannot avoid taking sides in metaethical disputes. In ethics, what counts as a *sufficient* justification always depends on what is the best justification someone thinks he can get; and if someone

thinks that he can directly intuit the truth about social justice with absolute certainty, or simply look it up in some Holy Book, he will hardly take great interest (unless he has no choice and has already become convinced of the virtue of liberal tolerance) in Rawls' pragmatic justification for a set of principles reached by a complicated process of wide reflective equilibrium, let alone accord them overriding moral authority.[15] Only insofar as Rawls' reflective equilibrium arguments are acknowledged to have genuine moral force, therefore, will they persuade competing interest groups in society to strike a bargain that does not simply reflect power differentials and threat advantages. So when Daniels claims, "I think it is a virtue of reflective equilibrium that it leaves open metaethical considerations"(282), he seems to neglect seeing the issue from the perspective of those with strong metaethical commitments.

(3) *The indeterminacy of pragmatic justifications*: Is it really plausible to think WREM necessitates regarding all and only those principles of justice as rationally acceptable which are justifiable in a purely pragmatic way? Are there any such principles at all? And if there are, do they resemble anything recognizable to people as giving each of them *justice*? And how convincing is Rawls' claim that his two principles would be singled out from among a host of possibilities by a pragmatic justification?

Two decades of criticism of Rawls' theory have made it clear that the search for a purely pragmatic mode of political accommodation is highly unlikely to converge on his particular principles of justice. As Arneson stresses, Rawls' hope of some core consensus in democratic society seems unduly optimistic:

---

[15] Rawls only touches on this objection to his approach: "One might say....that to develop a political conception of justice without presupposing, or explicitly using, a metaphysical doctrine, for example, some particular metaphysical conception of the person, is already to presuppose a metaphysical thesis: namely, that no particular metaphysical doctrine is required for this purpose."(1985, 240 footnote) His reply: "Following the method of avoidance, I should not want to deny these claims."(240). But *can* Rawls legitimately invoke the method of avoidance in answer to the charge that the claim of being able to avoid controversial meta-ethical assumptions is itself a controversial meta-ethical assumption?

...the Kantian ideal [of persons who give priority to preserving the conditions for the exercise of their moral powers to choose and to cooperate with others on fair terms] is present in democratic culture - but so are several other firmly rooted conflicting ideas, such as (a) competitive individualism, (b) patriarchal conservatism, and (c) Judeo-Christian benevolence....Their adherents would....politely decline the invitation to join in an overlapping consensus on Rawls's egalitarian liberal principles of justice or anything remotely close to them. (705).

The argument at this point in the debate becomes admittedly somewhat subtle and elusive, and it is not clear who is begging the question against whom. I take Rawls to be arguing that a pragmatic accommodation is in everyone's interest in precisely those situations where any particular group's attempt to get the better of all others would have disastrous consequences for all groups, and that his two principles of justice spell out the most rational terms of this accommodation. Arneson can retort that there are many possible modes of accommodation that are pragmatically rational, some based purely on existing power differentials and threat advantages, others - such as Rawls own proposal - laden with more or less controversial moral assumptions; and that Rawls has failed to give the various interest groups a compelling rationale to converge on his two principles as the terms of their mutual accommodation. Certainly, if history is the relevant test here - and for a thoroughgoing pragmatist it would seem to be ultimately the only relevant test - it seems clear that Rawls' theory has made few converts in the two decades of its existence, even among serious students of his work.

§4 It seems to be Rawls' view that, since we are not going to get agreement on a *moral* justification for a particular theory of justice, the task of the political philosopher at the present time is restricted largely to a sort of damage control in the face of this fact (cf. 1985, 227). But Rawls' radical pessimism about the prospects of some modest measure of moral agreement among informed people of good will seems to be at least as controversial as Rawls' optimism that, in the *absence* of moral

agreement, we can still get pragmatic agreement on his two principles of justice (or on any other usefully determinate set of principles). Therefore, Rawls pragmatic justification for his particular theory of justice can be considered to have failed on its own pragmatic terms. But even if Rawls is right that no agreement on a moral justification for a theory of justice is to be expected at the present time, it would not follow that the political philosopher is simply restricted to the role of mediator between interest groups, rubberstamping whatever forms of mutual accommodation may be feasible in a particular society at a particular time. If a strong reflective equilibrium-based case can be made for accepting a broadly consequentialist moral framework, and if this framework turns out to support a liberal social order, not simply as providing pragmatic terms of mutual accommodation, but as embodying morally justified principles of justice, then a liberal social order is quite likely to win the allegiance of many thoughtful people. Such allegiance, while falling far short of universal agreement, may contribute substantially to keeping a liberal regime stable.

The consequentialism I defend stands in sharp and obvious contrast to the sort of *modus vivendi* pragmatism about a just social order which tries to be agnostic about the possibility of a rational moral theory and wants to leave the clutter of clashing moral dogmatisms in place. A pragmatism of this kind seems likely to collapse into mere power politics. Even though accepting a consequentialist framework for moral reasoning will not *abstractly* fix the demands of justice in a very precise general way, it does make the reasons for the uncertainties inherent in principles of justice at the fundamental level clearly intelligible. If appreciated as perhaps our most credible alternative to moral skepticism and *Realpolitik*, a broadly consequentialist understanding of morality may motivate people to settle many disagreements about justice at the level of negotiated compromise instead of allowing them to degenerate into raging confrontations.

PART 3: CONSEQUENTIALISM AND RATIONAL DISAGREEMENT ABOUT JUSTICE

3.1 DISAGREEMENT ABOUT JUST-MAKING CRITERIA

After clarifying the concept of social justice and other key concepts, this chapter considers two central disagreements associated with the cluster of considerations relevant to determining justice: the criteria definition problem and the criteria inclusion problem.

3.1.1 The concept of social justice

§1 The concept of "justice" is almost invariably invoked whenever individuals or groups advance claims to society's resources. As a first step toward clarifying this concept, we must obviously be able to establish some parameters of relevance, some basic agreement about what we take people to disagree *about* when they disagree about justice, some widely assumed core meaning of the concept of justice as opposed to competing interpretations of this concept.[1]

The common assumption most people seem to make when they argue about what is just or unjust is that certain valued things are certain people's *due* or are *owed* to them for a variety of possible reasons, and that social distributions of those valued things should try to give individuals their due. To say that the core meaning of "justice" is "to give everyone their due" is unhelpful, however, unless the notion of something being someone's *due* is clarified. For what reasons, or on what basis, do people typically claim that something is their due?

---

[1] It is not a foregone conclusion that there *must* be a common idea unifying all uses of this concept to enable people to argue meaningfully about what justice requires. It is quite plausible that the concept of "justice" can be understood in the manner of Wittgensteinian family resemblance. Such an analysis of the concept allows room for disagreement about some of the exact items which make up the cluster of just-making criteria, even if these criteria are regarded as constitutive of the core meaning of "justice". There is also the more radical possibility that "justice" is used equivocally - that there are really several senses of this term linked not even by family resemblance. But the sense of justice that I am interested in, that defines my subject matter here, is only that invoked (i) by individuals claiming something as their due, and (ii) by attempts to work out theories about how best to respond to such claims.

This question is perhaps best approached negatively. People typically think that they have not been given their due (and speak of injustice) when it seems to them that, in the distribution of valued things, individual merits or achievements have been overlooked, needs disregarded, arbitrary inequalities introduced, rights violated, or expectations of reciprocity disappointed.

In general we could say that considerations of justice arise in contexts where there are benefits or burdens to distribute, and where there are too many claimants (for the benefits) or not enough volunteers (for the burdens). Typically, arguments to decide the justice of an individual's bundle of benefits and burdens appeal to a rather vaguely demarcated cluster of evaluative considerations, such as merit or desert, needs, equality, rights or expectations held in good faith. I shall call claims based on such considerations "justice-based claims" (without thereby implying that justice requires every such claim to be satisfied). In a preliminary way, then, we can define the core meaning of "social justice" as "appropriate responsiveness to the claims which members of a politically organized society make - as individuals or as groups - to a share of social resources in pursuit of their well-being, by appealing to desert, need, equality, rights or good-faith expectations."[2]

I now want to briefly clarify a set of important ideas: the distinction between a theory of justice and a mere conception; the "problem of social justice"; and the two separate levels at which I think the problem of social justice must be conceptualized. By a *theory* of justice I mean a systematic conception of justice which includes

---

[2] We can usefully distinguish justice at the personal, the social and the global level. *Personal* justice applies ideas about justice to the distribution of goods in small-scale, usually face-to-face contexts. (I follow the practice of using the term "goods" for both benefits and burdens.) *Social* justice applies ideas about justice to the context of goods whose distribution is thought to be a feasible and proper concern of political regulation. *Global* justice would go even further and apply ideas about justice to the distribution of goods among all people or even all sentient beings in the world. Although a person's initial understanding of the concept of justice is likely to arise out of experiences in the realm of personal justice, the transfer of our understanding between personal, social and global justice is clearly not all one way. The concerns of justice at these three levels grade into each other, but as will become clear, it would be a mistake to think that the distinction between them simply marks an expanding continuum involving only a quantitative increase in complexity (cf. especially Sec 5.1).

standards to assess the justice of important social distributions in a principled way, along with proposed justifications for these standards, and more or less detailed instructions for how to implement or credibly approximate them in practice. By a *conception* of justice I mean *any* interpretation of what justice requires, no matter how confused, that goes beyond the core meaning of justice as defined above. A conception of justice, at its simplest, may be embodied in judgments about specific cases, or in principles at varying levels of generality and determinacy, or - at its most complex - as a well-elaborated, comprehensive theory. The *problem of social justice,* which theories of justice try to solve, is how to resolve disagreements about social distributions in a way that has potential to become acceptable to rational people as an appropriate response to their justice-based claims and to lead to stable terms of mutual accommodation.

It is important to distinguish two fundamental levels of the problem of social justice, because many of the disputes about justice stem from confusion between these two levels and failure to appreciate the difficulties of achieving a fit between them. Let me call these the level of "feeling injustice" and the level of "doing justice", and explain briefly what happens at each level. (The asymmetry of terms here is intended, because I want to make the point that doing justice does not necessarily imply avoiding injustice, in every plausible sense of that term.) *At the level of feeling injustice,* individuals have strong expectations about what they can claim as their share of social resources in particular contexts, and they tend to feel anger, resentment or indignation (a "sense of injustice") when such expectations are disappointed. By the familiar process of Humean sympathy, personally unaffected bystanders may, of course, experience much the same sense of injustice as the actual victims of perceived injustice.

The expectations and sentiment about justice of all individuals *collectively* form a pattern of often conflicting justice-based claims to a society's resources. *At the level of doing justice,* there usually exist a range of individual, social and political

alternatives about how to respond to the pattern of justice-based claims in such a way as to achieve a balance among these conflicting claims in a particular domain. No matter which alternative is chosen, *some* individuals' justice-based expectations are almost always bound to be disappointed. Often the expectations which get disappointed are every bit as legitimate as the expectations which get met. One way to put this point would be to say that some injustices will usually be unavoidable in the process of doing justice. We could, of course, insist that the term "injustice" should not be used for what can plausibly be called "injustice" from the perspective of an individual, but only for what can plausibly be called "injustice" from some perspective that takes *all* justice-based claims of *all* individuals into account. One of the major challenges faced by a theory of justice is, in fact, to give usefully determinate conditions for diagnosing an injustice, and I shall raise objections both to making these conditions too permissive and to making them too restricted (Sec 3.3.2).

§2 As defined above, "doing justice" consists in an appropriate response to the claims to a share of society's resources which individuals make in pursuit of their personal well-being, by appealing to such criteria as desert, need, equality, rights or good-faith expectations. But any substantive claim about which criteria are just-making considerations is bound to be controversial.

Several distinctions are necessary to clarify the level of analysis in this chapter. The first and fairly obvious distinction is between those types of considerations that are held to be *sufficient* to judging justice in a specific domain and those types of considerations that are held to be *relevant* to making such judgments in general. I want to call a "just-making criterion" a consideration which is generally relevant to making judgments about justice. In contrast, I want to call a "principle of justice" a statement that spells out sufficient conditions for determining justice in some domain.

A note about the relationship between *criteria* and *principles*, as I define them here. A principle may obviously contain more than one criterion. The role of just-making criteria in a principle of justice is *either* to specify conditions a person must meet to justly claim certain goods *or* to specify conditions which a distributive outcome or process must meet to be just. What someone regards as the "demands of justice" results from the set of all judgments that this person is willing to make about what things are just or unjust on the basis of applying his interpretation of all just-making criteria to all distributive domains. The task of justice theorizing is to reduce, as far as possible, the degree of rationally permissible intersubjective variability in identifying the demands of justice.

We also need to distinguish *commonly invoked* just-making criteria from *ideal* criteria. Commonly invoked just-making criteria encompass a fairly standard list: (1) desert; also referred to as "merit" or "equity", and variously interpreted as ability, effort, contribution or achievement; (2) needs; (3) equality, in the sense of non-discrimination on the basis of specified traits; (4) rights; and (5) good-faith expectations. Fairly clear examples of what are commonly regarded as non-justice considerations are such political values as national security or prestige, economic productivity, or environmental conservation. It will be convenient to call demands based on just-making criteria "justice-based claims". But even if ordinary people draw the distinction between justice-based and nonjustice-based claims in a particular way, it obviously does not follow that a moral theorist cannot propose the *inclusion* of additional considerations in the list of just-making criteria or the *exclusion* of commonly invoked considerations from this list. From the fact that all individuals in society, or members of certain groups, invoke a particular list of criteria when they make claims to society's resources, it does not follow that a theory of justice will include *all* and *only* these criteria in formulating its proposal for what should count as an "appropriate response" to the overall pattern of expectations and sentiment about justice.

We can distinguish three connected sources of disagreement about just-making criteria. There is controversy over how to define the exact nature of such commonly invoked criteria as desert, need, equality, rights or good-faith expectations; controversy over exactly which items should be included in the list of just-making criteria; and controversy over what to do when different criteria either presuppose unavailable resources, specify conflicting uses for available resources, or conflict in other ways. I want to call the first issue *the criteria definition problem*, the second *the criteria inclusion problem* and the third, following Rawls, *the (criteria) priority problem* (cf. 40-45).

Each commonly invoked just-making criterion would warrant a book, and numerous books have, in fact, been written about each. My own brief analysis in this chapter aims only to clarify the basic nature of the criteria definition and inclusion problems to set the stage for the vastly more complex analysis of principles of justice in the next chapter. But two major objections to my way of proceeding need to be addressed right away: the objection that analyzing justice in terms of a complex of just-making criteria is circular; and the objection that rights are not plausibly regarded as one of several criteria or grounds of justice, but form an independent though perhaps overlapping moral category.

§3 First, the circularity objection. As mentioned, each just-making criterion raises the question of its own criteria for being justifiably invoked or properly applied. Answering *this* question takes us back to justice. So analyzing justice in terms of a list of just-making criteria is circular.

My claim is not, however, that we only need to enumerate a list of common just-making criteria, and we will know what justice requires. My claim is that the immensely complex idea of justice has normative constituents which are meaningfully distinguishable in terms of their separate functions in our moral vocabulary, and that exploring disagreements about them is a helpful way of

organizing our initial inventory of fundamental sources of disagreement about justice. Of course, a definitive normative account of what can legitimately be claimed on the basis of desert, need, equality, rights or good-faith expectations must await an adequate theory of justice. But in the absence of such a theory, and as a way of working towards it, how else could we get our bearings than from widespread understandings and deeply ingrained sentiments about the meaning of various just-making criteria? These understandings and sentiments, like the concepts that have evolved to express them, are of course riddled with inconsistency, vagueness, and other complexities. What could a (consequentialist) account of justice be, if not essentially a proposal for sorting out those complexities and inconsistencies in such a way that the point of morality is preserved? Morality's point - creating enough predictability and uniformity in people's behavior so as to generate the necessary sense of security and trust for beneficial mutual interaction - can *best* and in fact *only* be preserved, if the claims of individuals in pursuit of their personal well-being are taken very seriously. Taking them seriously, that is, achieving a reasonable compromise among the often conflicting demands of just-making criteria, will be a large part of the substance of any consequentialist morality.

§4 Second, the rights objection. Is it really plausible to consider rights to be merely one just-making criterion among several, or are rights perhaps best considered to form a separate moral category? What exactly is the relation between justice and moral rights? Which category is wider in scope - are injustices one possible form of rights-violations, or are rights-violations one possible form of injustice? Do these categories perhaps have overlapping scopes? Or yet another possibility: are these categories, though conceptually and functionally distinct, perhaps extensionally equivalent?

In our attempt to sort out conceptual (and associated normative) confusions, we must be very clear about what our analysis is supposed to accomplish. Obviously,

the task of conceptual analysis may be approached very differently by those who think that distinctions, such as that between justice and the rest of morality, or between justice and rights, are somehow carved into the nature of things or come to us from some Platonic realm of essences, and those who think that the ordinary usage of moral concepts and distinctions reflects a confused mix of irreducibly conventional, prudential, and historically or otherwise contingent elements. If one subscribes to the latter view, as I do, the question then arises whether one should approach conceptual analysis with "reporting" or "reforming" intentions. Is the analysis meant to be faithful to ordinary usage and trace out its inconsistencies, or is it meant to stay only as faithful to this usage as is compatible with the goal of introducing greater coherence into our moral vocabulary? My objective here is again the latter: to differentiate between moral concepts with a consequentialist rationale in mind, by trying to assign to each that rationale that seems to me to cohere best with ordinary usage. Specifically, I want to assign rights-talk a place that makes sense in a coherent consequentialist conceptual scheme. If we take this approach, it seems certainly implausible to subsume all justice-based claims under the category of rights. As Sumner remarks, "Though rights will surely be part of the story in any adequate theory of justice they are unlikely to be the whole story"(1987a, 136).

The concept of justice and that of rights can be functionally differentiated in terms of their consequentialist rationale. Rights-claims (from a consequentialist perspective) are fundamentally based in the recognition that exempting certain types of choices or interests of individuals or groups from the well-being test on every occasion will promote the greatest well-being in the long run.[3] Not all justice-based claims are, however, reducible to demands for the protection of types of interests or

---

[3] Although I shall develop my account of rights in terms of *individual* rights, there is nothing in the fundamental consequentialist rationale for rights that would seem to preclude the possibility of *collective* rights. Which collective rights particular social subgroups should have, or whether we should accept irreducibly collective rights at all, will of course again be subject to the consequentialist well-being test.

choices which, if recognized and socially protected as rights, would promote the greatest well-being overall. It is an open question if any particular claim based on desert, need, equality or good-faith expectations should be accorded the normatively potent (though not absolutely inviolable) status of a right. It is also an open question whether a particular right has anything to do with claims based on desert, need, equality or good-faith expectations (see below).

Campbell comments helpfully on the relation between moral rights and social justice:

> There are in life many minor injustices whose rectification may be very much an optional extra(20)....it can be argued that rights have a narrower scope than justice. Rights seem most at home in limited areas where individual interests are protected by definitive rules....This may be important for the rectification of injustices done to individuals, but it does not seem to have the same foothold where matters of wider collective or 'social' justice are concerned. Unjustified economic inequalities, the absence of educational opportunity and discrimination in employment are all grave social injustices, but they are not objectives whose attainment can always readily be effectively pursued by means of legally enforceable entitlements.... (37-9)

Although Campbell speaks of "legally enforceable entitlements" here, I think he clearly means to refer to *moral* rights protected by legal means rather than to purely legal rights in the conventionalist sense. Moral rights call for whatever social protection promises to be most effective (usually legal enforcement).

Of course, rights-claims will pick out *many* of the same types of actions for protection as claims based on desert, need, equality or good-faith expectations, although the degree of protection may be quite different. A consequentialist understanding of *moral rights* arises, we said, from the recognition that protecting certain types of actions from being subject to the well-being test on each occasion will, in the long run, most effectively promote well-being. A consequentialist understanding of *social justice* arises from the recognition that promoting well-being requires responsiveness to those claims to social resources which individuals make in pursuit of their personal well-being on the basis of criteria which they

allow others to use as a basis for similar claims and which they are capable of backing up with some of the fiercest sentiments known. Some of the justice-based claims people make will clearly be such that giving them rights-status will promote well-being in the long run. But clearly, some claim might satisfy the rationale for justice but not for rights.

Could the reverse also be true - so that the scope of rights-claims would *overlap* the scope of justice-based claims? In this case, some but not all rights-claims could be considered just-making criteria. I think the answer is no: consequentialists can safely subsume all rights-claims under the scope of justice-based claims. Rights-violations can generally be regarded as one possible form of injustice.

On the other hand, I think the temptation of regarding *all* injustices as necessarily rights-violations should be resisted. Given the function of rights in our moral vocabulary, a rights-based claim is meant to be normatively more potent, and capable of overriding, other moral considerations. Claims based on desert, need, equality and good-faith expectations will become rights if, and only if, giving them rights-status is likely to promote the greatest well-being overall. Certain needs-based claims may qualify for consequentialist rights-status in *any* society that can afford to meet these needs, but most justice-based claims will not qualify for rights-status, because that would give rise to far too many conflicting rights. This is not to deny, of course, that practically all justice-based claims and, in fact, all sorts of completely frivolous claims are popularly advanced in the normatively debased *language* of rights (cf. Sumner, 1987a, 2-4). But as consequentialists we have a fairly stringent criterion for generating rights which will disqualify many popular rights-claims.

It is plausible to think that all *rights*-violations constitute injustices, but clearly less plausible to construe every failure to satisfy *other* justice-based claims as an injustice. In spite of widespread readiness to pronounce all sorts of things unjust, the problem of finding plausible necessary and sufficient conditions for what constitutes an injustice turns out to be extremely complex (cf. Sec 3.3.2). If

consequentialists do not want to deprive the concept of "injustice" of its normative force by legitimizing its trivialization, then they cannot permit pronouncing failure to satisfy *every* claim based on a just-making criterion unjust, but only failure to satisfy those which survive an appropriate process of balancing the competing justice-based claims within some appropriately demarcated domain. But the function of according rights-status to *certain* justice-based claims is precisely to short-circuit the unpredictable vagaries of this balancing process to a large extent, and doing so may indeed be defensible on consequentialist grounds.

So by assigning rights a place among my proposed ordering scheme for the unruly family of commonly invoked just-making criteria I do *not* mean to deny that my list (or any such list) contains important asymmetries, considerable mutual overlap and a sort of normative pecking order. Claims grounded in basic needs, for example, will generally have far greater moral weight than mere good-faith expectations; and different needs and expectations will in turn have varying degrees of moral weight. But while I am prepared to admit that the normative weight of rights-claims usually allows them to override other justice-based claims, and that rights-violations should usually count as injustices, as a consequentialist I cannot, of course, go so far as to accord rights-claims *infinite* moral weight. A person's having a right *may* become compatible, morally, with her being prevented from exercising it, if the consequences of exercising this right would be sufficiently disastrous.[4]

It is plausible, we said, that all rights-violations should count as injustices (and since rights can conflict, that obviously means injustices can sometimes become morally permitted). But should consequentialists affirm the converse as well: should all injustices count as rights-violations? Should we perhaps regard

---

[4] Narveson seems to hold a wildly implausible absolutism about rights when he says, "If my right to drive a car is compatible morally with your forcibly preventing me from exercising it, then there is no point whatever in speaking of 'my right'"(1991, 331).

rights as the *cut-off criterion* to decide when disregard for some justice-based claim becomes an injustice?[5]

It is, of course possible, to use the term "injustice" in a sense such that a failure to satisfy a justice-based claim is called an injustice *only* if it constitutes a rights-violation. I want to reject this usage for three reasons: (1) Since there may be situations in which none of the competing justice-based claims have the normatively potent status of moral rights, this usage would mean that, in such a situation, it would be conceptually impossible for an injustice to be committed. (2) The cumulative weight of such justice-based claims of many individuals may, on occasion, even outweigh some individual's recognized moral right at some point, although rights *usually* will trump other justice-based claims. So it seems implausible to regard a violation of someone's rights as a necessarily sufficient condition for the occurrence of an injustice, as a "cut-off criterion" which would somehow specify when failure to satisfy a claim based on another just-making criterion becomes an injustice. (3) We could, of course, postulate a *particular* sort of right, the right never to be unjustly treated, and indeed many people want to claim such a right - but I have three consequentialist objections against recognizing such a right. *First*, postulating such a right would further debase the normative currency of rights, for two reasons: (i) it would make the actual existence of this right as indeterminate as the diagnosis of an injustice (and there are good consequentialist reasons for wanting to assign the normatively potent status of rights only to fairly determinate types of claims); and (ii) rights-violations would become as numerous as injustices (and there is a fairly clear consequentialist rationale for using the term "injustice" in a sense such that many injustices are morally defensible, either because they are unavoidable or because they are on balance morally preferable (see below). *Second*, postulating a general right of everyone not to be unjustly treated

---

[5] I owe this question to Michael McDonald, along with numerous other questions which greatly helped me clarify my position.

would be useless *as a cut-off criterion*, because invoking it would already presuppose that we know how best to balance the competing justice-based claims in some domain. Our consequentialist rationale for generating rights would be far too indeterminate to be up to the job of making rights a useful cut-off criterion for when to consider a failure to satisfy a justice-based claim an injustice. Certainly there seems nothing *conceptually* necessary about rights taking on this job. *Third*, making every injustice at the same time a rights-violation - a violation of the alleged "right not to be unjustly treated" - (after we already granted that every rights-violation can plausibly be regarded as an injustice) will further contribute to the conflation of these two functionally distinct concepts and to the resulting conceptual and normative confusions. It is an understandable attempt at securing double protection against being victimized by injustice; but raising every injustice to the status of a rights-violation is simply bound to debase the normative currency of rights-claims. It is true that this analysis creates a situation in which someone can be done an injustice, and yet I do not want to say that there is a failure to respect a right. But that is precisely what, from a consequentialist perspective, seems to me desirable, given the relatively subordinate moral weight which I think we must assign to many of the things individuals want to call, and should legitimately be able to call, "injustices".

From a consequentialist point of view, I see the division of conceptual labor between morality, justice and rights somewhat analogous to the division of academic labor between a university as a whole, the faculty of humanities and the department of philosophy, but with lots of cross-appointments of faculty members in different departments within the humanities. I see nothing inconsistent about saying that, regrettably, someone's rights have to be violated to avoid a major injustice. It *might* seem inconsistent for me to say, however, that even a minor injustice has to be committed in order to respect someone's rights because, on my account, rights are part of the input into the process of balancing competing justice-

based claims, and if other justice-based claims lose out over a particular right in a particular case where the appropriate process has been followed, then no injustice would seem to have been committed. But this would be a hasty conclusion. There are reasons for wanting to use the term "injustice" in a sense in which it is legitimate for us to admit that we may have had to commit an injustice against certain individuals, even if overall we have achieved the morally best result possible under given circumstances (see below).

In summary, then, rights seem to have all the requisite properties of a just-making criterion, but clearly not every claim based on a just-making criterion gives rise to rights. And considering rights to be a criterion among others on a list of overlapping, complex just-making criteria is quite compatible with admitting the special normative force of rights.

3.1.2 The criteria definition problem

Next we must wrestle with the criteria definition problem. It is tempting to pursue in detail the intricacies and interconnections of such complex concepts as "desert", "needs", "equality", "rights" and "good-faith expectations", but doing so would make us lose the main thread of our inquiry. I shall, therefore, restrict myself to a few clarifying remarks about each criterion from a consequentialist perspective and mention some plausible points of rational disagreement.

§1 Desert

Claims based on desert are central to real-life disputes about justice. What are desert-claims based on? The bases of desert, at least as understood in Western societies, are well analyzed by Sher, who differentiates six classes of desert-claims according to their justificatory basis (1987, 150-1). The first two classes include desert-claims based on *traits of a person*. They can be non-moral traits. It makes perfect sense to say, for example, that "The best-qualified applicant deserves the job". Or

they can be moral traits, as when we say, "The virtuous deserve to be happy." Sher's other four classes include desert-claims based on *features of a person's actions*. People's sense of desert is affected by such factors as whether the action involves effort, involves competition, is freely willed, or is morally blameworthy. Again, it makes perfect sense to say that "People deserve the fruits of their efforts", "People deserve prizes won in competitive performances", "People deserve the foreseeable consequences of their freely-willed actions", or "People deserve punishment for morally blameworthy actions" (cf. Sher, 1987, 150).

A consequentialist understanding of desert arises from the recognition that recognizing individuals' claims to social resources on the basis of conscientious efforts which serve a socially beneficial purpose will effectively encourage such efforts. For a consequentialist, then, there is no such thing as "desert" independently of a conception of what sorts of results count as socially beneficial. In merit-based hiring of job applicants, for example, what counts as merit is not fixed once and for all, but will change with our changing understanding of the social purpose that we want a particular profession or institution to serve. There will often be heated disputes, both because there can obviously be disagreement about the proper social purpose of an institution or profession, and also because job applicants who have worked hard to measure up to established criteria of merit have good-faith expectations that these criteria will not suddenly be changed.

One fundamental dispute about desert that seems to go on without resolution, at least in the philosophical literature, is disagreement about the very coherence of desert-claims in a world of causes. Sidgewick brought this issue into sharp relief. Asking whether rewards for a person's services should be apportioned to the effort made or to its utility for others, he argues that

> ...the actual utility of any service must depend much upon favourable circumstances and fortunate accidents, not due to any desert of the agent: or again, may be due to powers and skills which were connate or have been developed by favourable conditions of life, or by good education, and why should we reward him for these?....And certainly it is only in so far as moral

excellences are exhibited in human achievements that they are commonly thought to be such as God will reward. But by drawing this line we do not yet get rid of the difficulty. For it may still be said that good actions are due entirely, or to a great extent, to good dispositions and habits, and that these are partly inherited and partly due to the care of parents and teachers; so that in rewarding these we are rewarding the results of natural and accidental advantages, and it is unreasonable to distinguish these from others, such as skill and knowledge, and to say that it is even ideally just to reward the one and not the other. Shall we say, then, that the reward should be proportionate to the amount of voluntary effort for a good end? But Determinists will say that even this is ultimately the effect of causes extraneous to the man's self. (283-4)

Exactly this line of argument is invoked by Rawls to exclude desert altogether as a relevant criterion for determining the justice of society's basic structure. Rawls defends his skepticism about desert in trenchant words:

It seems to be one of the fixed points of our considered judgments that no one deserves his place in the distribution of native endowments, any more than one deserves one's initial starting place in society. The assertion that a man deserves the superior character that enables him to make the effort to cultivate his abilities is equally problematic; for his character depends in large part upon fortunate family and social circumstances for which he can claim no credit. (104)[6]

In more explicit form, Sidgewick's and Rawls' argument against including desert among just-making criteria could be summarized as follows:

1. A person's achievements are the result of ability, effort or circumstances.
2. A person's circumstances are undeserved.
3. A person's ability and effort are wholly the result of his endowments or motivations.
4. A person's endowment and motivations are wholly the result of heredity and circumstances.
5. A person's heredity and circumstances are undeserved.
6. If a person's achievements are wholly the result of undeserved factors, then those achievements are not deserved.
   Therefore, a person's achievements are not deserved.

---

[6] But Rawls sometimes vacillates in his deterministic assumptions. In "Social Unity and Primary Goods", for example, Rawls says: "As moral persons citizens have some part in forming and cultivating their final ends and preferences...But to argue this [that it is unreasonable to hold people responsible for unduly expensive tastes] seems to presuppose that citizens' preferences are beyond their control as propensities or cravings which simply happen. Citizens seem to be regarded as passive carriers of desires. The use of primary goods...relies on a capacity to assume responsibility for our ends."(168-9) I do not see how these statements could be made consistent with Rawls's radical skepticism about desert.

The fourth premise of this argument can be made extremely plausible with the help of empirical support from biology, psychology or sociology. It represents, of course, the familiar determinist thesis that a person's actions are merely links in an unbroken causal chain which started long before that person was born. From the thesis that no one can ever help doing what he actually does, it is tempting to conclude that no one ultimately deserves anything. But once the conclusion that ultimately no one deserves anything is combined with another alleged demand of justice which many find intuitively plausible, namely that people should be compensated for undeserved inequalities, a strongly persuasive chain of reasoning results:

     1. Justice requires that people should be compensated for undeserved inequalities.
     2. All inequalities are ultimately undeserved.
     3. So justice requires that all inequalities should be compensated.
     4. But justice also requires that the position of the worst-off group in society should be improved.
     5. Some economic inequalities may be necessary to improve the position of the worst-off group in society.
     Therefore, justice requires that all inequalities should be compensated, except when economic inequalities are necessary to improve the position of the worst-off group in society.

These arguments (together with arguments about the "priority of liberty") lie at the heart of Rawls' theory, and many deep rifts in contemporary justice theorizing can be traced back to disagreement about some premise or other of these arguments. Let me only give the briefest synopsis here. Nozick, for example, forcefully challenges Rawls' assumption that whatever results from something undeserved is itself undeserved. Nozick argues that unequal wealth may well be deserved, in the sense that no one is entitled to take it away from a person, if it results from something which this person simply has (such as natural endowments)

and which no one else is entitled to have (cf. 224-7).[7] Dworkin disagrees with Rawls that what people achieve as a result of effort and ambition is just as undeserved as what they achieve as a result of endowments, and therefore proposes social distributions which are, as far as possible, "endowment-insensitive but ambition-sensitive"(1981, 311). Most of Rawls' critics have argued that improving the position of the worst-off group in society is plausible neither as a necessary nor as a sufficient condition for identifying morally acceptable economic inequalities (see Ch 3.2). Some have denied that it is reasonable to prohibit tradeoffs between political liberty and economic gains when the loss of liberty would be small and the economic gains large (e.g. Barry, 1973.). Some have regarded the idea that people should be free to exercise their talents in the market as a much more plausible basic intuition about justice than that people should be compensated for undeserved inequalities (e.g. Nozick, Narveson), but many possible intermediate positions have also been defended.

There is near universal consensus in the literature on social justice that reason neither requires nor perhaps even permits us to accept Rawls' principles of justice. My aim here is not to trace the details of this protracted debate, which has been ably chronicled in a number of works, but only to locate a basic disagreement about desert.[8]

We seem left with essentially three alternative interpretations of "desert" about which rational people may plausibly disagree (even though perhaps with varying degrees of rationality): (1) No one ever really deserves anything because the capacities underlying a desert claim are themselves wholly undeserved. (2) People deserve things in proportion to their justified desert claims, even though these

---

[7] It is sometimes claimed that adopting an original position will necessarily blunt claims of desert. But the contrary seems to me true. Bargainers in the original position who don't know their individual endowments, but know enough about human nature to appreciate how demoralizing and resentment-provoking neglect of desert-claims is, will make sure to include a provision to reward desert in their principles.

[8] E.g. Kymlicka (1990); Mapel (1989); Campbell (1988); Pettit (1980).

claims are based on wholly undeserved capacities. A person deserves rewards on the basis of ability and effort, not because these are the result of something which is itself deserved, but because they are manifestations of who a person is and of values we want to socially uphold. A person's endowments and motivations are within the inner boundaries of his self, as well as expressing socially valuable qualities. (3) People deserve things in proportion to the extent that their effort and achievements result from the exercise of free will, properly construed. A person's abilities, effort or motivations are more than simply products of heredity and circumstances; they escape complete causal determination in the relevant "desert-undermining" sense.

The disagreement between those who hold (1) or (2) and those who hold (3) strikes me as remediable only if one of the knottiest philosophical conundrums, the free will problem, could be satisfactorily resolved. From a consequentialist point of view we want to say that, whatever the deep truth about human free will may be, recognizing claims of desert, in the sense of effort with socially beneficial results, promotes such effort, and therefore promotes well-being and is rationally justified.

§2 Need

Need has such a central place among commonly invoked just-making criteria, and such strong support from people's sense of justice, that theories of justice which exclude need (such as libertarian theories) provoke passionate opposition for that reason.

If "need" is to be a credible just-making criterion, however, two senses of this concept have to be kept apart. In the *relational sense*, need is simply a relation between means and ends: "Person A needs good B to achieve goal C." What someone needs, in this relational sense of "need", is purely a matter of finding the proper means to satisfy a person's preferences, no matter how frivolous they may be. Clearly, the relational sense of "need" is useless as a just-making criterion.

Braybrooke has tried to rehabilitate the concept of need in a sense in which it can be said that a person A needs goods of type B, not to achieve just any goal, but to facilitate that person's basic physical and social functioning. Braybrooke calls such needs "course-of-life needs". Course-life-needs are needs which, unlike preferences, human beings have in common. In distinguishing *such* needs from preferences, Braybrooke says,

> "...while it must be acknowledged that in preferences other people may be so
> different as to be hardly understood, much less sympathized with, needs
> unite human beings in their conception of themselves."(1987, 237)

Braybrooke lists the following "course-of-life" needs:

Needs based on physical functioning: the need to have a life-supporting relation to the environment; the need for food and water; the need to excrete; the need for exercise; the need for periodic rest, including sleep; the need for whatever else may be indispensable to preserving the body intact.

Needs based on social functioning: the need for companionship; the need for education; the need for social acceptance and recognition; the need for sexual activity; the need to be free from fright and harassment; the need for recreation. (Cf. 1987, 36)

Braybrooke is also careful to distinguish these "matters of need" (which every person has to some extent) from "minimum standards of provision" (which reflect the actual extent of each need, and which will vary from person to person and for the same person at different times). Braybrooke allows that his list of core needs can be cautiously expanded - technology, resources and generosity permitting - to accommodate other plausible candidates for course-of-life needs, such as the need for satisfying work or needs arising from the requirements of personality development. Whenever "fruitless" disputes occur about such expansions, Braybrooke recommends falling back on core needs. Against the obvious objection that satisfying some people's needs comes at the expense of other people's liberty, Braybrooke argues that an expansion of this list may itself be an exercise of liberty, provided the expansion happens as a result of social consensus (231ff). Braybrooke's clarification of all these issues is illuminating, even though there remains room for

rational disagreement about a number of questions. What exactly should be included in the list of course-of-life needs; what minimum standards of provision should be contemplated; and how exactly can we regulate the expansion of this list?[9]

Nevertheless, Braybrooke has made sense of the way needs can legitimately figure, and do in fact figure prominently, in disputes about justice. Course-of-life needs underlie many rights-claims and many expectations that are strongly backed by our sense of justice. It may well be an individual's strongest right and her most legitimate expectation that society's decision-makers will not enact policies which would sabotage her basic physical or social functioning, policies which would make it harder or even impossible for her to meet her course-of-life needs. Often the causal chains are complex, and deprivations may be unforeseen side effects of well-intentioned policies, but there are many blatant cases all over the world where governments cater to the preferences of the privileged while leaving even the most basic subsistence needs of the poor unmet. It is hard to see any reason why those whose course-of-life needs go unmet *through no commensurate fault of their own*, when there are potential resources to meet them, should feel any loyalty or respect for the laws and institutions of the country whose citizens they are.

§3 Equality

Equality is a notoriously vague and contentious concept. I want to begin by distinguishing two senses: the *inclusive* sense, in which "equality" simply denotes an ideal of just individual shares, so that *any* injustice can therefore be conceptualized as a deviation from equality in this sense; and a *non-inclusive* sense, in which "equality" is one just-making criterion among others, and not every injustice can be conceptualized as a deviation from equality in this sense.

---

[9] Although Braybrooke says, in reply to a critic, that "the book allows for uses of the concept of needs expanded as it were with unanimous consent", surely a strict condition like *unanimous* consent will effectively squelch any expansion. (Cf. 1988, 520)

Equality in the inclusive sense can be used either as a *purely formal* criterion (i.e. one that is compatible with any possible distribution whatever if it is suitably described) or as a more or less *substantive* criterion (i.e. one which rules out some distributions as unacceptable). The prime example of equality in the inclusive sense, used as a purely formal criterion, is Aristotle's principle of justice, according to which justice consists in treating equals equally and unequals unequally in proportion to their relevant differences (Cf. Bk V, *Nichomachean Ethics*). A good example of equality in the inclusive sense used as a substantive criterion is Honderich's conception of justice:

> A society should seek to secure, as far as is practicable, lives of equal satisfaction for all its members. It should do this by in general seeking to secure, as far as practicable, equality of income and wealth, equality of respect (where that is other than....the mere recognition of the relevance of all persons), equal political and legal freedoms, the full development of the different potentials of individuals by means of education and in work, equality in housing and environment, equal medical care and provision for old age. (183)

This is clearly an ideal of just individual shares, a master principle which defines what society owes each individual as a matter of justice. Like all such directly responsive, internally complex principles, it faces a multitude of problems. I shall consider these matters in the next chapter.

The sense of "equality" of interest to me here is the non-inclusive sense, equality as a just-making criterion which leaves room for other just-making criteria, although it may overlap them. Equality in this sense is best regarded as synonymous with non-discrimination on some specified basis, such as race, sex, physical disability, sexual orientation, age or other features which are considered irrelevant to the just distribution of some good at issue. "Equality" is used in the non-inclusive sense if a certain good is specified along with criteria on the basis of which discrimination in the distribution of this good is forbidden.

Types of equality which have figured prominently in social justice theorizing are: equality of freedom; equality of opportunity and equality of outcome; equality-in-meeting-needs; equality of life chances; equality of resources and equality of welfare (or well-being or satisfaction); and equality of consideration, respect, concern, attention or treatment. Some uses of "equality" clearly involve the inclusive sense and some the non-inclusive sense. Equality of life-chances, equality of resources or well-being, or equality of treatment seem synonymous with an ideal of just individual shares and therefore involve "inclusive" equality. Equality-claims to more specific goods, such as equality of access to medical care or education, commonly specify criteria, in the form of features of persons, on the basis of which discrimination in the distribution of such freedoms or opportunities is forbidden, and therefore commonly involve "non-inclusive" equality.

There are a variety of plausible general ways of characterizing what it means to treat people unequally or to discriminate against them. All are fraught with problems of interpretation, but they nevertheless provide some useful conceptual common ground. We could say that for something to constitute inequality or discrimination, two conditions must be met: there has to be (unjustified) harm inflicted on people; and some people must be harmed more than others in relevantly similar positions. Another roughly equivalent way of characterizing discrimination would be to say that discrimination exists whenever people are being placed at a serious disadvantage relative to some appropriate reference group, through no (commensurate) fault of their own, in the distribution of important social benefits. But perhaps the currently most widespread way of defining "discrimination" is to say that discrimination takes place whenever people are denied equal opportunity in acquiring X - where X stands for a variable domains of widely valued goods whose distribution can be socially regulated.

But what *is* equal opportunity in acquiring X? And how, apart from equality of outcome, can we tell if it has been achieved? I can only touch briefly on these

immensely complex issues here. For the sake of simplicity I want to restrict my discussion to one representative domain, desirable jobs. In this domain the shared core meaning of "equality of opportunity" seems to be the idea of distributing this good on the basis of *performance-related* traits rather than on the basis of race, sex, age, political or religious beliefs, sexual orientation, disabilities or such traits, unless they are demonstrably performance-related. "Performance-related" here must mean "furthering the social purpose of the job in question". Beyond this widely shared meaning of equal opportunity there are, however, three sharply different interpretations of this concept.

(1) Equal opportunity in the *pure performance sense* could be held to exist if individuals are selected purely on the basis of performance-related characteristics, no matter whether they acquired them under highly privileged or strongly adverse conditions. (2) Equal opportunity in a *performance/motivation sense* could be held to exist if people are selected purely on the basis of performance-related characteristics *and* if all people with sufficient talent and motivation are assured of at least minimally favorable conditions for acquiring performance-related characteristics. (3) Equal opportunity in a *performance/motivation/endowment sense* could be held to exist if people are selected purely on the basis of performance-related characteristics, *and* if people with sufficient natural endowment and motivation are assured of at least minimally favorable conditions for acquiring performance-related characteristics, *and* if people with similar natural endowments are assured of (roughly) similar motivating conditions to make them want to acquire those characteristics.

Cauthen describes this third sense of equal opportunity in these terms:

> ...every individual with similar gifts who makes the same effort should have equivalent life chances for success. Hence efforts must be made to overcome the hindrances of family and social background that inevitably cripple some in comparison with others. (77)

Rational disagreement is probably possible about how much effort on the part of society is required to equalize the life chances of people whose motivation to develop their gifts may have been crippled by their family and social backgrounds. Some defenders of equal opportunity in the performance/motivation/endowment sense argue that justice (and other considerations) require us to relax the standard performance-related qualifying criteria for members of groups in whose case discrimination may have denied opportunities or undermined motivation to acquire equal qualifications. Others argue that no relaxation of standards is involved if we can plausibly reinterpret the social purpose of the job, or our social goals in general, in such a way that membership in a particular racial, sexual or other group becomes itself a performance-related characteristic.

Proponents of this view often propose to measure equality of opportunity by whether members of different racial, sexual and other groups are represented in each category of desirable jobs in proportion to their percentage of the total population. Given what we know about motivational and genetic factors, there is a poorly defined point at which such attempts at strict proportionality interfere with people's authentic preferences as well as the autonomy of the family (cf. Fishkin, 1983). There is scope for rational disagreement about what equality of opportunity requires, but laying out the consequentialist pros and cons of each contemplated policy clearly can go a long way toward narrowing the scope for such disagreement.

§4 Rights

What do we mean be calling something a "right"? Why should consequentialists recognize rights at all? And which rights should we recognize?

(1) *What are rights?* Rights are meant to protect certain types of choices or interests of individuals or groups. Sumner's lucid account makes clear just how complex a task it is to define a right in all its dimensions:

The content of a right is what it is a right to do or to have done; this is given by the content of its core liberty or claim. The scope of a right has two ingredients: the subjects of the right and its objects. The subjects of a right are those who hold it; this is given by the holders of the various Hohfeldian[10] advantages (liberties, claims, powers, immunities) which are ingredients of the right. The objects of a right are those against whom it is held; this is given by the bearers of the various Hohfeldian disadvantages (chiefly duties and disabilities) which are correlated with ingredients of the right. It follows that the content and scope of a right have been completely specified only when the content and scope of its several ingredients, both core and peripheral, have been completely specified. Finally, the strength of a right is its ability to override, or susceptibility to being overridden by, competing moral considerations. The strength of a right has been completely specified when its weight has been given relative to every sort of consideration with which it might compete. (1987a: 124)

(2) *Why should rights be recognized at all?* From the point of view of promoting well-being, it is clearly best if certain types of actions are always protected than if we engage in complicated case-by-case calculations about whether interfering with such an action would be more likely to promote well-being in this particular case. As Campbell explains this consequentialist rationale,

As long as individuals have desires (however altruistic) on which they wish to act and as long as such acts are facilitated in a socially beneficial way by allowing the individual a range of legal powers with which to pursue his interests, then there is reason to have a rights-based system of rules. (188-9)

For a consequentialist, something counts as a right if giving this extraordinary protection to the particular type of action so protected constitutes a good consequentialist gamble. Rights ultimately protect certain actions (which often, but not necessarily, constitute vital interests of the individual agent) from routine interference, even interference for the sake of whatever honestly seems, at the moment, to government officials or other decision-makers to best promote well-being. Of course, such protection cannot be absolute; in extreme circumstances and

---

[10] Sumner clarifies and elaborates Wesley Hohfeld's pioneering (1919) analysis of the fundamental conceptual building blocks of *legal* rights for purposes of analyzing *moral* rights. (Cf. Sumner, 1987a: 18f)

after a special process of deliberation, interference with someone's rights may be justified after all to avert a major disaster.

(3) *What types of actions are justifiably protected as rights?* Some candidates for the status of rights are the right to life, to security of the person, to bodily autonomy, to certain forms of property, to the pursuit of happiness, to a rather open-ended list of civil and political liberties like freedom of speech and assembly, to a reasonably substantive equality of treatment in the allocation of sundry social benefits like jobs or educational opportunities, and a roster of welfare rights that are closely tied to elementary human needs. The welfare rights a society should recognize are of course relative to what the society can afford. There is room for rational disagreement, I think, both about what kinds of actions deserve protection as rights and how strong the protection should actually be. These questions are far too complex to be explored in the necessary detail here.

Combining such general rights with specific circumstances, one can derive a plurality of more specific rights, though rarely in an uncontroversially deductive fashion. Derivations of this kind are fraught with potential for rational disagreement.

The following reconstruction of the central argument in Shue's book on *Basic Rights* provides an example of an attempt to derive welfare rights from liberty rights and, in effect, to challenge the moral status of this distinction:

1. If a society is not committed to protecting a small number of basic rights which are necessary for the effective exercise of any other right, a society does not really protect any rights at all.
2. All societies should protect some rights like everyone's right to life, freedom or the pursuit of happiness.
3. Basic rights like the right to physical security (a right not to be murdered, tortured or assaulted) and the right to subsistence (a right to minimally adequate food, clothing, shelter, air, water and health care) are necessary for the effective exercise of any other right.

Therefore, any society should be committed to protecting people's right to physical security and subsistence.

Shue's argument links up with the disagreement about how best to interpret the concept of equality of opportunity. Shue's argument makes a strong case, I think, that the *pure* performance sense of equality of opportunity must be rejected. It does not show, however, that equality in the performance/motivation/endowment sense can be demanded as a moral right, as is sometimes claimed by those who call for proportional representation of various groups in all desirable social positions.

Desert-claims, needs and more or less urgently felt desires, demands for various forms of equality and expectations created in various ways can all too easily be expressed as, or translated into, rights-claims. Many disagreements about rights would be remediable if all the disputants shared a consequentialist moral framework and a rational conception of how rights must ultimately be justified within such a framework. But sometimes there are such intractably complex empirical issues involved that there is every reason to think that rational disagreement about the merits of conflicting rights-claims will persist indefinitely.

§5 Good-faith expectations

By "good-faith expectations" I mean existing expectations of individuals, held in good faith, about what is their due and which, if ignored, arouse a sense of injustice. I speak of "good-faith" rather than "legitimate" expectations because I want to avoid the implication that claims based on such expectations are necessarily justified demands of justice rather than merely one type of justice-based claims. This category obviously overlaps all the other just-making criteria. Many, but surely not all, expectations which people hold in good faith about what is their due are need-, desert-, equality-, or rights-based. Good-faith expectations may also be based simply on established practice, law, reasonably hoped-for reciprocity or some other considerations. As I shall use this term, "good-faith expectations" is meant to refer to a residual category of considerations that are relevant to justice. Such a residual category is needed because sometimes it is not clear which, if any, of the other just-

making criteria has been disregarded, and yet a person may feel a sense of injustice and may indeed have a plausible case for complaint on grounds of social justice.[11] An example might be a sudden raising of university admission standards which dashes a person's long-nurtured hopes for admission, or a drastic reduction in social insurance allocations. Or take the person who has spent years training to be a research scientist, encouraged in his belief that as a scientist he would be able to contribute greatly to society. Now he is ready to set up his research project. He has had good reason to believe that public funding would be forthcoming (the research has no sufficiently immediate commercial promise to interest private investors), but the government, in an effort to reduce the national debt, suddenly excludes the relevant category of research funding.

Many good-faith expectations, as well as many rights-, desert- and equality-based claims, appeal to reciprocity, to a sense that we should justly get something of value in return for giving something of roughly equal value. Other important types of expectations are based, not on reciprocity, but on established practice. It is expectations in this sense which, though underpinned by strong emotions, are often difficult to integrate with other just-making criteria.

Rescher aptly called the dilemma that expectations pose from the point of view of justice the "Reformer's Paradox":

> ...the actual, existing distribution...is a very important one, from the standpoint of distributive justice...And this existing situation carries within it an existing body of claims, claims which must, in the interests of justice, be taken into account. Justice limits utility at exactly the point of the "Reformer's Paradox": Given an imperfect existing initial distribution, any redistribution in the interests of arriving, from the standpoint of justice, at a superior distribution runs headlong into the pattern of existing claims that cannot - in

---

[11] But are all members of this residual class of just-making considerations necessarily "expectations"? If someone who lacks any expectation of getting x could nevertheless have a just claim to it, this residual class would obviously be misnamed. But is this a real possibility? Here, I think, we must again distinguish consequentialist from more metaphysical conceptions of justice. On a consequentialist account, if someone neither deserves x, nor needs x, nor has a right to x, nor expects to get x, nor is treated discriminatorily by not getting x, then I see no consequentialist rationale for thinking that it could possibly be a concern of justice that this person does not get x.

the interests of the very justice that provides the rationale for the entire enterprise - be brushed aside as an irrelevant obstacle. (121)

It is the stranglehold of long-standing expectation, among other things, which makes reforms of society's "basic structure" in the name of justice extremely difficult. Such reform tends to fail, in large part because an abrupt break with good-faith expectations mobilizes a tremendous sense of injustice in many people.

When people have fashioned life plans based on the way things always have been, disappointing their expectations tends to arouse their sense of injustice. An example of a social practice which is in conflict with most plausible conceptions of equality, desert, need and moral rights and which, from these perspectives, seems strikingly unjust, but is arguably legitimated by good-faith expectations (and a few other consequentialist considerations), is the practice of inheritance. But there is tremendous room for rational disagreement about what makes an expectation a legitimate just-making criterion, and for how much it should count, especially if it is not based on desert, need, equality or rights.

3.1.3 The criteria inclusion problem

§1 If we survey the range of things which human beings value, we find a number of values which could plausibly compete with commonly invoked justice-based considerations in constraining social distributions. In other words, we need to distinguish the set of considerations that are held to make some distribution *right*, from the proper subset of considerations that are held to make it *just*. What is of value to human beings, and which values can plausibly take on the status of a political rather than a mere personal value is, of course, itself contested.

But the range of plausible *political* values competing with justice may include - depending on what is already included in someone's conception of justice - freedom, happiness, peace and order, national security and prestige, economic productivity, artistic achievement, scientific progress, a sense of community and cultural integrity, cultural diversity, the preservation of valued traditions, the

conservation of the environment, or the needs of international development. It certainly seems possible that competing values could, on occasion, justifiably override considerations of justice. But some philosophers have regarded the value of justice as "indefeasible".[12] How could that be?

To answer this question, we must clarify the basic options available for handling the criteria inclusion problem. One can simply take stock of the considerations commonly invoked as relevant to distinguishing just from unjust distributions, draw up a rough list of standard "just-making criteria" and then busy oneself with clarifying the complexities of each item on this list. But one can also exclude commonly invoked considerations from one's list of just-making criteria or include some considerations that are not commonly invoked as relevant to distinguishing just from unjust distributions. Which option is chosen is not merely a semantic matter, but has potentially significant consequences, because by *including* within the scope of justice-based considerations many considerations commonly regarded as irrelevant to distinguishing just from unjust distributions (or by *excluding* from their scope many considerations commonly regarded as relevant), one in effect proposes to realign the unusually strong sentiments which normally underpin justice-based claims with a new set of criteria.

Such a proposal may or may not work. By diluting the commonly invoked list of just-making criteria with considerations normally thought to be completely irrelevant to giving individuals their due *qua* individuals, the concept of justice may simply lose some of its distinctive normative force. But because the cognitive and emotive aspects of people's sense of justice are to some degree separable, there is clearly *some* room for such realignment by rational or non-rational means of persuasion. Non-rational persuasion by means of propaganda, "nationalist" appeals to common ethnic identities and various brainwashing techniques can probably

---

[12] Cf. J.P. Day, "The Indefeasibility of Justice", in *Liberty and Justice* (1987). Day mentions Socrates, Kant and Mill as holding a Principle of the Indefeasibility of Justice.

achieve much greater realignments, however, than someone can hope to achieve by means of a theory of justice which tries to offer rational justifications. A theory of justice cannot stray arbitrarily far from the prevailing pattern of people's expectations and sentiments about justice if it is to be acceptable, or even *recognizable*, to them as a theory of justice, and therefore to have sufficient potential to harness their sense of justice. Still, within hard-to-define limits, reasoned appeals *can* substantially change *some* people's minds and sentiments about justice. There is both hope and danger in the fact that some theories of justice have had large influence on political leaders who then set the machinery of public indoctrination in motion.

§2 Depending on what one includes in one's list of just-making criteria, and on how much weight one attaches to these criteria relative to other political values, one can construe the relation between justice and other constraints on social distributions in several ways.

Before I examine this relation, I must clear up a potential confusion about what exactly is meant by "constraints on social distributions". Are there political values which are *not* constraints on social distributions? John Rawls seems to think so. Since his master principle of justice defines the distributive outcome which society's basic structure as a whole must achieve to be just, and since he argues that the demands of justice are "uncompromising"(4), all constraints on social distributions must already be incorporated and somehow harmonized within his master principle. But he also says,

> A complete conception defining principles for all the virtues of the basic structure, *together with their respective weights when they conflict, is more than a conception of justice*; it is a social ideal. The principles of justice are but a part, although perhaps the most important part, of such a conception. (9; emphasis added)

Unless Rawls were considered to commit an inconsistency here, the only way I see to interpret his remarks is that he thinks that not all "virtues of the basic structure"

function as *intended* constraints on social distributions. But what would be an example of such a virtue? I cannot think of any political value that might not, under some conceivable circumstances, become an intended constraint on a social distribution. It is possible, however, that there are political values which do not *ordinarily* function as such constraints, but are simply part of the cultural givens which no one questions. Examples might be cultural diversity or territorial sovereignty. To say that such values might not consciously or directly serve to justify distributions in a particular society cannot mean that they could not emerge as extremely important constraints on distributions under some conceivable circumstances. The question, "Are there political values which are not constraints on social distributions?", is therefore best answered as follows. There are political values which do not *ordinarily* function as *intended* constraints on distributions; but all political values constrain distributions in indirect ways and all could become intended constraints under conceivable circumstances.

§3 What, then, is the relation between the demands of justice and other ordinarily intended constraints on social distributions? Thinking about this relation in a systematic way, in terms of mutually exclusive and jointly exhaustive categories, will require us to distinguish four basic types of views about the list of just-making criteria and two types of views about the combined normative weight of these criteria relative to other constraints on social distributions. I shall try to assign reasonably descriptive labels to each type and mention their advantages and drawbacks, and important representatives of each.

I want to call the four views about the list of just-making criteria "the standard view" of justice, "the constricted view", "the expanded view" and the "mixed view", depending on whether a view accepts our standard list of commonly invoked just-making criteria as constitutive of justice-based claims or modifies this list in one of three ways: by excluding certain criteria, including additional criteria,

or both excluding and including criteria. Each basic view has two subtypes, which I want to call "the indefeasible view" of justice and "the defeasible view", depending on whether a view denies or allows that the demands of justice must be balanced against, and can therefore be defeated by, other political values which constrain social distributions. The following table summarizes the distinctions I wish to draw between different views about justice:

	Accepts commonly invoked just-making criteria as constitutive of justice	*Excludes certain commonly invoked just-making criteria*	*Includes certain additional just-making criteria not commonly invoked*	*Both excludes certain common criteria and includes new criteria*
Demands of justice need not be balanced against competing values	**The standard indefeasible view of justice** (Kant)	**The constricted indefeasible view of justice** (libertarians)	**The expanded indefeasible view of justice** a) the partly expanded view b) the catch-all view (Rescher?)	**The mixed indefeasible view of justice** (Rawls)
Demands of justice must be balanced against competing values	**The standard defeasible view of justice** (commonsense pluralism; eg. Berlin)	**The constricted defeasible view of justice** (Marx?)	**The expanded defeasible view of justice**	**The mixed defeasible view of justice**

Table I: A taxonomy of views of justice, according to
(i) whether the "standard list" of just-making criteria is accepted as constitutive of the concept of justice or modified; and
(ii) whether other political values are accorded sufficient weight to require balancing with the demands of justice

I. The standard view of justice

1. *The standard indefeasible view:* One can accept some list of commonly invoked just-making criteria as constitutive of justice-based claims and deny other values sufficient weight *ever* to override the demands of just-making criteria. This position is well captured by the phrase "Let justice be done even though the earth may perish", and it can be clearly ruled out on consequentialist grounds. Although such a view of justice is a major problem in politics, I know of no *contemporary*

justice theorist who defends it. Kant, of course, seems to have thought that not only the principles of justice, but all moral rules flowing from his Categorical Imperative, are absolutely inviolable. About retributive justice, for example, he says,

> The law concerning punishment is a categorical imperative, and woe to him who rummages around in the winding paths of a theory of happiness looking for some advantage to be gained by releasing the criminal from punishment or by reducing the amount of it - in keeping with the Parisai motto: "It is better that one man should die than that the whole people should perish." If legal justice perishes, then it is no longer worth while for men to remain alive on earth. (*The Metaphysical Elements of Justice*, 100)
> Even if a civil society were to dissolve itself by common agreement of all its members (for example, if the people inhabiting an island decided to separate and disperse themselves around the world), the last murderer remaining in prison must first be executed, so that everyone will duly receive what his actions are worth and so that the bloodguilt thereof will not be fixed on the people because they failed to insist on carrying out the punishment; for it they fail to do so, they may be regarded as accomplices in this public violation of legal justice." (*The Metaphysical Elements of Justice*, 102)

Despite Kant's many striking pronouncements about the uncompromising nature of justice and other moral imperatives, his derivations of the demands of justice from the two general formulations of the Categorical Imperative seem often obscure and controversial, and he does not seem to have come to terms with the crucial problem of how to handle conflicts between allegedly uncompromising moral imperatives.

Minor and even major injustices seem inevitable in the life of any society, and it is hard to see any rationale for accepting with equanimity whatever consequences may result from prohibiting trade-offs between justice and other values, such as national security or great artistic and scientific achievements. Consequentialists will tend to accept many minor injustices and focus on preventing major ones; but how to draw a principled distinction here between "major" and "minor" is obviously a knotty problem.

2. *The standard defeasible view:*This is the commonsense view of political pluralists who see no reason why ultimately all virtues of the good society must be capable of principled integration which avoids the need for intuitive, context-

dependent balancing of competing virtues. Viewing justice as an important but defeasible constraint on social distributions, and including essentially those general types of considerations in the list of just-making criteria which individuals commonly invoke in pursuit of their personal well-being and back up with their sense of justice would seem to have the advantage of defining the concept in a way that is relatively consistent with ordinary usage and therefore avoids both the fanaticism of the exclusive view and the many pitfalls of the catch-all view (see below). Such a definition would clearly not preclude piecemeal attempts to get people to change their perspective about the precise implications of particular just-making criteria in the light of a wide variety of consequentialist considerations. But it would make the demands of justice defeasible in the final weighing of distributive constraints imposed by all political values.

Admittedly, ordinary usage is ambiguous on the question as to whether it can ever be right to commit an injustice, perhaps because admitting that values of major importance can sometimes be mutually exclusive is extremely painful. But either we admit that the properly balanced demands of commonly invoked just-making criteria must be further balanced against the demands of other important constraints on social distributions; in which case committing injustices must sometimes become justifiable. Or we deny that committing injustices could ever be justifiable; in which case we can never allow trade-offs between the properly balanced demands of commonly invoked just-making criteria and other important constraints on social distributions. We cannot have it both ways, and often people are simply confused on this point.

We could, of course, *stipulate* that whenever the properly balanced demands of commonly invoked just-making criteria must reasonably be overridden by the demands of other ordinarily relevant constraints on social distributions, we will make a semantic adjustment in the ordinary meaning of justice and say that an injustice has not been committed after all. But from a consequentialist point of

view, such a semantic adjustment would seem to be a bad idea. First, such a meaning-change is likely to create conceptual confusion because it divorces the concept of justice from what individuals claim in pursuit of their personal well-being and back up with their sense of justice. Second, the legitimacy of such a meaning-change in any given case would obviously be contingent on someone's judgment of what constitutes the best trade-off between the properly balanced claims of just-making criteria and other political values. Such judgments are likely to be extremely controversial. Allowing a semantic adjustment in the ordinary meaning of justice, depending on the judgment of political decision-makers as to whether a particular trade-off was necessary, would seem to invite attempts to conceal the painfulness of such trade-offs and to dismiss justice-based grievances by glibly denying that any injustice has been committed. Allowing the suggested meaning-change would deprive us of the vocabulary to carry on this debate.

II. The constricted view of justice:

One can exclude from one's list of just-making criteria some or even all considerations which are commonly regarded as relevant to distinguishing just from unjust distributions. "From each according to ability; to each according to need" has been a hugely influential principle of this kind. Many theories of justice exist which single out need, or desert, or rights, or equality, as the only just-making criterion. Many libertarians constrict just-making criteria to a single basic right, namely the right to use one's own mind and body to make oneself better off as long as one does not thereby make others worse off. The constricted view of justice is motivated by a desire to make the seemingly chaotic diversity of justice-based claims manageable, but the price for severely constricting the list of commonly invoked just-making criteria is high. It will make one's view of justice seem lopsided and almost certainly deprive it of any sufficiently close relation to individuals' claims to social resources which are backed by their sense of justice. It will also have several

other drawbacks of the expanded view of justice, to be discussed below. (Ch 3.5 examines whether there could nevertheless be a sufficient rationale to adopt the libertarian view of justice.)

One can, of course, hold either a constricted indefeasible view or a constricted defeasible view, depending on whether one either denies or allows other political values sufficient weight to override the demands of just-making criteria on some occasions. The first alternative will appear to be a peculiarly lopsided form of justice fanaticism, while he second would seem to give up most of the gains achieved by simplifying the list of just-making criteria when it comes to balancing the demands of justice with competing constraints on social distributions.

III. The expanded view of justice

1. *The expanded indefeasible view:* This view includes in the list of just-making criteria some or all constraints on social distributions, many of which are not commonly regarded as relevant to distinguishing just from unjust distributions, and it regards justice as a value which can never override other ordinarily relevant constraints on social distributions. Here we must distinguish two varieties, which I want to call "the partly expanded view" and "the catch-all view".

(a) The partly expanded indefeasible view: It is possible to include in one's list of just-making criteria *some* considerations not commonly invoked as relevant to distinguishing just from unjust distributions without, however, including *all* considerations ordinarily regarded as relevant constraints on social distributions. The indefeasible variant of this view would seem to share the drawbacks of the standard indefeasible view as well as the drawbacks of the catch-all view to a lesser extent.

(b) The catch-all view of justice: This view simply expands the meaning of the term "justice" to incorporate *all* values that are ordinarily regarded as relevant to constraining social distributions. Rescher could be interpreted as making this move

when he includes "social utility" and "supply and demand" among his eight "canons of distributive justice" (cf. Rescher, 73-81). It can be argued that "social utility" or "supply and demand" involve many paradigmatic non-justice considerations, and regarding these values as constitutive of justice-based claims collapses a useful distinction that marks an important difference. Although it is ultimately a matter of semantics whether one uses "justice" as an umbrella term referring to an uncoordinated plurality of considerations many of which are not based on the claims individuals make to social resources and which they back up with their sense of justice, such a promiscuous use of this term would seem to me to have hardly any advantage and several drawbacks. It will lack any sufficiently close relation to individuals' claims to social resources which are backed by the sense of justice and will, therefore, necessitate another term for justice in this narrow sense. It will empty appeals to "justice" of their normative force. And it will almost certainly give rise to conceptual confusion and equivocation. As Campbell points out,

> If 'justice' is *defined* as the overall standard of social rightness, then logically no other value can stand prior to justice since all relevant values are subsumed under its umbrella.... It is possible arbitrarily to define justice as the prime social value and then to go on and fill in its content with whatsoever is thought to be morally most important in social distribution, and perhaps also in the aggregation of benefits and burdens. But this dogmatic approach has the effect of undermining our efforts at conceptual clarification by removing the constraints imposed by the informal logic of the language of justice in actual political debate, thereby rendering dangerously misleading any subsequent appeals to our 'intuitions' about what we think is 'just' or 'unjust', for such intuitions are rooted in our operative rather than in our stipulative normative conceptions. (8)

We can give the term "justice" any semantic content we like. But what we clearly cannot do is to save ourselves painful trade-offs between political values by simply conceptualizing all of them as considerations of "justice". If someone like Rescher succeeded in broadening the ordinary usage of the term "justice" in the way he suggests, we will still be left with the problem of how to conceptualize the type of

claims which individuals press as their due with the sort of passionate conviction characteristically associated with claims of desert, need, equality, rights and good-faith expectations.

2. *The expanded defeasible view:* This view expands the list of just-making criteria without, however, going so far as to include all relevant constraints on social distributions, and it does not regard "justice" as an indefeasible value. It shares to a lesser extent the drawbacks of the catch-all view. The expanded view of justice, whether in its indefeasible or defeasible form, seems to me to lack any convincing rationale to outweigh these serious disadvantages. The possible exception is an expanded view which explicitly includes productive incentives as a just-making criterion. I shall discuss the consequentialist pros and cons of *this* version of the expanded view of justice below.

IV. The mixed view of justice

This view both excludes from the list of just-making criteria considerations which are commonly regarded as relevant to distinguishing just from unjust distributions, and includes considerations which are not commonly regarded as just-making criteria.

1. *The mixed indefeasible view:* This mixed view regards the value of justice as indefeasible, that is, the demands of justice need not be balanced against competing constraints on social distributions. This, of course, is the view held by John Rawls (and also the view conveyed by various "ideals of the good society" proposed as theories of *justice* in the wake of Rawls' theory). The mixed indefeasible view, as well as the indefeasible versions of the standard, constricted and expanded views of justice, all share the characteristic that they could, without inconsistency, be proclaimed in exactly the same ringing words with which Rawls begins his book:

> Justice is the first virtue of social institutions, as truth is of systems of thought. A theory however elegant and economical must be rejected or revised it if is untrue; likewise laws and institutions no matter how efficient

and well-arranged must be reformed or abolished if they are unjust. Each person possesses an inviolability founded on justice that even the welfare of society as a whole cannot override. For this reason justice denies that the loss of freedom for some is made right by a greater good shared by others. It does not allow that the sacrifices imposed on a few are outweighed by the larger sum of advantages enjoyed by many. Therefore in a just society the liberties of equal citizenship are taken as settled; the rights secured by justice are not subject to political bargaining or to the calculus of social interests. The only thing that permits us to acquiesce in an erroneous theory is the lack of a better one; analogously, an injustice is tolerable only when it is necessary to avoid an even greater injustice. Being first virtues of human activities, truth and justice are uncompromising. (3)

Why does this paragraph seem so startling? The reason is that most readers will naturally assume the standard meaning of justice in terms of commonly invoked just-making criteria and jump to the conclusion that the writer proposes to defend what I have called the standard indefeasible view of justice. Obviously, if someone introduced beliefs about justice prior to telling his audience that he has switched to an unorthodox view of the meaning of this term, he might be accused of creating misleading expectations. And Rawls has indeed switched to an unorthodox view which assumes not only a radical departure from any ordinary list of just-making criteria, but also an even more radical departure from what is ordinarily regarded to be the immediate subject of justice-based considerations. For him, the "primary" (immediate or direct) subject of justice is not the distributive share of an individual or group, but certain fundamental aspects of the entire set of distributive outcomes in society. Rawls can say that, for him, the demands of social justice are inviolable *only because* he identifies these demands with all those considerations which he considers relevant to constraining social distributions in ordinary circumstances. So, of course, do proponents of the catch-all view. But the difference between any two views of justice could hardly be greater than that between Rawls' mixed indefeasible view and Rescher's catch-all view.

Suppose it were possible to find a set of constraints on social distribution which, under ordinary circumstances,[13] coordinates the demands of *all* political values in a way that can be justified to individuals as giving each what he can reasonably demand in the name of justice and therefore promises to harness most people's sense of justice in support of it. It would be natural and clearly defensible to identify this set of constraints as "principles of justice". This, of course, is Rawls' use of the term when he thinks of his theory not just as an alternative to the utilitarian account of *justice* (as outlined, for example, in Mill's *Utilitarianism*, Ch. V or Sidgwick's *Methods of Ethics*, Bk III), but, at least in the realm of political decision-making, as an alternative to utilitarianism as a whole.

Although both the catch-all view and the mixed indefeasible view include *all* ordinarily relevant constraints on social distribution in their respective lists of just-making criteria, the difference between these two conceptions of justice is now clear. Whereas the catch-all view considers just-making criteria to form an uncoordinated, perhaps incommensurable plurality of criteria whose demands must be balanced according to the intuitions of political decision-makers in specific contexts, the mixed indefeasible view welds some commonly invoked just-making criteria together with other constraints into a sort of social algorithm or at least draws the outlines of a social blueprint. By a "social blueprint" I mean distributive constraints on those social arrangements reasonably regarded as being of fundamental importance, constraints whose acceptance is required for any rational person, and which avoid any need for unguided intuition at this fundamental level.

If we use "justice" in this sense, then, on the face of it, we might seem to need yet another term to describe the criteria which individuals invoke when they claim something as their due in pursuit of their personal well-being and back it up with their sense of justice. But of course, Rawls' radical idea was that principles sufficient

---

[13] Of course, national emergencies, such as the threat of military invasion, will justify a suspension of these constraints.

for constraining the distributive outcome of the entire system of basic social institutions could be found and justified as plausibly giving each individual his or her due *qua* individual (given that there is tremendous dispute about what this amounts to in any case). If Rawls' project had succeeded, all would be well indeed. We would now have a social blueprint which would finally have laid major disputes about justice to rest. But that has not happened. Instead, Rawls' proposal for a social blueprint at the level of society's basic structure has triggered something of an avalanche of alternative visions of the good society, now standardly offered as theories of "justice". The net result is perhaps more confusion than ever about the demands, and indeed the core *meaning*, of justice.

It will be convenient to have additional terminology on hand to clearly distinguish the type of mixed defeasible view of justice which Rawls proposed from other views of justice that may be superficially similar. Theories of justice differ in a fundamental way, according to whether they attempt to be *directly* or *indirectly* *responsive* to individuals' justice-based claims. Directly responsive theories, which take justice to be concerned with individual distributive shares, defend a standard of ideally just individual holdings of socially distributable goods, and they interpret justice as requiring redistributions of such goods from those who, according to this standard, have too much to those who have too little. Indirectly responsive theories like Rawls', which take justice to be concerned with society's basic structure, defend a standard of ideally just distributive outcome patterns or processes for certain important social goods, and they interpret justice as requiring a system of social institutions which will enforce these distributive outcomes or processes.

2. *The mixed defeasible view:* This view both includes and excludes criteria from the standard list, but requires that the demands of justice be ultimately balanced against other relevant constraints on social distributions. Since this view would combine the serious drawbacks of both the constricted and expanded views, and it would seem in the end to require intuitive balancing against other constraints

on social distributions, it would need a very strong rationale. I cannot think of any examples of this theoretically possible view.

§4 Contemporary justice theorizing is a conceptual swamp. Among many conceptual confusions, perhaps the most fundamental confusion has to do with the meaning of the term "justice" itself. Which of two very different meanings of the term "justice", if any, should justice theorists adopt? Should we allow the legitimacy of equating "principles of justice" with the goal of finding a social blueprint for society's basic structure organized as a whole? Or should we equate "principles of justice" with the vastly more modest aim of constructing principles which strike a consequentialist balance among people's justice-based claims in fairly context-dependent ways? This conceptual decision will, of course, depend on the most reasonable answer to the question of whether a social blueprint for society's basic structure is really possible. Before answering, let us get very clear about what such a blueprint would involve.

(1) It would have to single out certain features from the total set of a society's distributive outcomes as uniquely relevant to assessing the moral acceptability of these outcomes.

(2) These features would have to be sufficiently straightforward that it becomes actually *verifiable* whether the total set of distributive outcomes prevailing in a given society at any given time has or doesn't have these features.

(3) The blueprint would have to suggest, at least in rough outlines, some feasible procedure for deciding which existing social institutions should be reformed in what ways so as to give the total set of distributive outcomes those features which would make it morally acceptable.

(4) One particular proposal for how to single out these features and how to reform society in ways so that the total set of distributive outcomes would actually have these features must be justifiable to all rational people as the best practicable

response to the claims individuals make in pursuit of their well-being *qua* individuals, in order to have credible potential to enlist people's sense of justice in support of this proposal.

If there is realistic hope of justifying principles which meet these conditions - which *both* provide determinate constraints for the *overall* distribution of the benefits and burdens of social co-operation *and* can somehow be justified as best giving individuals what they can reasonably demand in the name of justice - then the rationale for adopting the mixed indefeasible view of justice seems compelling. But if there are fairly compelling reasons for thinking that principles at such a high level of generality are destined to be normatively empty abstractions, or to involve unmanageable complexities, or to fit poorly with people's justice-based demands, or to be hopelessly controversial for other reasons, then the rationale for sticking with something like the standard defeasible view of justice seems equally clear. (I shall examine this issue in the next chapter.)

§5 But even if the search for a uniquely rational social blueprint turned out to be hopeless, another strongly contested part of the criteria inclusion problem still poses a major obstacle to accepting any particular basic view of justice: the question of productive incentives. *Should concessions to people's self-interest, for the sake of providing incentives to economic productivity, be considered to fall inside or outside the list of relevant considerations that help make a distribution just?* There seem to be strong arguments in support of both alternatives. It is not even entirely clear whether we should say that a view that includes our standard list plus productive incentives as an additional criterion is best regarded as an example of the expanded view of justice, or that the standard view of justice is somewhat indeterminate on the question of productive incentives.

The basic case for *including* productive incentives among just-making criteria is very simple. Suppose we lived in a society of economic equals. And make the

realistic assumption that everyone is fairly poor in this society and that almost everyone could become considerably better off economically if the talented and ambitious were given special economic incentives to exercise their talents and ambition in productive ways. If doing justice consists in *both* satisfying individuals' justice-based claims in an impartial way *and* to the greatest possible extent, and if the exercise of talent and ambition will generate the resources to satisfy these claims to a greater extent, then giving talented and ambitious people a larger-than-average share of society's wealth *as an incentive to exercise their talents and ambition* cannot reasonably be considered unjust. Suppose we adopt a Rawlsian scheme and tolerate only those economic inequalities which make the least advantaged five percent of society better off. Nevertheless, a certain number of individuals who, *as individuals*, might have done better under the old egalitarian order or under some alternative social order suffer hardship under this one, hardship which cannot be plausibly rationalized as ultimately self-inflicted and *deserved*. Are these individuals justified in claiming to be victims of social injustice? The sensible answer, for a sophisticated moral theorist, would seem to be no. Clearly we don't want to call a social order unjust which satisfies the justice-based claims of individuals to the greatest possible degree, even though some individuals who might have done better under a different order will suffer special hardship under this one. After all, it is pointless to call a social order unjust if it is impossible for any alternative order to be more just.

Yet, upon reflection, the case for *excluding* productive incentives from among just-making criteria seems extraordinarily resilient. When we accept a particular form of capitalism as, on balance, preferable to its alternatives, are we really committed to accepting all the possibly resulting inequalities as perfectly *just*? Wouldn't that invite rationalizations of paradigmatic injustices? Is it really just that the rich should get tax exemptions in the perhaps justified hope that they might then have an incentive to invest locally and create jobs, while at the same time the

less well off must bear a heavy tax burden? Should we not say that such inequality is perhaps morally justifiable although unjust, rather than that it is justifiable and therefore not unjust? It is not at all obvious why it should satisfy the least advantaged in the slums of capitalist countries to be told that their destitution in the presence of extravagant and often unearned wealth is just, simply because some sophisticated moral theorist may be sincerely convinced that ours is the best economic system overall and that the percentage of the destitute under any alternative system would be higher. We would not regard it as respecting the separateness of persons to tell an individual who claims to be victimized by social injustice: "Yes, you suffer undeserved hardship, and under some other system you as an individual would probably be better off, but our system produces the most wealth *on average* - so you aren't really done an injustice." Should we regard it as fully respecting the separateness of persons if we now tell this same individual that the least advantaged five percent of the population would probably be even worse off under any alternative economic system, and that she should therefore desist from claiming to be a victim of injustice? From the point of view of this individual, haven't we only trivially changed our tune and left her just as badly off as before?

Suppose some hypothetical economic system makes the major ingredients to a good life dependent on paid employment and assigns some percentage of the population to a state of involuntary unemployment, low self-esteem, and suicidal unhappiness through no fault of their own, so as to create a reserve pool of labor which will exert downward pressure on inflationary wage demands. Now it seems hardly good enough simply to tell such individuals that they should consider themselves fortunate to live in a just society and assure them that they have no reason for complaint on grounds of justice. They will, in any case, continue to use the language of justice, or evolve some substitute for it, in an attempt to call attention to their fate and to motivate whatever assistance to them may be feasible. But suppose we have convinced the least advantaged in a Rawlsian type of society

that they cannot possibly have any legitimate grievance on grounds of justice. Why should other individuals higher up the socio-economic ladder agree that justice has been done to *them*, if social arrangements are systematically designed for the maximum benefit of the bottom five percent? Is it unjust to invest social resources in space exploration unless, as a result, the condition of slum dwellers promises to be ameliorated more effectively than through any alternative use of these resources? Could it be justifiable, on grounds of justice, that the relatively least advantaged, who may already be quite well off in absolute terms, should hold the other ninety-five percent back from realizing possibly important life goals in scientific or artistic endeavors of no foreseeable benefit to the least advantaged?

§6 These are complex questions, and obviously relevant to settling the question we are now addressing: whether we should build productive incentives into the heart of the concept of justice, or whether we should sharply distinguish between considerations of justice and productive incentives and openly consider balancing such considerations to be a matter of painful tradeoffs. It is clear that those who consider a uniquely rational social blueprint for society's basic structure possible *must* go against the grain of many people's firmest intuitions and include productive incentives among just-making criteria, even at the risk of possibly trivializing the undeserved destitution of many individuals as simply the inevitable byproduct of doing *justice*. Therefore, supporters of *in*directly responsive conceptions of justice have no choice.

But supporters of directly responsive conceptions of justice, who conceive of principles of justice as spelling out an ideal of individual shares of either society-wide or domain-restricted scope, do have a choice: they can hold fast to the standard defeasible view of justice and exclude productive incentives from the list of just-making criteria (except insofar as those incentives can be construed as rewarding desert). Or they can adopt an *expanded* defeasible view of justice and include

productive incentives into the list of just-making criteria, even where such incentives lead to windfall profits for people who clearly do not deserve them.

Suppose we find reason to adopt a directly responsive conception of justice. Which view of justice is *then* superior, the expanded view (expanded to include productive incentives) or the standard view (which excludes productive incentives)? Or is there perhaps room for rational disagreement here? I think this issue can be settled in favor of the standard view. Importing productive incentives as a just-making criterion into the heart of the concept of justice would complicate the proper use of this concept to such an extent as to radically reduce its usefulness for ordinary individuals who want to give voice to what they regard as their due. But that productive incentives are best excluded from the list of just-making criteria does not mean that they can also be excluded from the list of all morally relevant considerations. Committing an injustice for the sake of the morally desirable consequences of allowing productive incentives can sometimes be morally justifiable. (I will argue for a distinction between *bona fide* principles of justice and morally justified principles of justice in Ch 3.3).

But, it may still be objected, must we not include productive incentives in our list of just-making criteria on grounds of logic alone? Productive incentives create resources which allow us to satisfy justice-based claims to a greater extent than would otherwise have been possible. If we exclude them from our list of just-making criteria, we are left with the paradox that we may have made society as just as we can humanly make it, and yet we grant some individuals that they may have legitimate grievances on grounds of justice.

A consequentialist response to this argument is as follows. It is not at all a foregone conclusion that we should try to rid the language of justice of this paradox. What purpose is served by the language of justice? Justice, we said, is fundamentally a two-level or dual-purpose concept. Its basic purpose is *both* to give voice to what individuals regard as their due *qua* individuals *and* to articulate principled

constraints for rearranging aspects of the social order in greater conformity to what individuals regard as their due. Any such concept, *whether it includes productive incentives as a just-making criterion or not*, must contain the deep paradox that even a social order which satisfies justice-based demands to the greatest humanly possible extent will not be able to silence charges of injustice. A principled response to individuals' justice-based claims requires balancing the justice-based claims of different individuals, and balancing means aggregating, and aggregating implies that the claims of some individuals, through no special merit of theirs, will win out over equally legitimate claims of others. Therefore we will in any case have a paradox at the heart of the language of justice: we are forced to acknowledge legitimate individual grievances even in cases in which we think that social affairs are arranged as justly as we can possibly make them. But allowing a dual-purpose concept into our vocabulary that contains precisely this paradox has important consequentialist pros and cons. On the one hand, it may discourage complacency that everything already works for the best in the best possible social order and may help motivate the perennial search for better social arrangements. By hypothesis we may imagine to have reached some Rawlsian optimum, but we can probably never be sure what this optimum is or whether we have reached it. On the other hand, allowing the concept of justice to retain this paradox may also promote social strife and encourage expectations among some groups which simply cannot be met. Whether it would be better, on balance, to include productive incentives along with, or instead of, desert in our list of just-making criteria may be a matter for rational disagreement. My point here is that whether moral theorists should follow Rawls' lead and try to purify the concept of justice of its paradoxical implications is not a mere logical question which clearheaded conceptual analysis can resolve, but a question of what sort of social dynamics we want to encourage for purposes of motivating reforms toward greater well-being.

Our brief survey of commonly invoked just-making criteria as well as of different views about the criteria which are held to be constitutive of the meaning of justice makes evident the formidable obstacles faced by any theory of justice which does not simply appeal to some group's partisan intuitions about what justice requires, but actually tries to provide a compelling rationale for combining commonly invoked just-making criteria (and other considerations) in some normatively determinate way which any rational person must accept.

In view of the varied and imprecise nature of just-making criteria and all the different options of combining just-making criteria into a conception of justice, how can we *ever* convincingly argue for or against the justice of a distribution? We must examine what level of comprehensiveness we can reasonably demand from usefully determinate principles of justice, and what sorts of justification can be given for such principles.

3.2 DISAGREEMENT ABOUT THE COMPREHENSIVENESS OF PRINCIPLES

This chapter explores the level of comprehensiveness which is feasible for principles of justice.

3.2.1 Basic properties of principles of justice

§1 The last chapter dealt with *just-making criteria*, as I called those considerations which are considered relevant to making judgments about justice in general. In this chapter I want to examine the role of such criteria in *principles of justice* - by which I mean propositions which specify sufficient conditions for what justice requires in a particular domain. What important types of principles of social justice can we distinguish and what fundamental problems do they face? Although Rawls has done a great deal to give us the analytic vocabulary to deal with these questions, I shall have to coin a number of new terms as we go along. But first, let me clarify the important difference between a principle of justice and a mere justice-based claim.

We defined a "justice-based claim" as a claim to a share of social resources made by an individual in pursuit of personal well-being, based on a cluster of criteria which that individual would allow others to base analogous claims on, which typically includes such criteria as desert, needs, equality, rights or good-faith expectations. Principles of justice spell out what justice requires with respect to some domain by proposing a reasonable compromise between the conflicting justice-based claims of different individuals in that domain. The largest possible domain for a principle of *social* justice is of society-wide scope and includes all individuals and all goods whose distribution is thought to raise considerations of social justice. More restricted domains are demarcated by types of goods, institutions and practices, social roles, cultural subgroups, geographic locations and other individuating characteristics. Since one domain may be included in the scope of another, a principle which proposes *sufficient* conditions for determining justice within a

particular domain may only provide *relevant* conditions for determining justice in some larger domain. Moreover, since domains may overlap in complex ways, there is obviously ample room for conflict between domain-restricted principles.

This situation seems unsatisfatory. The ultimate goal of justice theorizing has, therefore, generally been regarded as finding principles of society-wide scope which jointly provide *sufficient* conditions for deciding whether anything whatever that might be claimed to be socially unjust is unjust. Rawls, for example, says:

> Thus justice as fairness seeks to identify the kernel of an overlapping consensus, that is, the shared intuitive ideas which...turn out to be *sufficient* to underwrite a just constitutional regime. (1985: 246-7; emphasis added)

Principles of justice, whether of society-wide or domain-restricted scope, state abstractly what conditions are sufficient for achieving a proper balance or reasonable compromise between competing justice-based claims to social resources (cf. Rawls, 10). But what makes a balance *proper* or a compromise *reasonable*? This chapter aims to illuminate the complexities associated with this question.

Principles are often divided into two sets of categories: "formal" versus "substantive"; and "procedural" versus "material". The distinction between formal and substantive principles refers to the degree of normative guidance a principle provides, and raises important questions (see §2 below). The distinction between procedural and material principles can be briefly explained as follows. Principles of justice are purely procedural if they define constraints on a distribution process and specify any resulting distributive outcome as just, as long as it results from this process. Principles are purely material if they define a just distributive outcome, without specifying by what process this outcome must or should be achieved. Many principles will, of course, combine procedural and material constraints.

But there is a confusion lurking here that needs clearing up. Rawls' claim is familiar that social justice must be conceptualized as a matter of just procedures or "pure procedural justice", because "there is no independent criterion for the right

result"(86). Rawls claims that trying to achieve justice in a society by defining just individual shares and then keeping track of whether each individual has more or less than their just share at any given time gets us into hopeless complexities. But given this claim, it seems initially confusing when, on the face of it, Rawls' own principles turn out to be obviously material. Is it perhaps tempting to think that he calls them "procedural" because they are (partly) derived from a contractarian choice procedure; but that would add yet another layer of confusion. The solution to the seeming paradox of Rawls' obviously material principles being called "pure procedural" is that they specify a distributive outcome at the level of society's basic structure and not at the level of individual distributive shares. Hence Rawls' principles are material when it comes to the design of the basic structure, but purely procedural when it comes to distributing particular goods to particular individuals. Rawls argues that we cannot find workable criteria for the share of society's resources which a given individual could justly claim and which a just society could therefore be said to *owe* this individual (cf. e.g. 87-8). But he thinks we may still be able to find workable criteria for designing a scheme of basic social institutions which does a reasonably good job overall of responding to concerns about justice, and which has the potential to enlist people's sense of justice in support of it. So Rawls' principles deserve to be called "pure procedural", not because they are derived from a hypothetical contractarian procedure, but because they are meant to respond to the justice-based claims of individuals purely by setting up society's basic structure in a certain way, and not by regulating individual distributive shares directly. Individuals can use Rawls' principles only to argue that the total system of social institutions should be reformed to more nearly achieve this society-wide pattern of distributive outcomes. But once this pattern has been achieved, Rawls' principles can no longer be used to justify complaints on grounds of justice, no matter how destitute a given individual's position may be.

§2 What sorts of principled responses are possible to the justice-based claims which individuals advance to a share of society's resources? An analysis of principles of justice must focus on two questions: how comprehensive the proposed principles are; and how convincingly the principles are justified. Ideally, we would like a theory of justice to offer us *maximally* comprehensive principles (see below) and to show their acceptance to be necessary for *any* rational person. We might also wish them to be *universally* valid, valid for all societies throughout history, at least where reasonably favorable circumstances obtain (Hume's famous "circumstances of justice").[1] But contemporary justice theorists have rarely been so ambitious. Rawls, for example, regards his theory as applying to the basic institutional arrangements of a "more or less self-sufficient" society and assumes "strict compliance" with his principles (8-9; 457). In recent writings he has further limited the scope of his theory as applying only to reasonably affluent pluralist democracies and left largely unclear, even in the case of these societies, what parts of their existing basic structure, if any, he finds objectionable on grounds of justice and what he would put in their place (cf. 1985, 223-51; 1988, 252).

What would it mean for a principle to be "maximally comprehensive"? By a principle's level of comprehensiveness (or what I shall also call the size of its normative payload) I mean how much substantive normative guidance it gives, which is a function of two quite distinct but frequently confused variables: generality and determinacy. Since the distinction between a principle's generality and determinacy is absolutely crucial, but rarely made explicit, let me make it explicit here. A principle which applies only to a narrow range of cases (low generality) and rules out few alternative courses of action (low determinacy) has a small normative payload. A principle which applies to a wide range of cases and rules out most alternatives obviously has a large normative payload. Of two principles which are

---

[1] Cf. Hume, *An Enquiry Concerning the Principles of Morals*, 15-25.

identical with respect to their determinacy, the principle which has greater generality has a greater normative payload. Of two principles which are identical with respect to their generality, the principle which has greater determinacy has a greater normative payload. The important exceptions are principles which have zero determinacy, e.g: "Always do what is right and avoid what is wrong." Such principles have no normative payload whatever, no matter what their generality.

We can define the level of comprehensiveness or the size of the normative payload of a principle in terms of the number of situations in which the principle could conceivably be applied in an action-guiding way. (Philosophers of science often define the empirical content of a hypothesis analogously, in terms of the number of situations in which the hypothesis could be tested.) It is the failure to distinguish between generality and determinacy which encourages confusions, even among philosophers, between principles with high (even universal) generality - e.g. "Treat equals equally" - and principles which actually give us determinate normative guidance.

We can now define our topic more clearly. In asking, "What level of comprehensiveness is achievable for principles of justice?", we are asking what sort of problems arise when we try to make principles *both* as general as possible *and* as determinate as possible. Maximally comprehensive principles of social justice would be principles which have both the greatest conceivable generality (their scope is society-wide) and the greatest conceivable determinacy (they tell us exactly what to do, down to the most specific level of practical decision-making, by always ruling out all but one alternative). What is the relationship between a principle's level of comprehensiveness and the difficulty of justifying it? Are less comprehensive principles generally less difficult to "justify" (whatever that comes to) than more comprehensive principles? Can justified principles of justice achieve a high level of comprehensiveness, or do they often merely combine high generality with low determinacy, and therefore deliver a very small normative payload?

As a step toward getting clear about these questions, we must distinguish two types of principles of justice: master principles and what Rawls calls "commonsense precepts" (e.g. 36). Commonsense precepts are imprecise, often highly context-dependent considerations which people invoke in everyday life, both when they *make* justice-based claims and when, as decision-makers, they balance competing justice-based claims in an intuitively satisfying way. Such precepts are unsatisfactory, because they often conflict with other precepts and leave immense scope for self-serving bias. It will do little good to confront "the chaos of competing claims" with a chaos of competing precepts.

Master principles of justice are proposals to infuse this chaos of intuitive precepts with at least some degree of rational order. A master principle consists of an internally often complex set of principles that (1) individuate a fixed and usually quite short list of distributive goods along with criteria according to which these goods are to be distributed; (2) are offered as jointly sufficient conditions to determine what is socially just or unjust; (3) have, or are intended to have, the property of internal consistency, that is, they do not mandate mutually exclusive distributions (although they may leave difficult balancing problems unresolved); and (4) are defended as justifiable in ways which any reasonable person (i.e. any person who is interested in finding publicly defensible principles of social cooperation) must accept. Each of these objectives poses serious difficulties.

It will make our discussion more concrete to have on hand several prominent examples of what I call "master principles", excerpted from the contemporary justice literature:

> a. Rawls' master principle of social justice:
> "(1) Each person has an equal right to a fully adequate scheme of equal basic liberties which is compatible with a similar scheme of liberties for all." [This scheme includes: "freedom of thought and liberty of conscience; the political liberties and freedom of association, as well as the freedoms specified by the liberty and integrity of the person; and finally, the rights and liberties covered by the rule of law."]
> (2) Social and economic inequalities are to satisfy two conditions. First, they must be attached to offices and positions open to all under conditions of fair equality of opportunity; and second, they must be to the greatest benefit of the least advantaged members of society."
> (1987a: 3; cf. 1977: 160-1)

b. Nielsen's master principle of social justice:

"(1) Each person is to have an equal right to the most extensive total system of equal basic liberties and opportunities (including equal opportunities for meaningful work, for self-determination and political and economic participation) compatible with a similar treatment of all...

(2) After provisions are made for common social (community) values, for capital overhead to preserve the society's productive capacity, allowances made for differing unmanipulated needs and preferences, and due weight is given to the just entitlements of individuals, the income and wealth (the common stock of means) is to be so divided that each person will have a right to an equal share. The necessary burdens requisite to enhance human well-being are also to be equally shared, subject, of course, to limitations by differing abilities and differing situations." (289)

c. Narveson's master principle of social justice:

"...our sole basic duty is to refrain from utilizing the fundamental resources of others without their consent; and those resources include, at a minimum, the bodies and minds of those others. The implications I will suppose we can draw from this are these, at least:

(1) We have no fundamental general duty to provide others with such goods as the necessities of life, let alone some particular proportion of all the socially distributable goods there are;

(2) we have in general a duty not to interfere with the operations the 'market', so long as it is truly a free market and not something else masquerading as same;

(3) we should always in general prefer voluntary social arrangements to involuntary ones whenever this is definable and feasible.

(4) Governments, in particular, are severely restricted in what they may properly do..." (1988, 166)

d. Phillips' master principle of social justice:

"In a just social order, there must be

(a) welfare rights to the provision of such goods and services as free and comprehensive medical care and education, public housing, and guaranteed employment; plus

(b) a redistribution of wealth and income to assure that everyone is enabled to obtain an adequate level of these and other necessary goods and services; and

(c) the implementation of rectification policies to rectify earlier systematic injustices.

After these stipulations have been met,

(d) any inequalities that remain in the distribution of income and wealth are permissible if and only if they do not lead to concentrations of wealth that are demonstrably detrimental to people's exercise of their generic rights" (427)

e. Sterba's master principle of social justice:

Principle of Need: Each person is guaranteed the primary social goods that are necessary to meet the normal costs of satisfying his basic needs in the society in which he lives. (The principle of need applies under the assumption that there are sufficient social and economic resources in the society to maintain the guaranteed minimum, but not necessarily without requiring a minimal contribution to society.)

Principle of Appropriation and Exchange: Additional primary social goods are to be distributed on the basis of private appropriation and voluntary agreement and exchange.

Principle of Minimal Contribution: A minimal contribution to society is required of those who are capable of contributing, when social and economic resources are insufficient to provide the guaranteed minimum to everyone in society without requiring that contribution or when the incentive to contribute to society would otherwise be adversely affected, so that persons would not maximize their contribution to society.

Principle of Saving: The rate of saving for each generation should represent its fair contribution toward realizing and maintaining a society in which the members can fully enjoy the benefit of its just institutions. (61)

f. Nozick's master principle of social justice:

"Ignoring acquisition and rectification, we might say: From each according to what he chooses to do, to each according to what he makes for himself (perhaps with the contracted aid of others) and what others choose to do for him and choose to give him of what they've been given previously (under this maxim) and haven't yet expended or transferred." (160)

§3 One caveat before we proceed. My analysis of principles in this chapter may seem to pile distinctions upon distinctions. I remain convinced, however, that this analysis proceeds at the right level of abstraction - the only level that promises to advance our understanding of social justice beyond the ideologically partisan. Despite this chapter's complexities, I am conscious of having unravelled only part of the inherent complexity of the subject matter.

I now want to distinguish four types of master principles, using two cross-cutting distinctions. The first distinction, already introduced, is between *directly responsive* and *indirectly responsive* principles; the second between *intuitionistic* and *non-intuitionistic* principles. A table will be useful to provide an overview.

	Principles which spell out a political response to individuals' justice-based claims by specifying what is to count as individuals' just distributive share:	*Principles which spell out a political response to individuals' justice-based claims by specifying what is to count as a just process or outcome of the total scheme of social institutions:*
Principles which do not spell out a way of balancing conflicting just-making criteria, but leave such balancing up to largely unguided intuition:	**Directly responsive, intuitionistic master principles of justice** (e.g. Nielsen, Phillips, most egalitarian theories, most rule consequentialist theories)	**Indirectly responsive, intuitionistic master principles of justice** (An empty category, because a principle cannot credibly claim to specify a just scheme of social institutions while leaving the balancing of conflicting criteria to intuition).
Principles which (claim to) avoid the need for intuitive balancing of criteria by specifying a single ultimate maximizing criterion or lexical ordering:	**Directly responsive, non-intuitionistic master principles of justice** (An empty category, because no directly responsive theory can credibly claim to be able to avoid the need for intuitive balancing of conflicting criteria)	**Indirectly responsive, non-intuitionistic master principles of justice** (Rawls, Sterba (?), libertarian theories)

Table II: Types of master principles of social justice

All principles of justice must, by definition, formulate an "appropriate", "reasonable" or "credible" response to what individuals claim as their due in pursuit of their personal well-being, based on their sense of justice. Indirectly responsive principles respond to the justice-based claims of individuals by specifying what is to count as a just scheme of social institutions and simply identify an individual's just share as whatever outcome such institutions happened to produce in that individual's case. Directly responsive principles respond to the justice-based claims of individuals by specifying what is to count as individuals' just distributive share and worry only later, if they worry about it at all, about the institutional means by which a society is to secure each individual's just share at any given time.

Indirectly responsive principles, which can either be material or procedural, will have some such general form as the following:

> "*Justice requires that the total scheme of social institutions be designed in such a way that the distribution of good A follows a process or achieves an outcome that meets criterion x, and the distribution of goods B, C, D, and E follows a process or achieves an outcome that meets criteria x, y..... in lexical order.*"

The clearest examples of indirectly responsive *procedural* master principles are libertarian principles, which specify the free market as the fundamental social institution which all other institutions are designed to uphold, and which simply specify that all goods should be distributed according to the basic right of individuals to use their own minds and bodies to make themselves better off without thereby making others worse off - which is to say, in practice, according to supply and demand in the market. The paradigm examples of indirectly responsive *material* master principles are, of course, Rawls' master principle and various traditional forms of utilitarianism. Although it may not be immediately obvious from

comparing Rawls' master principles with other non-libertarian principles, his principle is not only the paradigm example of a non-utilitarian indirectly responsive material master principle, but also the only remotely credible candidate of any such principle. Other justice theorists proposing more complex material principles may also claim to fix on the basic structure rather than on individual distributive shares, but such claims seem to me to involve a serious misunderstanding of what it means for a principle to fix on the basic structure of society organized as a whole, and of what extreme simplification would be required to get any sort of credible normative determinacy out of such a principle (see §5-8 below). Although Rawls' theory is surely not the only possible indirectly responsive liberal-egalitarian theory of justice, all other non-libertarian theories yet proposed seem to me most charitably reinterpreted as directly responsive.

Traditional forms of utilitarianism are distribution-insensitive, although they claim to be able to account for most of people's considered judgments about justice by appealing to such ideas as the principle of diminishing returns or the frequently bad side effects of what are intuitively considered injustices. Traditional utilitarians only specify one basic just-making criterion: that the total set of goods in a society should be distributed according to whatever will maximize utility - a term that can be cashed out as average or total happiness or pleasure, preference-satisfaction, or intrinsic value of whatever sort. Some utilitarians postulate equality as a subsidiary criterion which kicks in whenever several alternative distributions are equally compatible with maximizing utility. I shall not rehearse the well-known criticisms of traditional utilitarianism's inability to respect the separateness of persons.

§4 Directly responsive principles, which can be procedural or material, will have some such general form as the following (which can obviously be further complicated in many ways):

"Justice requires that good A be given to individuals who meet criterion x, goods B, C and D to individuals who meet criterion y, and good E to individuals who meet criteria z and w."

Given such a principle, it should theoretically be deducible, for any given individual who may meet some criteria but not others, what his just share consists in. How difficult this calculation is would seem to depend primarily on the scope of a principle's application, the determinacy of the particular just-making criteria invoked and the extent to which balancing problems arise in particular contexts of application (see below).

As noted in the previous chapter, in the case of directly responsive principles, the term "justice" does not generally encompass all political values that make up a vision of the good society or even only those that ordinarily constrain social distributions. Rather, the demands of justice will generally have to be balanced against the demands of those other values. In the case of indirectly responsive principles, however, there would seem to be no room for such balancing. The whole point of taking the total scheme of social institutions as the subject of justice is to avoid the complexities of having to keep track of whether each individual has more or less than their just share at any given time with respect to some good, while yet doing a sufficiently credible job overall of responding to concerns about justice so that there is a realistic hope of enlisting enough people's sense of justice in support of this scheme to achieve social stability. But having developed a blueprint for such a scheme, could someone like Rawls argue that we can regularly adjust this scheme to accommodate competing social objectives? The answer seems to be clearly no, except in national emergencies or other extraordinary circumstances.[2] Someone who takes individual shares as the subject of justice can

---

[2] I have already remarked on Rawls' ambiguity on this point. His introductory paragraph on the "role of justice"(1) affirms the thesis of the indefeasibility of justice; but a few pages later he says that "A complete conception defining principles for all the virtues of the basic structure, *together with their respective weights when they conflict*, is more than a conception of justice; it is a social ideal."(9, emphasis added)

easily argue that we may regrettably have to commit an injustice against particular individuals to avoid even greater injustices against others or to achieve more important social goals. But someone who takes the basic structure as subject cannot argue that we can regularly modify the just scheme of basic social institutions to accommodate competing considerations, both because once established such a scheme is not easily modified, but more importantly because permitting unprincipled ad hoc modifications would seem to undermine the ever-precarious ability of such a scheme to do a sufficiently credible job overall of responding to the pattern of justice-based expectations and sentiments so that there is a realistic hope of enlisting enough people's sense of justice in support of this scheme. It would therefore seem that, unlike a directly responsive principle, an indirectly responsive master principle of justice cannot simply leave aside traditional non-justice considerations, but must somehow amalgamate them all in one coherent principle which can feasibly constrain the design of society's basic institutional arrangements.

§5 The second distinction between different types of master principles, which cross-cuts the first, is the distinction between *intuitionistic* and *non-intuitionistic* principles. Intuitionistic principles contain several different just-making criteria and leave the balancing of at least *some* of these criteria up to unguided intuition (cf. Rawls, 34-45). Clear examples of strongly intuitionistic principles are Nielsen's and Phillips' master principles, despite some superficial attempts at a priority ordering.

Non-intuitionistic principles can specify either one or several criteria. Examples of single-criterion principles, or as Rawls calls them, "single-principle conceptions with one ultimate standard"(41), are necessarily non-intuitionistic because they propose to determine the justice of all distributions *ultimately* on the basis of a single criterion. They may, of course, specify any number of subsidiary criteria, but they are ultimately single-criterion principles as long as they specify a common currency, such as happiness-maximization, in terms of which the relative

importance of each of these sub-criteria may at least in principle be assessed when they conflict. Prominent examples are the simplest utilitarian principles which define the morally right (including the just) as promoting the greatest total or average happiness or preference satisfaction; and libertarian principles, such as Nozick's or Narveson's, which define the socially just as whatever is distributed on the basis of uncoerced individual decisions or free market transactions. But non-intuitionistic principles can also specify several ultimate criteria and avoid conflict between them by specifying a "lexical ordering". At least in principle, a lexical ordering of principles obviates the need for intuitive balancing.[3] Examples of non-intuitionistic principles which make a serious attempt at lexical ordering are Rawls' and, far less credibly, Sterba's, Nielsen's and Phillips' master principles as well.

3.2.2 Sources of rational disagreement about principles

§1 Now that we have analytically informative distinctions, we can confront the question of what problems different types of master principle face. What type of master principle offers the best hope for providing a rational basis to settle conflicts about justice? What types of master principles can be confidently eliminated as being beyond the scope of the rationally permissible? Or must we perhaps fall back altogether on common-sense precepts in our arguments about justice?

These are urgent questions; and considerably easier to ask than to answer. One way of justifying an answer to these questions, and probably the most satisfactory way, would be to examine the whole range of contemporary theories of justice in detail, to achieve the widest possible reflective equilibrium of relevant considerations. No such monumental effort is feasible here. An interesting project might be structured around a comparative study of the difficulties associated with a variety of theories. In fact, several critical surveys of contemporary theories of

---

[3] See Rawls, 42-3, for an account of lexical ordering.

justice already exist (cf. e.g. Kymlicka, Campbell, Pettit and others). My aim here is not to simply follow in their footsteps, but to move beyond these critiques toward some general conclusions about *important sources and limits of rational disagreement about justice*. In an effort to assess the prospects of various master-principle approaches to social justice, I shall first try to define the main problems of each type of master principle in the abstract and then assess the way John Rawls tried to surmount these problems. I focus on Rawls' theory, not because its faults are particularly glaring, but because, on the contrary, it would be hard to surpass many of its virtues and especially Rawls' awareness of balancing problems. Rawls' work still has a lot to teach us about the limits of reason in deciding on the rationally required content for master principles of social justice. If Rawls' master principle, despite the ingenious simplifications it incorporates, proves unworkable, other master-principle approaches of society-wide scope which do *not* incorporate these simplifications are also likely to fail. In fact, many master principles proposed in the wake of Rawls' work seem to me to stumble over precisely the sort of difficulties which Rawls' theory was designed to circumvent.

Rawls' motivation for seeking an indirectly responsive, non-intuitionistic master principle (which incorporates massive simplifications even at the risk of putting it fundamentally at odds with many people's sense of justice) is best appreciated in the light of the difficulties faced by alternatives. Four major problems faced by master principles of social justice are the justification problem, the priority problem, the aggregation-distribution problem and various remaining commensuration problems. Let me clarify the nature of these problems.

§2 The justification problem consists in giving a compelling rationale - compelling, that is, for any person with an interest in publicly defensible rules of social cooperation - for accepting a particular master principle over many competing principles of justice at various possible levels of comprehensiveness. Different

theories seek to resolve this problem by appealing to a range of mutually incompatible meta-ethical positions. Although explicit appeals to God's will or design have gone out of fashion in Western analytic philosophy, neo-Aristotelian appeals to the human good, realist appeals to natural rights or categorical imperatives, contractarian and rights-based constructivism, unabashed intuitionism, and especially invocations of intuitively appealing forms of equality, as well as a pragmatic liberalism based on a mixture of relativism and meta-ethical despair are all very much in evidence. The best one can do in matters of ultimate moral justification, I have argued, is to seek a wide reflective equilibrium which tries to work toward a principled understanding of the basic nature of morality by assessing all the *prima facie* plausible meta-ethical alternatives in the light of considered moral judgments and relevant background theories, including the scientific world view. I have tried to establish some presumption that a wide reflective equilibrium supports a form of consequentialism, although much detail remains to be filled in. According to this moral framework, a theory of justice must be judged, relative to alternatives, by the extent to which it shows promise to help accomplish morality's main job: to create enough uniformity and thus predictability in people's behavior so as to generate a sense of security and trust to allow mutually beneficial interaction. More precisely, the ideal consequentialist goal is to help achieve that combination of satisfied interests of individuals in their own well-being which takes seriously the distinction between persons. We have found reason to think that this goal can only be achieved if we take very seriously the claims individuals make to social resources in pursuit of their personal well-being, and which they back up with their sense of justice. That means that nothing can count as a justified master principle of justice if it does not respond to the pattern of expectations and sentiment about justice in a way so as to have credible potential to harness enough people's sense of justice in support of it to make this principle the basis for a stable social consensus about justice. Another necessary condition for a

rationally acceptable principle of justice is that it must not involve unworkable balancing problems which would completely vitiate its normative determinacy. How well a principle meets these conditions is a matter of degree, at least up to a hard-to-define point at which a principle may asymptotically approach zero determinacy. Of those principles which meet these necessary conditions equally well, taking the principle with the greatest comprehensiveness would seem to make the most rational sense, because the point of a theory of justice is, after all, to pull people's conflicting judgments about justice as far up to the level of principles as possible. What can count as having "justified" a principle of justice is obviously not a subject that admits of very great precision.

While our consequentialist moral framework rules out *some* proposed master principles as complete non-starters, it requires us to assess the comparative merits of many others in the laborious way in which I will try to assess the merits of Rawls' principle. Despite Rawls' self-proclaimed anti-utilitarianism, there is nothing inconsistent about assessing the rationale for Rawls' two principles from within a consequentialist framework. Since I defend indirect consequentialism, it is an open question what sort of principles of justice will most plausibly promote the ideal consequentialist goal. Rawls' two principles are certainly among the *prima facie* contenders. If I ultimately reject Rawls' principles, it is not simply because the justification Rawls gives for them conflicts with a consequentialist framework, but because they fail Rawls' own necessary conditions - which are also plausible conditions from a consequentialist perspective - for any adequate theory of justice.

§3 **The priority problem** is how to weigh the relative importance of the various just-making criteria invoked by an internally complex principle whenever these criteria conflict (cf. Rawls, 40-45). Such conflicts can take several possible forms: two criteria may demand the same scarce resource; or one criterion may qualify a person for obtaining a certain good and another may disqualify him from

obtaining it. Solving the priority problem exacts a steep price. It means that one must either accept a single-criterion conception of justice or else specify a lexical ordering. Single-criterion conceptions of justice end up with a criterion which either gives merely semantic unity to an underlying pluralism of values or else has no plausible potential to enlist many people's sense of justice. Lexical ordering tends to be vulnerable to obvious counterexamples, because it specifies an absolute order of priority for principles and prohibits any trade-offs between a minimally lesser satisfaction of principles of higher priority even for the sake of a hugely improved satisfaction of principles of lower priority. Indirectly responsive principles *must* solve the priority problem (and other balancing problems) through radical simplifications, because principles which specify constraints for the design of society's basic structure while leaving open the question of how to balance conflicting just-making criteria would have no normative determinacy whatever, and *de facto* amount to having no principle at all (see below).

Solving the priority problem is far less urgent for directly responsive principles. I have called those directly responsive principles which solve the priority problem (and pay the price for solving it) "non-intuitionistic", and those which don't solve it "intuitionistic". It is, of course, a matter of degree how intuitionistic a directly responsive principle is - that is, to what extent it leaves the balancing of criteria up to the moral intuitions of relevant decision-makers. It is important to see why it is *less* crucial for directly responsive principles than for indirectly responsive principles to solve the priority problem and other balancing problems. Allowing some degree of balancing of conflicting just-making criteria by unguided intuition is less damaging for principles which specify the justice of individual distributive shares rather than of society's basic structure for several reasons. The local context in which an individual's share will have to be determined will generally impose additional constraints on the intuitions of decision-makers; grievances can be dealt with locally, often in face-to-face contact between the parties most affected;

procedures can often be more easily agreed upon within such contexts because the people involved have a direct personal stake in resolving disputes about justice; and the injustices committed will be less momentous because confined to more or less restricted domains. Material master principles which specify a plurality of criteria for determining just individual shares in a plurality of goods will have to be supplemented with purely procedural principles and allow for a great deal of context-dependent interpretation. But in the end, procedural principles (supplemented by a guaranteed social minimum) may often be the only type of master principle we can get in many domains. Nevertheless, leaving the interpretation of a directly responsive, intuitionistic master principle up to contextual balancing of conflicting criteria is far from ideal and, as Rawls stressed, may often be little better than having no principle at all (cf. 34-40).

§4 **The aggregation-distribution problem:** All principled ways of responding to individuals' competing justice-based claims in a credible way will necessarily involve both distributive and aggregative considerations. The problem then becomes how to balance these two types of considerations. The problem (which the thought experiment of imagining oneself living *all* lives in sequence was meant to resolve in theory) gets more and more intractable in practice as a principle's domain of application expands over a heterogeneous range of goods and a heterogeneous range of beneficiaries of those goods. Even if we do not include additional values into our list of just-making criteria except those considerations which are commonly invoked as relevant to determining justice, we still have two objectives: satisfying individuals' competing justice-based claims to the greatest possible extent, and satisfying them in a way that respects the separateness of persons. And those objectives may easily conflict. Satisfying the important claims of even the poorest fraction of the population may mean forgoing the satisfaction of a greater amount of less important claims of the rest of the population. The larger and more

heterogeneous the domain to which a principle applies, the more intractable this balancing problem would seem to become, because we have to make a multitude of qualitative judgments of the relative importance of different claims of different individuals. Only by factoring this problem into small parts, that is, by either breaking the problem of social justice up into a plurality of particular domains or (possibly) by using Rawls' device of judging the justice of the total scheme of social institutions from a representative social position, can we hope to be sufficiently sensitive to the disparate justice-based claims of enough individuals to harness their sense of justice for social cooperation.

Rawls tried to simplify the aggregation-distribution problem in a bold and highly original way. Instead of giving up on the idea of a master principle of society-wide scope, he introduced one of the most controversial simplifications of his theory, the idea of a representative social position.[4] In trying to balance satisfaction of the greatest aggregate of justice-based claims against the satisfaction of the claims of greater importance, he asks us to assume the position of a representative worst-off person and and to arrange the total scheme of social institutions in such a way that the justice-based claims of this person are satisfied to the highest possible extent (cf. Rawls, 64f). There is obviously a lot to be said for judging the demands of justice from the perspective of those whose needs are greatest, given that any attempt to judge them from *everyone's* perspective simultaneously gets us into the complexities of the thought experiment of imagining oneself living all lives in sequence. Still, for persons whose sense of identity is highly tied up with their social class, Rawls' strategy of maximizing the expectations of one particular social position (that of the representative worst-off person) hardly seems to guarantee

---

[4] Rawls, of course, describes *two* types representative social position: the position of equal citizenship, from which we are to judge whether the liberty principle is satisfied; and the position of the worst off representative person, from which we are to judge whether the difference principle is satisfied (cf. 64f). I shall only concern myself with the social position of the worst off representative person in what follows.

an acceptable level of respect for the separateness of persons without assuming an extremely dubious chain connection of benefits rippling upward through the different social strata as a result of maximizing the lot of the worst off.

From the perspective of distribution-sensitive consequentialism, in contrast, we can generally pronounce outcomes acceptable which would both produce more total satisfaction of justice-based claims more equally distributed than any other feasible outcome (assuming a safety net for the very worst-off). We can also say with confidence that an outcome which increases some plausible index of aggregate satisfaction at the expense of unequal distribution will eventually become just as intolerable as an outcome which increases the equal distribution of satisfaction at the expense of every plausible index of aggregate satisfaction. But within broad limits there seems to be no general way to rank, in the abstract, alternative combinations of aggregate satisfaction and degrees of equal distribution. Here perhaps any theory of justice must admit limits to rational decidability in the abstract and defer to more context-dependent reflective equilibria, which is to say, defer to the wisdom of decision-making bodies faced with concrete distributive problems.

§5 **The commensuration problem** is the set of balancing problems left after the priority and aggregation-distribution problems are resolved. Such additional balancing problems arise when we try to compare the value of the different holdings of all the different individuals in society in such a way that we can tell, at any given time, what redistributions are required from those whose holdings exceed their just share to those whose holdings fall short of it. The problem of tracking "the endless variety of circumstances and the changing relative positions of particular persons", as Rawls so clearly realized, makes it seem like a forgone conclusion that any directly responsive principle of society-wide scope will be unworkable in practice. Rawls emphasizes the virtue of his approach in the face of the commensuration

problem:[5]

> Now the great practical advantage of pure procedural justice [i.e. letting a just basic structure take care of individual shares, without individuals being able to claim redress in the name of justice] is that it is no longer necessary in meeting the demands of justice to keep track of the endless variety of circumstances and the changing relative positions of particular persons. One avoids the problem of defining principles to cope with the enormous complexities which would arise if such details were relevant. It is a mistake to focus attention on the varying relative positions of individuals and to require that every change, considered as a single transaction viewed in isolation, be in itself just. It is the arrangement of the basic structure which is to be judged, and judged from a general point of view. Unless we are prepared to criticize it from the standpoint of a relevant representative man in some particular position, we have no complaint against it. Thus the acceptance of the two principles [governing the basic structure] constitutes an understanding to discard as irrelevant as a matter of social justice much of the information and many of the complications of everyday life. (87-8)

We can divide the commensuration problem into three basic parts, and different master principles are mainly distinguished according to how they cope with each different part of it.

(1) Take a principle of society-wide scope which defines just individual shares. Determining whether someone has a just grievance requires us to determine whether a particular individual's share at a given moment is actually just. First, we would have to find ways to establish the total value of an individual's holdings of relevant distributive goods relative to the value of the holdings of all other individuals in society in such a way that we can tell whether he exceeds his just share or falls short of it. The holdings of different individuals will, of course, be composed of different types of goods and different amounts of each type, and the same good will have different value for different persons. If we are to compare the value of one person's holdings relative to another person's, we need to have a

---

[5] I cannot discover that he gave it any particular name in *A Theory of Justice*. In "The Priority of Right and Ideas of the Good", however, Rawls calls something closely related to it "the problem of interpersonal comparisons of citizens' well-being": "given the conflicting comprehensive conceptions of the good [individuals have], how is it possible to reach a political understanding of what is to count as appropriate claims?"(1988, 255-6).

common measure to be able to add up the value of different goods a person may have.[6] We could, of course, reduce the index problem to the extent that we specify a separate criterion for different goods; but that has obvious drawbacks. It will escalate the priority problem, insofar as different goods ultimately depend on the same resources to produce them. Moreover, it would obviously be unjust not to take into account, for example, that some people may have more money, but other people may choose to compensate for that by having more leisure. The fact is that there is no good index to compare the value of diverse goods, from the point of view of justice. The market value of goods is notoriously problematic, partly because the market value of some important goods for which there now exists no regular market is hard to determine, and partly because markets are distorted by the preferences of the already affluent. We could use the amount of happiness the person derives from each good to index the total value of each person's holdings and then compare his holdings to everyone else's, to judge whether his distributive share was above or below average in value. But interpersonal comparisons of mental states are also extremely problematic. One popular way of reducing the index problem without escalating the priority problem or giving rise to obvious injustice is, of course, simply to specify standards of minimum provision for certain strategic goods, regardless of what other goods a person may have, and often regardless of whether someone finds himself short of such goods through his own fault.

(2) Having indexed the total value of a person's holdings according to some common measure, one would then in principle have to do the same with every other person's holdings, because redistributing to those who have less than their just share may require taking from those who have more, and to do that we have to know who is in which relative position at any given time (or perhaps over a whole lifetime). Either case requires keeping track of the endless variety of individuals'

---

[6] This part of the commensuration problem was called the "index problem" by Rawls (cf 93f).

circumstances and changing relative positions which Rawls mentions. Although computer technology has allowed great progress in keeping updated records of individuals' changing circumstances, a centralized information bank on all individuals of this sort obviously gives rise to other troubling moral issues.

(3) Having determined who has too much and who too little, we now have to devise effective, cheap, peaceful and fairly continuous ways of redistribution, to reestablish or maintain a just balance of benefits and burdens for each individual. There are many additional problems which will arise in considering various options for accomplishing redistributions on a massive, continuous scale.

The magnitude of the commensuration problem a directly responsive master principle faces will vary, in complex ways, with its scope of application, the number and concreteness of distributive goods it individuates, as well as with other particulars of a principle. As Mapel sums up the difficulties of directly responsive master principles of ambitious scope that try to "provide standards for redressing the overall situation of each member of society",

> [They] must find some coherent and practicable way of distinguishing the natural, social and personal factors that together presumably determine what individuals are capable of claiming. Of the many obvious difficulties....I will note only the most salient: (1) the impossibility of distinguishing between the holdings that have been acquired by actual members of society and the holdings that might have been acquired in another society or a 'state of nature'; (2) the impossibility of establishing the counterfactual claims necessary to justify policies of direct compensation to individuals for discrimination or unfairness occurring any time in a more or less remote past; (3) the necessity of nevertheless rectifying the cumulative effects of individual transactions, both free and fraudulent; (4) the impossibility of regarding market demand as an objective measure of the worth of individual contributions, together with the impossibility of commensurating the value of most goods and services at all except through the operation of markets; (5) the conceptual and practical difficulty of assessing welfare, developmental or otherwise; and (6) the requirement that principles of justice nevertheless allow reasonable economic efficiency and political stability. (147)

Directly responsive master principles of justice of ambitious scope may enunciate noble-sounding manifestos of an individual's just share. But if all the above

difficulties stand in the way of their implementation, they may simply serve to arm many people with yet another mischief-causing rhetorical device for pressing unsatisfiable claims in the name of justice. It clearly took genius not only to conceive the possibility of a radically different approach, but to develop it against seemingly hopeless odds to the degree of sophistication Rawls did. Our question now is whether Rawls' indirectly responsive master principle can, in fact, escape balancing problems of unworkable proportions without, at the same time, simplifying beyond the bounds of what is acceptable if his principle is to have credible potential to enlist rational people's sense of justice in support of it.

3.2.3 Rawls' master principle

§1 From the perspective of conceptual intelligibility alone, the number of possible principles of justice is obviously huge, because of the immense diversity of possible goods to which some combination of just-making criteria might conceivably be considered relevant. A master-principle of social justice must individuate those types of goods it considers strategic enough to become the concern of social justice, and must align these goods with a particular combination of just-making criteria whose demands should be met.

The fundamental normative idea embodied in Rawls' master principle is simply that inequalities are unjust unless they work for everyone's advantage: "Injustice, then, is simply inequalities that are not to the benefit of all."(62) He distinguishes between political and economic "primary goods" and argues that, given a certain minimum level of material prosperity, any inequalities in people's "right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all" could never work to everyone's advantage (250; 541f). But when it comes to social and economic inequalities, he thinks that a just system of social institutions must allow even large inequalities, because certain incentive structures motivate individuals to exert effort and develop their talents to

create economic wealth which, despite producing relative disparities, will make all individuals better off in absolute terms.

Rawls supports these conclusions essentially with two arguments linked in a wide reflective equilibrium: a *narrow-reflective-equilibrium argument* that claims that these conclusions cohere with our considered moral judgments, "duly pruned and adjusted", about what is just or unjust in particular cases; and a *social contract argument* which derives these conclusions from an original position whose description is in turn supported by considered moral judgments and relevant background theories of various sorts. Since Rawls' arguments are now well known yet difficult to state concisely, I shall not rehearse them in detail here.[7] The following summary of the essentials of two decades of criticism of Rawls' theory is meant to be intelligible only to those already reasonably conversant with the literature.

What are the main ways in which Rawls' theory of justice is widely thought to fall short in justifying his proposed master principle as a principle of justice which any rational person must accept? The social contract argument seems to many to fall short on two main counts. It is argued that Rawls fails to show that his description of the original position is in reflective equilibrium with everyone's considered moral judgments about what constitutes necessary and sufficient conditions for a fair choice situation. Moreover, even if his description of the choice situation were in reflective equilibrium with everyone's considered moral judgments, Rawls fails to show convincingly that in this position rational choosers would necessarily pick out his favored principles.

Rawls' narrow-reflective equilibrium argument (which tries to work from our considered judgment in particular cases to general principles) is also widely seen to fall seriously short. Rawls' favored principles cannot be shown any more

---

[7] Summaries are to be found in Barry (1973, 1989); Kymlicka (1990) or Campbell (1988), and in almost any other text on contemporary political philosophy. But there is no substitute for Rawls' own work.

convincingly to be in reflective equilibrium with different people's considered judgments than a multitude of competing principles within a very broad range. In fact, most people's considered judgment remains stubbornly opposed to Rawls' rejection of desert as a just-making criterion. Finally, it seems now commonly accepted that Rawls *wide* reflective equilibrium argument is ultimately inconclusive. He fails to show convincingly that there is a sufficiently determinate set of principles reached via the original position which overlaps another sufficiently determinate set of principles reached in narrow reflective equilibrium with most people's considered judgments in exactly such a way that this overlap contains the following two principles:

> (1) Each person has an equal right to the most extensive scheme of equal basic liberties compatible with a similar scheme of liberties for all.
> (2) Social and economic inequalities are permissible provided that (a) they are to the greatest expected benefit of the least advantaged; and (b) attached to positions and offices open to all under conditions of fair equality of opportunity. (1977: 160-1)

I will not offer a detailed rehearsal of all the seemingly devastating and easily accessible arguments of Rawls' critics, but I do want to make a general point about criticizing Rawls. Although I think it is true that, even considered in combination, Rawls' arguments do not come anywhere near justifying his favored principles as uniquely required by reason, *much* of the criticism he has received seems to me to miss the point of his theory. Much of this criticism depends on taking Rawls' derivation of the two principles from the original position in isolation and also on the fact that we simply do not have sufficiently precise criteria by which to decide what *counts* as having shown a master principle of justice to be required by reason. But an extremely plausible necessary condition for any rationally acceptable theory of justice is that it should capture the pattern of expectations and sentiment about justice in a way that has credible potential to enlist enough people's sense of justice to make it the basis for a stable social consensus about justice *without*, however,

involving balancing problems that would completely empty it of any normative determinacy in practice. If the degree of a theory's rational acceptability is judged by this criterion, then it won't do for critics to argue that Rawls' theory hasn't captured their considered judgments about justice and then propose in its place yet another noble-sounding ideal of just individual shares which cannot be justified over dozens of similar ideals and which, in addition, involves utterly unworkable balancing problems.

§2 Rawls' objective, as I understand it, was to construct a theory which captures a sufficient part of the pattern of expectations and sentiment about justice so as to have credible potential to make enough people, upon reflection, accept this theory as sufficiently responsive to their justice-based claims to allow it to serve as a basis for an overlapping social consensus about justice, at least in large, reasonably prosperous pluralist societies. "However attractive a conception of justice might be on other grounds", Rawls says,

> it is seriously defective if the principles of moral psychology are such that it fails to engender in human beings the requisite desire to act upon it. (455)

However philosophically elegant, ingenious or provocative a theory of justice may otherwise be, it must have credible potential to allow justice to play the social role for which a theory of justice is needed in the first place:

> to enable all members of society to make mutually acceptable to one another their shared institutions and basic arrangements, by citing what are publicly recognized as sufficient reasons. (1980, 517)

And it is hard to see what could give a theory of justice such credible potential to do that, if not a high degree of responsiveness to the existing pattern of expectations and sentiment about justice along with persuasive justifications for disappointing some of these expectations.

At the same time, an acceptable theory of justice must incorporate as much

simplification as is necessary to keep the priority, aggregation-distribution and commensuration problems within manageable bounds. As Mapel and others have convincingly argued, the alternative master principles offered by Rawls' non-libertarian critics such as Dworkin, Galston or Ackerman pose insurmountable balancing problems and require "second-best" approximations which are unrecognizable travesties of the original principle.[8] And having an unworkable theory of justice may well be worse than having no theory at all, because the unsatisfiable expectations raised by such a theory may only serve to stir up a destabilizing sense of injustice. There is only one class of theories of equally ambitious scope which escape these problems, namely libertarian theories whose only just-making criterion is a particular interpretation of free choice and which leave it completely up to the market to commensurate the value of different goods for different individuals. But avoiding unworkable balancing problems is only one plausible necessary condition for an acceptable theory of justice; capturing a sufficient part of people's expectation and sentiment about justice to have credible potential to become acceptable to them as an account of *justice* is another. In this respect, I think libertarian theories fall short (see Ch 3.5).

So Rawls seems easy to criticize but, given his extremely plausible objective, he is certainly not refutable simply by looking at each stage of his argument separately and showing that neither his description of the original position, nor his favored principles, are uniquely picked out by these arguments. Rawls' project was to achieve a wide enough reflective equilibrium which simultaneously takes every conceivably relevant consideration into account to decide which master principle of society-wide scope captures people's expectations and sentiment about justice to the greatest extent possible while requiring a minimum of controversial simplifications

---

[8] Cf. e.g. Campbell (1988); Mapel (1989) and Kymlicka (1990) for detailed critiques of these and other theories. As Kymlicka remarks, "How can we fairly implement [Dworkin's] insurance scheme? We tax the rich, even though some got there purely by effort with no natural advantage, and support the poor, even though some...are there by choice without any natural disadvantage."(82).

to keep the balancing problems manageable. He claims to have achieved such a wide reflective equilibrium and has presented a awe-inspiring architectonic of philosophical insight to back up this claim.

§3 I see only two ways to criticize Rawls without completely begging the question about what counts as a necessary condition for a rationally justifiable theory of justice. First, one can attempt to develop an even more sophisticated architectonic than Rawls does, within an even wider reflective equilibrium, which will make apparent that a very different master principle than Rawls' best balances the potential to capture people's expectations and sentiment about justice with a sufficient minimum of controversial simplifications to avoid unmanageable balancing problems. Second, one can attempt to show that his theory fails on its own terms and cannot plausibly meet the necessary conditions for a rationally acceptable master principle. The first attempt is likely to achieve at best inconclusive results because there are fundamental indeterminacies involved, both in clearly defining the various terms of the objective and in deciding what background beliefs are ultimately relevant. Although I can see no better *ultimate* method of theory acceptance than wide reflective equilibrium, it will by now seem obvious why this method, applied to the problem of social justice, is likely to allow for a great deal of conceptual, normative and empirical disagreement among informed people acting in good faith. Once a master principle plausibly meets the two necessary conditions for being rationally acceptable, there would seem to be no conclusive way of showing which master principle would be uniquely picked out by a wide reflective equilibrium which draws on all the relevant arguments and best available information.

But can *any* master principle of society-wide scope plausibly meet the two necessary conditions? Here is the second way of criticizing Rawls: asking whether he has succeeded in his own terms. *Does* his theory solve the priority, aggregation-

distribution and commensuration problems in a practically workable way while, at the same time, having the potential to capture enough people's expectations and sentiment about justice to have credible potential to become acceptable to them as doing *justice* and thus to serve as the basis for a stable social order? This is my guiding question in the rest of this chapter.

§4 It will be well, first, to remind ourselves of the essential components of Rawls' master principle. The four essential components of any master principle of justice are its primary subject; its scope or domain of application; the particular goods it individuates as the concern of justice; and the just-making criteria it invokes. Obviously, the potential number of *intelligible* combinations of these components is staggering. Every selection one makes from the field of the intelligible must have a rationale.

Rawls chooses a very high level of abstraction for all four components. His choice of the basic structure of society as the primary or direct subject of justice is familiar. His master principle is meant to have society-wide scope; that is, there is nothing left for social justice to be concerned about as long as this principle is satisfied. The basic structure he fixes on is not individuated at the level of everyday decision-making in society's existing institutions, but at the vastly more abstract level of a four-stage design of basic institutions in an ideally just society. The goods whose distribution society's basic structure is designed to constrain are individuated, not at the level of a plurality of very concrete goods, but again at an abstract level of what Rawls calls "primary goods". Rawls' five types of primary goods are basic rights and liberties, powers and opportunities, and the social bases of self-respect. Primary goods are supposed to be neutral all-purpose means for carrying out any rational life-plan (1988, 257). And finally, Rawls' invokes three types of (lexically ordered) just-making criteria. In the domain of political goods, he invokes the criterion of (formal) equality of rights to the greatest system of equal basic liberties compatible

with similar liberties for all; in the domain of social and economic goods, the criterion of "maximin" inequality (inequality regarded as justified only if it maximizes the social and economic positions of the representative worst-off person) combined with the criterion of "fair" equality of opportunity in competing for offices and positions.

Rawls has an impressive array of rationales for the particular levels of abstraction he chooses. I want to review them in the briefest of terms. Rawls' rationale for making society's basic structure the focal concern of justice seems to be essentially threefold: (1) People have hopelessly conflicting intuitions about what is their just share in particular situations, which gives rise to "a chaos of competing claims"(99); and attempts to determine just individual shares by measuring each individual's bundle of goods by some ideal standard and then redistributing resources in a way to conform to such a standard gives rise to hopeless complexities. (2) It is the total scheme of basic social institutions which pervasively conditions the prospects of individuals and often causes conflict at the level of individual shares; and only by knowing what sorts of basic institutions are just can we make the necessary changes. (3) Only against the background of a system of social institutions which create certain opportunity structures and expectations does it even make sense to conceptualize an individual's desert, rights, or legitimate expectations (cf. 304-5; cf. also "The Basic Structure as Subject", 1977).

Next I want to briefly consider Rawls' basic rationale for fixing on the particular bundle of strategic goods he calls primary goods. Any master principle of justice faces an obvious dilemma when it comes to the individuation of goods whose distribution it regards as important enough to be the concern of justice. If it singles out only one type of good, say freedom or happiness, there will be the obvious objections that making only this good the concern of social justice is unacceptably lopsided, or unintelligibly abstract, or a purely verbal unification of an unruly plurality of goods. If a master principle of justice allows many different types

of goods to be the distributive concern of social justice, at the extreme anything that at least some people can plausibly value, there will be an exponentially growing index problem, unless a separate just-making criterion is specified for each good (but that gives rise to an unmanageable priority problem). Rawls' choice of a short list of goods is an obvious compromise between the vexing problems of the single-good and the many-goods approach. Rawls notes that a theory of justice can capitalize on "a partial similarity in the structure of citizens' permissible conceptions of the good":

> ...comprehensive conceptions of the good, however distinct their content and their related religious and philosophical doctrines, require for their advancement roughly the same primary goods, that is, the same basic rights, liberties, and opportunities, as well as the same all-purpose means such as income and wealth, all of which are secured by the same social bases of self-respect. (1988: 256-7; cf. *Theory*, 90-95)

The list of primary goods is meant to include all and only those things which almost anyone wants, and wants to the greatest possible extent, whatever else he may want, because they are necessary or useful means to pursuing almost any plausible conception of personal well-being (92).

Rawls' rationale for a lexical ordering of his principles parallels, in a way, his rationale for primary goods. Just as, in the individuation of social goods, the choice between a single-good approach and a many-goods approach presents a dilemma, so in the individuation of just-making criteria, the choice between a single-criterion approach and a many-criteria approach presents a dilemma. Single-criterion approaches (which may, for example, propose to distribute whatever goods they individuate purely on the basis of desert, or purely on the basis of need, or purely on the basis of particular rights, or purely on the basis of market success) face the objection of being unintelligibly abstract, unacceptably lopsided or unified only in a verbal sense. Many-criteria approaches will, of course, face a priority problem (for example, what to do when considerations of need conflict with considerations of desert), a problem which increases exponentially with the number of criteria. Rawls

well-known proposal to deal with this dilemma is again to seek a compromise between too many and too few criteria, and to solve the priority problem among those few criteria he thinks relevant to deciding the justice of the basic structure by trying to justify an ingenious lexical ordering for their application (cf. 42f; 61f; 82f; 151f).

Last I turn to Rawls' rationale for judging the justice of a distribution in terms of what will maximize the "expectations" of a representative member of the least advantaged group in society. Why should we judge the justice of the basic structure by what is desirable from the perspective of one particular social position, when strictly speaking there are as many perspectives as there are individuals? Privileging the perspecitve of one of countless social positions in this way - is that not itself an injustice? We have already touched on Rawls' dual rationale for this radical step for the sake of necessary simplification. It is natural for a person to think of social justice from his particular perspective - that is, to interpret just-making criteria in the light of goals and aspirations which reflect the social context in which that person has constructed his self. Justice-based claims are projections of such embedded selves. Social class is an especially strong determinant of what claims will be backed by a person's sense of justice, because people are naturally more likely to compare themselves with their approximate socio-economic equals than their inferiors, and to adjust their justice-based expectations accordingly. But strictly speaking, there are as many different positions in society from which justice can be judged as there are people, and depending on whose perspective we imagine ourselves to be taking, we would intuitively assess the relative merits of the conflicting justice-based claims of different individuals very differently. Rawls argues that thinking of social justice instead in terms of what will maximize the prospects of a "representative social position" can alone make the commensuration problem manageable by reducing it to the index problem (44; 64f; 95-100). Moreover, it simplifies the aggregation-distribution problem by giving us a single, coherent

perspective for assigning weights to conflicting aggregative and distributive considerations. Mapel lucidly explicates Rawls' rationale:

> By focussing on abstract social positions, rather than individuals, it is no longer necessary to follow and compare the claims of each and every individual within society. Instead, we can leave the matter of individual shares largely to pure procedural justice, as long as the least-advantaged are taken care of. The device of representative positions also enables us to deal with the problem of balancing philosophical first principles, by 'factoring the problem of social justice into small enough parts.' Since it is no longer necessary to try to maximize expectations with respect to more than one point of view...., the problem of balancing general aggregative and distributive considerations has been simplified considerably. These practical adjustments may well give us pause, but [they] seem necessary if we are to avoid the difficulties involved in trying to be fair to each and every individual and each and every social group that might possibly raise a claim on resources. The device of representative social positions seems the only solution to the problems. (Mapel, 98-9; cf. Rawls, 44; 95-100)

Given, then, that an indirectly responsive master principle can solve balancing problems *only* by judging the justice of distributive outcomes from the perspective of a particular social position, which position shall a defender of such a principle choose? The social position which Rawls picked to assess the justice of social and economic institutions is, of course, the position of the representative worst off person: both because the justifiably felt injustice seems most acute for this group and because, Rawls thinks, judging distributive outcomes by whether they maximize the prospects of the worst off segment of society allows for a "chain connection" of benefits rippling upward through society.

This completes my brief overview of Rawls' rationale for thinking of social justice in terms of designing social institutions around the goal of distributing a set of primary goods according to lexically ordered principles which stipulate the equal distribution of political rights and liberties, equality of opportunity in the distribution of access to offices and positions, and the maximization of an index of social and economic goods with respect to the worst off representative person. Although no such overview can possibly capture the subtleties of Rawls' own

argumentation, it will do for our purposes here.

§5 Is Rawls' attempt to meet the two necessary conditions for a master principle of justice by resorting to such simplifying devices as a focus on society's basic structure, primary goods, lexical ordering and representative social positions at least a qualified success? Does Rawls succeed on his own terms - providing a theory of justice of society-wide scope which has plausible potential to capture enough of the pattern of most people's expectations and sentiment about justice to become acceptable to them as a theory of *justice* and to enlist their sense of justice in support of these principles, while at the same time avoiding unmanageable complexities which would prevent his theory from giving concrete practical guidance?

Concerns about the normative determinacy of Rawls' theory

A. The basic structure: Rawls' seemingly compelling rationale for choosing the basic structure as subject raises troubling questions of its own. First, it is far from clear whether a sufficiently principled distinction between the basic structure and the rest of the social order can be drawn. Rawls defines "the basic structure" as "the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation."(7) What, then, are major social institutions?

> By major social institutions I understand the political constitution and the principal economic and social arrangements. Thus the legal protection of freedom of thought and liberty of conscience, competitive markets, private property in the means of production, and the monogamous family are examples of major social institutions. (7)

But Rawls himself admits, "Now admittedly the concept of the basic structure is somewhat vague. It is not always clear which institutions or features [of an ideally just society] should be included."(9) Brian Barry clarifies Rawls' concept of society's "basic structure" as follows:

[The basic structure] is made up of the institutions that together determine the access (or chances of access) of the members of a society to resources that are the means to the satisfaction of a wide variety of desires. These resources can be grouped under three headings: power, status, and money. Examples - this is not an exhaustive list - of institutions that fall under the basic structure of the society are the following: the rules that allocate fundamental legal rights and privileges (equal or unequal) to members of the society; the rules that specify how access is gained to political decision-making power; the rules that permit concentrations of private decision-making power, such as that exercised by business corporations, real-estate developers, and so on, in market societies; the institutions (including the educational system and rules of apprenticeship or licensure) that determine access to professions, crafts, or other sought-after occupational positions; and the whole complex of institutions (including some already mentioned but also including rules for the acquisition and inheritance of property and the system of taxation and transfers) that determine the distribution of income and wealth in a society. (1989:146)

Now indeed we have a clearer picture of what the "basic structure" involves. But if the basic structure breaks down into many separate domains, surely the question arises whether Rawls' focus on the basic structure is really a simplification. Is it likely that we can tell whether such a large variety of different "institutions" actually produce the pattern of distributive outcomes which Rawls' principles require? And how is the division of distributive labor among these many institutions to be decided?

If Rawls' master principle were meant to give guidance at the level of all these specific institutions which make up society's basic structure, then on the face of it they would seem to leave decision-makers largely at a loss. How can a connection be made between maximizing the prospects of the least advantaged representative person and particular policy decisions within institutions? In reply to such worries Rawls introduces "an elaboration of the original position", a "four-stage sequence that clarifies how the principles for institutions are to be applied"(194-201). Rawls' basic idea is "a division of labor between stages in which each deals with different questions of social justice"(199), accompanied by a complex process of gradual lifting of the veil of ignorance. Campbell gives a helpful

summary of Rawls' own rather difficult exposition of the four-stage sequence:

> ...in the first [stage] the two principles of justice are chosen; in the second there is a constitutional convention to set up a system of government; in the third ordinary laws are legislated and in the fourth these laws are applied by judges. During this process the veil of ignorance is progressively lifted somewhat so that at the second stage the constituent assembly knows the nature of the society in question, at the third stage the legislators know the basic economic facts about their society and at the fourth stage the veil of ignorance is totally removed so that for the first time citizens know their own circumstances and characteristics. (81)

But Rawls' instructions are puzzling and counterfactual - the four-stage sequence is of course meant to be a thought experiment (200). Even if decision-makers seeking normative guidance from Rawls' two principles were able to place themselves behind a partial veil of ignorance, what determines how high the veil is to be lifted at each stage? Rawls says that

> The flow of information is determined at each stage by what is required in order to apply these principles intelligently to the kind of question of justice at hand, while at the same time any knowledge that is likely to give rise to bias and distortion and to set men against one another is ruled out. (200)

We will agree that applying *any* principles of justice always requires absence of self-serving biases. It is not clear to me, however, how stipulating intelligence and the absence of such bias will help someone test the justice of contemplated and existing economic institutions and the justice of decisions made within them *by appealing to Rawls' difference principle*. According to this principle, inequalities are justified only as long as they are to the long-run advantage of the least advantaged (and do not interfere with fair equality of opportunity in qualifying for offices and positions).

But how is Rawls' test for the justice of any contested distribution in a given society supposed to work? We must ask if the thought-experiment of the four-stage sequence shows the basic structure of a given society to be in essential conformity with Rawls' difference principle. Now there would seem to be two possibilities: either a contested distribution is itself part of the basic structure, or it is not. If it is not, and merely flows from a basic structure that is itself just, then - according to

Rawls' theory - no charges of social injustice are legitimate, as long as the basic structure is just. But what if the contested distribution is itself part of the basic structure (which, given a sufficiently broad understanding of "basic structure", is probably the usual situation in the case of important distributions)? Now when it comes to deciding whether a contemplated policy conforms to the difference principle, who bears the difficult burden of proving that a contemplated or existing inequality among people will or won't bring about socioeconomic changes that can reasonably be expected to make the worst off group in society better off? And, to focus on an easily overlooked aspect of this question, what time frame should we contemplate? How long can benefits be delayed in the name of producing the greatest "long-run" advantage? Obviously, some policies will make some particular occupants of the worst-off social position even worse off in the short run but may promise to make future occupant of this position vastly better off in the end (e.g. by making clear that they can't rely on government handouts). By doing what will make some representative worst-off group in society better off at any given time, we may well have blocked opportunities for subsequent cohorts of the relatively worst-off to achieve equal or greater benefits they might otherwise have achieved.[8] And there may be all sorts of alternatives with different pay-off schedules, often requiring extremely difficult judgments of probability, which obviously leave tremendous scope for rational disagreement. As Mapel remarks,

> Even if the parties possessed a social science with perfect predictive powers....the problem of balancing the relative benefits and burdens of various reform strategies over time would remain."(120)

Our question is how anyone, at whatever level of decision-making within society's most important institutions, can get normative guidance from Rawls about

---

[8] In the present Canadian context, for example, recognizing the rights of Native peoples to control over large territorries, political sovereignty and large cash payments, we may well make one of the most disadvantaged groups in Canada better off by changing the face of the country in ways which the most disadvantaged among future generations of Canadians will live to regret.

how to reform specific social institutions so as to make the overall system of such institutions more just. Perhaps all that can be said in the end about this question, or all I can say, is that we must largely trust the impartiality and good sense of decision-makers to balance conflicting claims within particular contexts, within the parameters set by a democratically elected government and legally protected rights. But given the apparent ambitions of Rawls' theory, one might have hoped for clearer normative guidance than Rawls seems to offer us in the end, when he asks us to decide the justice of economic arrangements by going through the thought-experiment of the four-stage sequence.[9] The normative indeterminacy of Rawls' theory seems almost palpable when he says, in "The Basic Structure as Subject",

> ...the two principles of justice do not insist that the actual distribution reflect any observable pattern...What is enjoined is that (permissible) inequalities make a certain functional contribution over time to the expectations of the less favored...Institutions must organize social cooperation so that they encourage constructive efforts. We have a right to our natural abilities and a right to whatever we become entitled to by taking part in a fair social process. The two principles of justice define the relevant fair process and so whatever distributive shares result are fair. (1977: 164)

B. Primary goods

Rawls restricts his focus to the just distribution of so-called "primary goods", which are meant to be all-purpose means to pursuing any rational conception of the good, and again his rationale seems impeccable. His list of such goods contains six standard liberal political freedoms (including the right to own personal property, but not *necessarily* ownership of the means of production) and five social and economic

---

[9] In fact, Rawls himself sometimes seems to disregard the complexities of the four-stage sequence, for example when he says,
> "It is fairly straightfoward to ascertain what things will advance the interests of the least favored. This group can be identified by its index of primary goods, and *policy questions can be settled by asking how the relevant representative man suitably situated would choose*"(320; emphasis added).
Surely this passage cannot be construed as meaning that each specific policy question within society's basic institutions can simply be settled by asking what will favor the least advantaged? But how else can it be interpreted? I do not see how this passage could be reconciled with Rawls' insistence that society's basic structure, rather than particular distributions, are the subject of justice.

goods: wealth, income, power, authority and the bases of self-respect (cf. 62; 90-95).

Several troubling questions arise about this list of goods. Is it really possible to find a common index of aggregation for these different goods to determine when, in fact, the difference principle has been satisfied? Are they really all-purpose means to any rational life plan, or do they reduce a heterogeneous variety of goods verbally to a few simple categories? Do these primary goods cover *all* the means necessary for well-being? Let me only address the first of these worries here.

Although Rawls' acute awareness of balancing problems is unsurpassed in the literature on social justice, it is far from clear to me whether his proposed solution allows us to identify who the least advantaged in society are from whose perspective we are supposed to judge if the difference principle is satisfied. What we are supposed to do, if I understand Rawls correctly, is to compare individuals who hold varying bundles of social and economic primary goods and identify the representative least advantaged person. But social and economic primary goods encompass, as Mapel stresses,

> the powers and prerogatives of offices and positions of responsibility in all of the main economic and political institutions of the basic structure, income and wealth, and the social bases of self-respect. Leaving aside the numerous problems in identifying and measuring each of these goods, there remains the problem of assigning a relative weight to each part of the index. (Mapel, 105)

But if we are not clear about how to index primary goods - if *anyone* can attach to each (compenent of a) primary good whatever relative value seems to that person intuitively plausible - how are we going to reach any sort of social consensus about what will maximize the share of the worst off? Rawls' ingenious simplications reduce the commensuration problem (having to rank the relative value of all possible bundles of goods for all individuals) to a mere "index problem" (having to rank only the relative value of different bundles of goods possessed by plausible candidates for the social position of worst-off representative group). But the scenario of many different groups vying for the status of being especially disadvantaged is all

too familiar in our contemporary society. The two related problems here are that Rawls' index problem is solvable for *anyone* only in the crudest intuitive way and that, therefore, there will be nothing approaching a unique solution for *everyone* which could serve as the basis for a social consensus.

In fact, if even Rawls' parsimonious list of primary goods gives rise to troubling problem of intersubjective indeterminacy when we try to compare varying bundles of such goods to determine who is actually worst off, one hesitates to take Rawls to task for not including other important goods. In a later paper, Rawls himself suggests expanding his list of primary goods by including such things as "leisure time, and even certain mental states such as the absence of physical pain", provided we "respect the constraints of simplicity" and keep the goal of a political conception of justice in mind: "to serve as the focus of an overlapping consensus"(1988: 256-7). There are obvious pressures to expand the list of primary goods; but if Rawls thinks he must give in to them, I am puzzled how he can think that the index problem for primary goods is still sufficiently solvable to give us normative guidance.

The normative indeterminacies introduced by the highly subjective calculations required by the index problem, together with the indeterminacies associated with the thought-experiment of the four-stage sequence of deciding on the basic structure of society, would seem to make the normative content of Rawls' theory of justice highly interpretation-sensitive. Rawls leaves me at a loss as to how to generate convincingly determinate answers to contested questions of social redistribution. Kymlicka is even more blunt in assessing the normative payload of Rawls' theory:

> Other than a rather modest proposal to limit inheritances, Rawls gives us no idea how to confront the injustices in our society, or indeed whether he thinks there are substantial injustices to be confronted"(1990: 89).

§6 Concerns about Rawls' ability to enlist people's sense of justice

Even if Rawls' theory carries no determinate normative payload once we try to calculate our way through the complexities of the index problem and the four-stage sequence, can at least the spirit of it inspire and harness people's sense of justice? Our answer here can be brief. From the point of view of many thoughtful people's considered beliefs about justice, the very idea of the difference principle - arranging the economic system for the greatest benefit of the worst-off group - seems fundamentally misconceived. Even in Rawls' original position, behind the veil of ignorance, the contracting parties would hardly agree to principles which, at least in spirit, seem to require correcting certain inequalities which strike most people as clearly just (for example, someone's failure to acquire, through laziness, what the diligent person acquired through effort); and which seem to sanction inequalities that strike most people as clearly unjust (for example, a vast increase in inequality which would marginally improve the lot of the worst off in a society where the relatively worst off are already quite affluent). Even from the point of view of the original position it would not seem irrational to agree to a principle that does not permit the relatively well-off to enjoy a huge gain whenever the relatively less well-off, who may be quite affluent in absolute terms, suffer a marginal loss as a result. In fact, it has been a persistent criticism of Rawls' derivation of this two principles from the original position that he does not take the risk-taking side of human nature sufficiently into account.

We must conclude, I think, that Rawls' master principle, even if it were usefully determinate, would have dubious promise to enlist the sense of justice of enough people to serve as the basis for a stable social consensus about justice.

3.2.4 Limits of the master-principle approach

§1 What general lessons can we draw from the shortcomings of Rawls' work? Can other theories succeed where his seems to have failed? What can be

achieved in the field of social justice theorizing?

First, another caveat. I face a dilemma. On the one hand, I wish to claim that no other theory has succeeded *in the seemingly very plausible terms which Rawls set for a theory of justice:* finding a master principle of justice of society-wide scope which can give normative guidance by managing to avoid hopelessly complex balancing problems while at the same time having credible potential to harness people's sense of justice sufficiently to make it a basis for an informed social consensus. On the other hand, any such sweeping claim should clearly be backed up by solid evidence in the form of a critical study of every theory yet proposed. Although anyone familiar with the critical literature on contemporary justice theorizing will probably find it safe to say that no one yet succeeded where Rawls failed, I am clearly not justified in dismissing alternatives to Rawls' theory without discussing them. But one cannot fairly criticize a theory of justice without going into considerable detail, and my project here is to engage in synoptic reflections about the possibilities for rational disagreement about justice - not simply to write yet another book focussed on showing the inadequacy of existing theories. My caveat is, then, that the following tentative and somewhat disjointed speculations should only be accepted with that degree of confidence which they seem to warrant in the light of the reader's own background knowledge of contemporary theories of justice.

§2 What scope seems possible for a master principle of justice? Preferably we would like to have a distributive social algorithm which any rational person must accept for assessing the justice of all important distributions with some reasonable precision. Rawls was right, I think, that such an algorithm would have to be indirectly responsive, that is, focus on fundamental social background features which condition people's individual situation in ways that do not require individualized distributive assessments and corrective measures. Such an algorithm would also have to be either non-intuitionistic or at least extremely

parsimonious in the just-making criteria and primary goods it individuates, to avoid giving rise to unmanageable priority and commensuration problems. I can only see two workable suggestions of how, in principle, such a distributive algorithm could be constructed. Rawls' suggestion was, of course, to start with a picture of the distributive pattern which a system of just institutions should collectively aim at, institutions which then set up a procedural hierarchy within which distributive decisions at lower levels inherit just constraints from the levels above them. I cannot see a way to avoid the problems which effectively emptied Rawls' master principle of any useful normative payload. The other suggestion, that of libertarians, is to rely almost exclusively on markets to coordinate allocations to individuals. But it seems extremely hard to justify the libertarian principle in a way that has the potential to harness people's sense of justice sufficiently to prevent them from wanting to invoke the concept of justice to press additional claims based on desert, special needs, rights and other expectations. And there would seem to be little point in claiming that justice requires us to accept market outcomes to determine individual distributive shares (supplemented perhaps by a guaranteed subsistence income), and at the same time allowing all sorts of additional redistributions in the name of justice. This would simply open the floodgates yet again to the chaos of competing claims.

It is hard to see how *any* master principle of society-wide scope, whether focussed on individual distributive shares or society's basic structure, could escape what we might call *Rawls' dilemma*. Either a master principle embodies the necessary simplifications to keep the priority, aggregation-distribution and commensuration problems manageable; but then it would seem unable to capture enough of people's expectations and sentiment about justice to make this principle a promising basis for social consensus. Or a master principle captures enough of people's expectations and sentiment about justice to make it a promising basis for social consensus; but then it would seem to face unmanageable balancing problems

which make the principle normatively empty in practice. It is fair to say that libertarian theories typically get impaled on the first horn of Rawls' dilemma, egalitarian theories on the second. Rawls own theory, in an attempt to avoid either horn, seems to get impaled on both.

But if we cannot have master principles of justice of society-wide scope, what can we have? It is clear that commonsense percepts which people intuitively apply are always our fall-back option. People are not about to give up justice talk. The need to express, in some vocabulary, claims to a share of social resources which individuals perceive as their due is too deep a part of human nature. But surely there is room to move between a master principle of society-wide scope and commonsense percepts. These percepts can be given some degree of order which would at least mitigate "the chaos of conflicting claims" and which rational people would therefore have to accept, even if these ordering principles do not allow us to assess either the justice of individual shares on a society-wide basis or the justice of society's basic structure organized as a whole. But perhaps we shouldn't think of justice as having *either* society's basic structure *or* individual shares *on a society-wide basis* as its proper subject.

What exactly is the relationship between the basic social arrangements and what can *intelligibly* be called unjust? Although Rawls was undoubtedly right that many disputes about justice arise out of pervasive features of basic social arrangements, perhaps it is misleading to equate these background features with background "injustice". Although an indirectly responsive master principle aims to stand above our warring intuitions about justice and wants to arbitrate between them, it seems such a principle must fail to capture the complexity of causes which underlie people's sense of justice and instead gives rise to intractable, fruitless disputes among rival master principles. Perhaps our understanding of justice flows from context-dependent considerations in a way that makes us lose our grip on what we mean when we talk about the injustice of the basic structure of society as a

whole. Mapel's remarks seem helpful here:

> The alternative to dropping the idea of the basic structure, understood as a single, global unit of analysis, is not one of ignoring background injustice altogether. Rather, what we must give up is the idea of redressing the effects of background injustice all at once, from some comprehensive perspective. (124)

Master principles of justice geared toward society's basic structure run the danger of rejecting evolved and deeply entrenched features of society which are nourished by people's basic value orientations and are usually impossible to change all at once without destructive consequences. The historically well-substantiated risks of revolutions and radical reforms create a strong consequentialist presumption against them, but these risks do not, of course, rule them out completely under all circumstances. In some cases, as probably in the case of the communist Soviet Union, extremes of cultural and economic decay do justify a wholesale revamping of basic principles of social and economic organization. But usually there is a bitter price to pay for any such abrupt break with the past. Society's basic structures rarely get revamped to conform to someone's master principles of justice, and when they do, the result often turns out to be very different from what the reformers envisaged. To insist that we cannot sensibly hope to make headway on questions of social justice unless we first agree about which basic structure it would be just for society as a whole to adopt may only discourage piecemeal change while holding out a futile hope for magic consensus.

This is not to say, of course, that one is not entitled to have a particular vision of the best distributive pattern within society as a whole, but only that it is not clear how implementing any such vision could be justified, to every rational person, as a demand of justice. If a master principle is interpreted, in a vastly more modest sense, as *rationally permitted*, it will articulate only one of many possible visions of the good society, and will be useless for settling disputes about *justice* among defenders of competing visions.

Obviously, if we cannot even justify some basic moral framework as required by reason, the question of what justice requires becomes moot; but I have tried to argue that a form of consequentialism is indeed required by reason. A consequentialist moral framework does not seem to pick out one very definite vision of the good society as required for any rational person to accept, although I think it allows us, given minimally favorable circumstances, to confidently embrace liberal democracy. But given the epistemic uncertainties of consequentialism, we must be extremely careful about condemning whole economic systems, social structures, and civilizations on the grounds that they do not conform to someone's expectations and sentiment about "justice".

But beyond that, there exists, at least in Western pluralist democracies, an uneasy sort of evolved consensus on certain institutional fundamentals: representative government, a list of constitutionally guaranteed rights and freedoms, due process in legal matters, a market economy supplemented by an economic safety net, publicly funded basic education and publicly subsidized higher education, and a whole lot more. The basic structure of such societies can be already said to embody an open-ended, strongly pluralistic "master principle" in the way they regulate the distribution of a multiplicity of goods among their members. And yet, when it comes to the distribution of many economic goods in particular, social strife is rampant and in certain cases escalating, as members of various groups become more militant in demanding "justice" for their cause. Liberal societies respond to such matters largely in an *ad hoc* way, allowing decision-makers at various levels of responsibility to balance conflicting claims of diverse pressure groups in whichever way seems best designed to appease them. This situation is far from satisfactory.

§3   If we are not simply to resign ourselves to "the chaos of competing claims", at what level of generality and determinacy can we pitch principles of

justice? We want principles which are firmly rooted in people's context-bound sense of justice and at the same time clearly justifiable on the basis of consequentialist considerations. It seems clear that any justifiable master principle will have to be quite strongly pluralistic in both its criteria and its goods, and will therefore seem to face insuperable balancing problems if its domain of application is too heterogenous. In any case our principles will have to rely heavily on unguided intuitive balancing of context-dependent considerations in the light of whatever wisdom, impartiality and good will political decision-makers can muster. There seems to be only one basic way to reduce such balancing problems, and that is by thinking of principles of justice as applying to restricted domains. Although this will reduce the three balancing problems only by introducing a domain demarcation problem and a domain balancing problem, these problems may be more manageable, to the extent that there exists a fairly settled understanding of different domains of goods. An important type of goods are social roles, and part of promoting a settled understanding of different distributive domains will be to promote a stable system of reasonably clear role differentiation among members of a society and to attach expectations to each role which can be clearly justified in terms of its function. Walzer, for example, individuates eleven distributive domains or "spheres of justice", each with its own set of roles and goods which are constitutive of that domain, and each with its own principles of justice which must be applied by intelligent, responsible, caring persons who are both familiar with the "social meanings" of the goods at issue and occupy an impartial viewpoint.[10] Walzer's eleven domains are community membership; security and welfare; money and commodities; office; work; leisure; education; love and kinship; religion; and recognition. Walzer's approach, which relies on settled social meanings, faces many obvious limitations and needs to be supplemented with a consequentialist moral

---

[10] Cf. Michael Walzer, *Spheres of Justice* (1983). Walzer also has an interesting principle of the autonomy of different domains which raises many problems. I cannot pursue them here.

framework if it is not to lapse into social relativism. But it correctly describes, I think, how societies must try to channel individuals' justice-based claims into manageable patterns of expectations to which a satisfactory political response is possible.

3.3 DISAGREEMENT ABOUT THE JUSTIFICATION OF PRINCIPLES

This chapter explores the structure of justification for principles of justice within a consequentialist framework.

3.3.1 The basic nature of consequentialist justification

§1 Given the many competing principles which have been proposed by different theories of justice, what sorts of considerations can *ever* make compliance with any such principle morally required? What exactly is meant by "justifying" a principle of justice within a consequentialist framework, and how much room do consequentialist justifications leave for rational disagreement? When is failure to satisfy someone's justice-based claims an injustice, and are all injustices morally wrong? Answers to these questions are urgent to inject clarification into many confused debates about justice.

I will again focus on concerns at the metaethical level, on conceptual structure rather than on judging the justice of specific distributions. To recapitulate briefly, I tried to establish a presumption that a wide reflective equilibrium is likely to support a constructivist, non-foundational, distribution-sensitive consequentialism, according to which morality's main job (promoting enough uniformity and predictability in people's behavior to create the necessary sense of security for a climate of mutual trust and cooperation) is best accomplished by thinking of moral constraints as aiming at an interpersonal goal which combines both aggregative and distributive values. A broadly consequentialist approach to morality requires us to evaluate a system of behavioral constraints by whether its components are individually necessary and jointly sufficient for plausibly promoting the most desirable combination of satisfied interests of individuals in their own well-being. In determining which combination of satisfied interests of individuals is most *desirable,* each of us is to be in principle guided by the idea of giving to every individual's interest in his or her well-being the same consideration

which each of us would give it if each of us had this individual's life to live and judged from the perspective of that individual's values. In judging a person's interests from that person's perspective, however, the pursuit of purely subjective interests is to be constrained within the bounds of what can plausibly count as "well-being" (see Sec 3.4.2). Although many proposed forms of consequentialism are rightly rejected as fundamentally irreconcilable with our considered judgments about the wrongness of routinely sacrificing the vital interests of the few for the marginally greater happiness of the many, this particular form of consequentialism can arguably achieve a self-consistent synthesis of fundamental distributive and aggregative concerns, although necessarily at the expense of a massive aggregation-distribution problem at the level of fundamental theory. But the ideal consequentialist goal is not to be understood as a deductive moral foundation, justified independently of our considered judgments in particular contexts and bestowing justification on these judgments uni-directionally. It is better understood as a central node in the web of contextual principles selected in a wide reflective equilibrium between considered judgments and background theories. The gap between normative indeterminacy at the abstract level and useful determinacy at the practical level can be bridged only by thinking of moral justification as seeking maximum consistency among a hierarchy of principles at different levels of generality, each of which is a plausible approximation to the ideal consequentialst goal in its particular domain.

§2 But can even *maximum* consistency be sufficient? Or must moral justification be cashed out in terms of some stronger relation between principles, such as strict deducibility, or perhaps derivability or coherence in some less intimate sense, or perhaps only some loose degree of mutual support? It is not easy to characterize the relations that hold between moral principles of different content, generality and determinacy. Moral principles do not arrange themselves as a

deductive hierarchy whose implications somehow reach, like tentacles, into every corner of our lives. The model that doesn't fit morality at all well, yet stubbornly persists in some minds, is that of an interpreted axiomatic system which requires us to know only logic and a small number of basic moral formulas to derive, in a more or less mechanical way, all required moral theorems. These theorems would then simply require us to have only factual information to generate the relevant moral constraints in each real-life situation.

I will not analyze the ways here in which any minimally sensible account of moral reasoning must conflict with this sort of model. For one, the model overlooks a major source of rational disagreement we have not considered: the problem of description-relativity. Obviously, we can only make principled moral judgments if a state of affairs is understood first to be of a type to which a particular principle applies. Whether a particular principle is applicable depends on how a state of affairs is described, but there are often various ways, at varying levels of generality and determinacy, to describe the same state of affairs. Any principled approach to moral issues runs, therefore, into the fundamental problem that the same principle will apply to a state of affairs under one description but not apply to it under another. Applying moral principles to real-life situations - connecting an abstraction in the mind up with our experience of the world - always requires a creative leap of the imagination, tremendous selectivity in one's focus on relevant feature of experience, and a judgment in reflective equilibrium that will often leave room for rational disagreement.[1]

While I have not worked out any alternative model for the "ecology" of our moral constructs, something like the following seems to me worth stressing. Our collective understanding of moral concepts evolved from a largely subconscious inductive process which made human beings coordinate their behaviors in certain

---

[1] A helpful discussion of the problem of description-relativity applied to political philosophy can be found in Shapiro, *The Evolution of Rights in Liberal Theory* (1986), 290f.

ways to satisfy desires they naturally had. This coordination can be greatly helped by the ability of human beings to group situations and behaviors into types with certain common features, and to communicate with one another about such matters. Moral principles can thus be expected to contain, as their chief constituents, commonsense characterizations of types of behaviors along with commonsense descriptions of types of situations in which these behaviors should be displayed.

What relationships is it reasonable to expect between different principles of this sort? There are many different ways to carve up the experienced world conceptually, depending partly on one's interests and purposes. Since human interests and purposes vary, the resulting moral concepts will not neatly carve up the world into mutually exclusive and jointly exhaustive domains of application, but into a patchwork of overlapping domains and subdomains. Moral concepts or principles can only be marked off from one another – for example, justice from beneficence or gratitude – to the extent that the functional roles they play are distinguishable. And the functional roles of different moral concepts, as we saw in the case of rights and justice, often overlap greatly with respect to the sort of behaviors they actually require of people. The functional role of the concept of justice in a consequentialist moral vocabulary, we said, is twofold: to give voice to the claims to a share of social resources which individuals or groups make in pursuit of their well-being, backed by their sense of justice; and to articulate a political response to the pattern of such claims within some domain. Why such a sense of justice might have evolved in human beings is an interesting sociobiological question. At least in general outline, it is not difficult to imagine how survival pressures might select in favor of types of community-forming animals whose interest in their own well-being causes them to make demands on communal resources, but who are at the same time willing to recognize other community members' analogous interests and to constrain the pursuit of their own demands accordingly.

As long as behavior-coordinating principles, arising from various different purposes we have, do not contradict by prescribing two or more behaviors simultaneously that cannot be carried out together, we are justified by our purposes in doing as each prescribes. If "justifying" a moral principle means exhibiting its share in carrying out morality's job of channelling human behaviors in ways that promote well-being , then rather than thinking of moral principles in terms of a more or less well-ordered hierarchy, we should think of them in terms of each principle regulating sub-autonomous domains of goals and purposes. If morality is to accomplish its job, principles need to be more or less tightly coordinated so as not to lead to frequent conflicts between prescribed behaviors. And that means that lower-level principles must be at least consistent, or loosely cohering, with higher-level principles, not that they must be strictly implied by them. If some such account of the underlying nature of moral principles is correct, then it is hard to see what else it could mean to "justify" a principle of justice than to make plausible that it *both* responds appropriately (in a sense to be clarified) to the pattern of expectations and sentiment about justice in some domain *and* also happens to fit in reasonably well with an overall system of principles whose wellbeing-promoting tendencies are already inductively well-established.

§3 But what is this system of principles which has already met the consequentialist well-being test? What is the fundamental content of a consequentialist system of personal and political morality? My aim here is not to develop such a morality in great domain-specific detail, but simply to sketch in outline how it can be launched. I shall assume that we are designing a morality for a complex, pluralist society and move quickly through some very complicated issues (some of which will be taken up in the next chapter).

What system of principles, standards, norms or constraints on the behavior of individuals, groups, and governments can by justified in a wide reflective

equilibrium, given our understanding of such a system's main job, of the ideal consequentialist goal, of the nature of typical human interests and other features of human psychology, of prevailing moral traditions, of the causal structure of the world, and other relevant background knowledge? What system of principles will have the most desirable consequences?

In justifying a particular principle as having more desirable consequences than its alternatives, we must ask not merely how desirable its total consequences would be *if* everyone complied with this principle. Assuming perfect compliance may, at best, be a very rough initial approximation to eliminate obviously unacceptable principles from the rest. Consequentialists, for whom only the consequences a principle will produce in the real world matter, must also ask three further questions: (1) Will the proposed principle still be the most desirable alternative, given the likely extent to which people will, for many complex reasons, fall far short of perfect compliance? (2) Will the proposed principle still be the most desirable alternative, given some often overlooked costs associated with its effective propagation, such as disrupting people's established habits, creating anxieties, and requiring teaching and enforcement efforts? (3) Will the proposed principle still be the most desirable alternative, given that it should lead to as few conflicts as possible with any of the other principles which are to make up society's total system of moral norms?[2]

So strictly speaking, we cannot justify a principle of justice simply by deciding, in reflective equilibrium, whether perfect compliance with it would plausibly promote well-being better than its alternatives. The project of consequentialist constructivism requires us to ask a far more complicated question: "Will the total system of social norms of which this principle is to become a part produce the most

---

[2] Cf. Richard Brandt, *A Theory of the Right and the Good* (1979), for the distinction between ideal criteria of right and wrong and a practical moral decision procedure, in which those criteria are watered down so as not to put the morally conscientious person at too unfair a disadvantage in a society in which these criteria are not generally complied with.

desirable consequences, given the extent to which people will probably fail to comply perfectly with this system, and given also that there are costs of maintaining this system?" Let me call this *the ultimate conquentialist test*, as opposed to the ideal consequentialist goal.

Although the information and computations required to assess the consequences of a proposed principle and all its alternatives along all these dimensions would seem to be staggering at first, in practice the approximate range of principles which are plausible candidates for passing the ultimate consequentialist test is not at all hopeless. We know, for example, that even otherwise ideal principles won't do us any good if enough people cannot be motivated to comply with them at a reasonable enforcement cost. To achieve sufficient compliance with moral principles requires most people's active participation in maintaining compliance pressure on people around them. To achieve such compliance pressure, given that human beings are largely creatures of habit, will require us not to tax their reflective capacity concerning fundamental moral matters too heavily, but to respect the inertia of people's childhood conditioning as one of the facts of commonsense moral psychology. So any acceptable system of moral norms will have at least three crucial pragmatic characteristics. It will contain only principles simple enough so that most people inhabiting the domain of application of these principles will understand them. And it will contain only principles that do not generally require people to make the sort of self-sacrifices which would conflict with their (reflectively rational) self-interest to such an extent as to motivate widespread non-compliance.[3] And it will contain a strong presumption in favor of requiring

---

[3] A rational morality must ultimately be reconcilable with self-interest, not in the implausible Socratic sense that doing the morally wrong thing could never make anyone better off, but in the sense that, first, the system of moral norms must plausibly aim at promoting the most desirable combination of satisfied interests of individuals in their own well-being, and second, the system of moral norms must be such that, with a not too costly system of rewards and punishments, ultimately most people, most of the time, regard it as their *best bet* to comply with these rules and cultivate, out of enlightened self-interest, the disposition to be morally responsible. Of course, no moral system can ensure that doing the morally wrong thing will not make someone better off, given the *status quo* as the baseline for comparison. All that is required for an effective morality is an appropriate system of rewards and

moral innovations to be made piecemeally and to be clearly defensible as consequentialist improvements, rather than making a clean sweep of existing practices and beginning *de novo*.

Given these pragmatic conditions on acceptable moral principles, plausible candidates for the core of a consequentialist morality are not hard to identify. We can make a fairly confident judgment that the system of moral principles should include the following: "Tell the truth", "Don't steal", "Keep your promises", "Obey the law", "Don't engage in unprovoked physical attacks", "Take care of your children" or "Don't discriminate against people on the basis of race or sex". Let us follow Haslett in calling these the "primary norms of personal morality"(cf. 88ff). One conspicuously absent norm, which clearly fails the ultimate consequentialist test for primary norms, is the act-consequentialist principle, "Always do, on every occasion, what you think will have the most desirable consequences." This principle fails the test, and is therefore not among the primary norms of a rationally justifiable personal moral code, because it is far too epistemically demanding and would lead to the sorts of disagreements, failure of coordination, and rationalizations for freeloaders which could not accomplish morality's main job.[4]

Primary norms of political morality will specify the central task of government, namely to solve coordination and free rider problems by enforcing a system of legal sanctions. Primary norms of political morality will further consist of norms specifying acceptable conduct for government officials and restricting governments and individuals in what they can permissibly do to other individuals or groups. In the contemporary context, many of these norms will be stated in the language of rights. Nothing about consequentialism *necessitates* rights-talk (except

---

punishments which makes it a good bet for most individuals to comply with the important moral rules even in the face of most temptations.

[4]Harsanyi, among others, has given an excellent account (which I shall not try to repeat here) of the advantages of rule over act consequentialism in terms of coordination, expectation and incentive effects (cf. 56-60).

perhaps a right to life and certain basic freedoms). The point of having a morality is clear; but whatever conceptual division of labor we evolve to carry out morality's basic job - whether we use talk about rights, goals, virtues, or other moral concepts - is relatively immaterial and partly a matter of historical contingencies.

Norms of acceptable governmental conduct, in addition to rights, would very likely have to include the requirement that governments must themselves obey the law, keep their agreements, give aid to other countries, or take care to some extent of the interests of animals and future generations. About the details of these norms there can obviously be considerable rational disagreement, but hardly about the question as to whether we should have norms governing these matters at all. And the consequentialist rationale for enforcing norms about a certain matter is, of course, that allowing individuals to act *against* such norms in order to promote what seems to a given individual the greatest overall well-being on a particular occasion is, given our epistemic limitations, our propensity for self-interested rationalizations and other people's expectations of us, almost always a bad consequentialist gamble.[5]

Controversy will frequently erupt about the area of governmental action which is left over after governments have observed all constitutionally or otherwise predefined norms - an area which is necessarily governed by some version of the act-consequentialist norm after all. Within limits, usually spelled out in a country's constitution, consequentialists want governments to promote the greatest well-being according to their best judgment in the context of current opportunities. In fact, precisely because there can be, within limits, many conflicting interpretations of what the specifics of this goal amount to, governments are necessary as arbiters in any society. Because a government-of-the-day's power of discretion is so great and mistakes are potentially fateful, however, another plausible norm of consequentialist political morality is that the individuals forming each particular

---

[5]See Haslett (135) for detailed arguments to this effect.

government should be subject to a system of checks and balances which periodically includes free elections.

In addition to primary norms, a complete system of personal and political morality must, of course, also include secondary norms which tell us what to do in four special types of cases: (1) *emergency cases* (and *golden opportunity* cases) where it seems virtually certain that we could prevent immense evils (or achieve tremendous good) by ignoring ordinary norms; (2) *conflict cases* or *borderline cases* where two or more norms seem to sanction incompatible courses of action; (3) *oddball cases* which are so infrequent that it is not worth complicating our system of explicit norms for their sake; and (4) *cases without precedent*, which are types of cases which we had no occasion to contemplate in the past and for which our existing norm system is not prepared. Although I cannot pursue this matter here, it is possible, I believe, to come up with plausible consequentialist norms to resolve these types of 'hard' cases in a principled way most of the time (cf. Haslett, 100ff; also Dworkin, 1985). I now turn to the question of how principles of justice are embedded within a general system of consequentialist moral principles.

3.3.2 *Bona fide* principles of justice and morally justified principles

§1 We must distinguish two types of principles of justice, according to the stage they have reached in the process of consequentialist justification: *bona fide* principles and morally justified principles. *Bona fide* principles of justice constitute an "appropriate" response to conflicting justice-based claims in a given domain by striking a reasonable compromise between them. Morally justified principles of justice are principles that require a morally responsible person's compliance. A *bona fide* principle of justice is not generally the same as a morally justified principle, because there can be many competing *bona fide* principles in a given domain, and only after all *bona fide* principles have been subjected to the trade-off considerations necessary to determine the morally best can one of them be regarded as morally

justified.

A *bona fide* principle of justice is a proposal for achieving a reasonable compromise among different individuals' justice-based claims to social resources within some sufficiently well-bounded distributive domain. *Bona fide* principles can be more or less comprehensive. The most comprehensive principle conceivable would be a master principle of global scope which encompasses all distributive domains in existence and would balance conflicting justice-based claims to global resources both within and between its many subdomains to any desired degree of precision. Given the intractable problems arising with a principle's increasing level of comprehensiveness, however, no such "ultimate master principle" seems feasible. Most *bona fide* principles of justice will simply be proposals for achieving a compromise among conflicting justice-based claims in a restricted domain.

I do not think there can be anything very exact or systematic about individuating distributive domains. Domains are simply those contexts of cooperative human activity which shape the identity of individuals inhabiting them. Many large, inhomogeneous domains have large numbers of important subdomains; for example, the domain of paid employment. Domains are demarcated primarily according to evolved understandings about the nature of different goods, institutions and practices, the functional differentiation between social roles, ethnic groupings, or geographic boundaries. The social understanding of a domain's boundaries, even though not of just distributions within it, must be reasonably well settled for domain-specific principles of justice to become viable. Although the consequentialist outlook is fundamentally cosmopolitan, the choice of a politically organized "society" or "nation state" as marking off an especially salient domain is justifiable for many purposes.[6]

---

[6] Peter Danielson suggests (personal comm.) that the question of international redistribution is an especially important test for consequentialism. Contractariansm, it would seem, can keep the obligations of affluent nations toward the Third World within reasonable bounds, whereas consequentialism would require a tremendous transfer of wealth from rich to poor countries. This may seem to some a *reductio ad absurdum* of consequentialism. I cannot digress into this complex issue here.

The major balancing problems afflicting principles of justice can be alleviated by dividing the problem of justice into problems *within* different domains, even if their exact nature and boundaries are often themselves in dispute. Where social understandings are unsettled, what matters is that individuals have opportunities to voice their dissent and change prevailing conditions within a domain or, if necessary, to exit one particular domain for another that suits them better (cf. Skhlar, 166f). Different domains may compete for a common stock of scarce resources, but admitting this obvious fact does not undermine the claim that we can only make the problem of justice manageable by dividing it into problems within separate domains. For example, problems of justice arising in the domain of allocating scholarship funds among graduate students can generally be settled by clearly articulated domain-specific procedural and material principles which structure the expectations of individuals inhabiting this domain. Problems of justice arising in the wider domain of deciding a university's budget can also be settled, usually after lively debate, by a domain-specific combination of principles. And problems of justice arising in the even wider domain of deciding a government's spending priorities can in turn be settled by principles appropriate to that domain, such as, for example, some established variant of a fundamentally democratic process constrained by constitutionally protected consequentialist liberty and welfare rights. But the process of creating constraints which are acceptable to people's sense of justice works both top-down and bottom-up. Subdomains must adjust their principles as a result of pressures from wider domains, wider domains must adjust

---

But it is not obvious to me that consequentialists in affluent countries must rush in and help the starving with all the resources they have. Arguing about principles of global redistribution within a consequentialist framework will require (1) a strong probability that a proposed set of measures will indeed restructure the complex causal web of the world's environmental, demographic and economic problems in an desirable way; and (2) a factoring into this probability a realistic estimate of the motivational barriers resulting from substantial sacrifices required of individuals in the name of moral idealism. Proposals to expand the scope of theories of social justice beyond nation-states - such as, for example, Pogge's proposal to apply Rawls' maximin principle globally - are likely to impose such strains on most people's moral commitments in affluent parts of the world as to predictably fail to harness their sense of justice in support of such proposals, and thus also to fail the ultimate consequentialist test.

their principles as a result of pressures from subdomains. At every stage in this first-order contextual process it is important to be sensitive to sources of limited rational disagreement which require pragmatic compromise-seeking informed and constrained by an awareness of the sources and limits of first-order disagreement.

§2 We defined a *bona fide* principle of justice as a proposal for achieving a reasonable compromise among those conflicting claims to social resources made by different individuals in pursuit of their personal well-being which are backed by their sense of justice. Let me now explore the most reasonable spelling-out of the idea of a "reasonable compromise" within a consequentialist moral framework. A proposal for achieving a compromise among conflicting claims to social resources in a given domain is reasonable if and only if (1) it can be implemented, or credibly approximated, by some feasible distributive mechanism (which implies among other things that it does not allocate resources to conflicting uses); and (2) it is *responsive* to existing patterns of expectations and sentiment about justice within the relevant domain. A third condition must be added, for obvious consequentialist reasons: (3) The existing pattern of expectations and sentiment about justice are the result of adequate information available to people and *not* the result of effective brainwashing techniques which have made segments of the population willingly embrace slavery or its equivalent.

These seem to me plausible consequentialist conditions for making something a *bona fide* principle of justice, though they contain elements of vagueness, and judging when they are met will often leave *some* room for rational disagreement. How, most importantly, are we going to explicate the exceedingly vague responsiveness condition? It is clear that a *bona fide* principle of justice must be, to some hard-to-define degree, responsive to existing patterns of expectation and sentiment about justice in order to be recognizable as a principle of *justice* at all.[7] At

---

[7] The charge is often levelled against Gauthier's "mutual advantage" contractarianism that it is not a

one extreme, a principle would be strongly responsive to existing patterns of expectations and sentiment about justice if it could somehow satisfy every expectation and conform to everyone's sentiments. At the other extreme, a principle would be weakly responsive to existing patterns of expectations and sentiment about justice if the balance it proposes among conflicting justice-based claims has some minimally credible potential that enough people's sense of justice could realign itself with this balance eventually to produce tolerably favorable conditions for social cooperation. What is the most reasonable compromise between these alternatives?

A CND consequentialist (unlike certain contractarians) will insist that the idea of a "reasonable compromise" be spelled out so as to embody a plausible conception of impartiality. We might say that a balance between the demands of different just-making criteria is adequately responsive to existing patterns of expectations and sentiment about justice if it corresponds to the decision that could plausibly be reached by a tribunal whose members have an in-depth understanding of the domain at issue, have personally nothing to gain or to lose from their decision, and are guided by a desire to satisfy different individuals' competing justice-based claims to the greatest extent compatible with respect for the separateness of persons.[8] This idea is lucidly expressed by Brian Barry:

> What, after all, is the virtue of impartiality, as we seek it in, paradigmatically, judges and arbitrators other other public officials? An impartial decisionmaker is one who acts 'without fear or favor'. This includes, of course, not taking bribes, anticipating rewards, or succumbing to threats. But it also means the absence of more subtle departures from objectivity; impartiality excludes favoritism based on friendship, similarity of race or class, and so on...The reasons for deciding should be general, publicly statable,

---

theory of justice at all, but merely a description of *realpolitik*. See Kymlicka, 1990, p. 131, who cites Rawls, Barry and Sumner as making this point. This charge is not, of course, a refutation of Gauthier who claims that his theory is the best we can have by way of rational agreement on principles of social cooperation.

[8] I address below the complicating factors for any such criterion of "adequate responsiveness" arising from the fact that different representative committees might well reach different distributional decisions.

and publicly defensible.... (1989: 290)

This idea of procedural impartiality can be readily (though usually far from perfectly) operationalized in a multitude of interesting ways which I cannot discuss here. But what makes something a *bona fide* principle of justice, rather than merely an expression of personal opinion or sentiment, seems now fairly clear. It must be a proposal for achieving a reasonable compromise, relative to some sufficiently well-bounded domain, among different individuals' conflicting justice-based claims to social resources, where "reasonable compromise" means that (1) the proposal can be satisfactorily approximated by feasible distributive mechanisms; (2) it corresponds to a decision that could plausibly be reached, after due deliberation, by a group of (intelligent) decision-makers of good will who are guided by a thorough understanding of the nature of justice-based claims in the domain at issue (and of neighboring domains), and guided also by the desire to satisfy these claims to the greatest possible extent compatible with respect for the separateness of persons; and (3) the pattern of people's expectations and sentiment has been shaped by adequate information and not by effective brainwashing techniques which have made segments of the population willingly embrace slavery or its equivalent. As is evident from this definition, there is no avoiding the aggregation-distribution problem at any level of moral decision-making, no calculus that would make qualitative judgments unnecessary, although the difficulty of making these judgments to the satisfaction of all or most individuals affected will become quite manageable for experienced decision-makers in suitably restricted domains.

3.3.3 Defining injustice

§1 If something like this is indeed the most plausible set of conditions for *bona fide* principles of justice, then there will, almost inevitably, be a range of such principles relative to most viable domains. So there can be rational (first-order) disagreement about which principle should be implemented, given that only one

can. Now given that justice permits, but does not require, the implementation of any particular *bone fide* principle from within this range, what does it mean for something to be an "injustice"? For something to be justifiably called an injustice, it is clearly not sufficient that some justice-based claim has to be disregarded or that someone's sense of injustice has been aroused. Something that we must demand from *any* minimally plausible account of justice is that it spell out sufficiently restrictive conditions for calling something an "injustice" so as not to debase the normative currency of appeals to justice. At the same time, the conditions must not be so restrictive that non-justice considerations routinely overrule any reasonable compromise between conflicting justice-based claims. Here consequentialists face a difficult conceptual choice.

Should an injustice be diagnosed whenever a distribution violates some *bona fide* principle or other within the permissible range of such principles in a given domain? Or only if it violates *every* principle within this only approximately delimitable range? Or only if, in addition to violating every relevant *bona fide* principle, it meets further conditions, such as passing the ultimate consequentialist test which tries to harmonize justice-based claims within a domain with all-things-considered morality?

I shall label these three options for the use of the concept of injustice "the least restrictive", "the moderately restrictive" and "the most restrictive" account of *injustice*. Which should a consequentialist accept? There would seem to be strong pros and cons on all sides of this conceptual issue. It would seem plausible to adopt the least restrictive account (allowing the violation of whatever *bona fide* principle some individual might have preferred to be called an injustice - at least to that individual), because all rival principles within this permissible range are equally *bona fide* principles of justice, and choosing one does violate all the others, and it may make all the difference to an individual which principle is actually acted on. But it would seem equally plausible to adopt the moderately restrictive account

(allowing only distributions that violate every *bona fide* principle within a given domain to be called injustices), because a particular distribution can only satisfy one permissible rival principle, and it is clearly a matter of rational disagreement which principle within this range is preferable, and we do not want to debase the normative currency of injustice-claims. And, finally, it would also seem plausible to adopt the most restrictive account (allowing only distributions which violate a *bona fide* principle that has met the test of all-things-considered morality to be called injustices), because in the popular mind the term "injustice" strongly connotes a moral wrong, and violating a *bona fide* principle is morally wrong only if this principle does not fail the ultimate consequentialist test, and this account would seem least likely to debase the normative currency of injustice-claims.

How do we want the concept of injustice to function in our moral vocabulary? Which use of the term "injustice" should consequentialists legitimize and therefore encourage people to adopt? There is no ideal solution. On balance I incline to the moderately restrictive account, despite some very serious drawbacks. The moderately restrictive account of injustice makes it *both* less likely than the least restrictive account that the unjust and the morally wrong will be very far apart, *and* less likely than the most restrictive account that legitimate uses of the concept of injustice will be radically divorced from what individuals claim as their due in pursuit of their well-being. This account implies that, in order to make an injustice-claim legitimate, *more* is required than simply the violation of some plausible *bona fide* principle or other in some domain, but *less* is required than being able to defend this injustice-claims against the demands of all-things-considered morality. On the one hand, this account requires someone who is displeased with a distributive decision to look at a whole domain of application and ask whether the violation of his preferred *bona fide* principle was not perhaps required to avoid violating another such principle. It requires an individual who feels unjustly treated to consider whether perhaps a distributive decision did not go his way because others

had as good a claim as he did to having the decision go *their* way. This may put unrealistic demands on some individuals, but making such demands, trying to habituate individuals to judging justice from a wider perspective, seems eminently desirable if appeals to justice are to have a harmonizing rather than divisive effect.

On the other hand, the moderately restrictive account of injustice does not divorce uses of this concept so radically from the particular contexts of individuals' justice-base claims that it would lose its connection with them, would have no potential to harness most people's sense of justice, would require them to wade into the complexities of all-things-considered morality and would get us into all the difficulties of ambitious master principles of justice of society-wide scope.

§2   Given that Rawls' project of tying the concept of injustice fundamentally to an ideal of society's basic structure has not proved feasible, I see no alternative for a consequentialist account of justice than to follow Walzer in tying the necessary and sufficient conditions for justifiably claiming an injustice fundamentally, although not indefeasibly, to those spheres of evolved social understandings in whose context the identity of individuals is constructed and embedded. Justice-based claims are fundamentally projections of such embedded selves. The pattern of people's expectations and sentiments is structured by their understanding of what is proper within particular domains of goods and roles. It is true that individuating domains poses problems, including problems that can themselves be intelligibly conceptualized as problems of justice, and that different domains often draw on a common pool of scarce resources and interact with each other in complex ways. But following Rawls in going beyond individual domains in our diagnosis of social injustice and trying to assess the totality of distributive domains all at once would only make us lose our grip on questions of *who* can reasonably demand *what* in the name of justice, or of what we mean by "justice" at all.

But isn't my balancing view of justice simply a form of social relativism,

taking as just whatever evolved social meanings and patterns of expectations and sentiment prevail in a particular social context? There is admittedly a strong element of conservatism in the account outlined above, and it has this much in common with relativism that the specific content of morally justified principles of justice may vary from society to society, and from social context to social context, if patterns of expectations and sentiment vary. But *all* accounts of justice will have this feature to some extent, unless they propose some ideal of what justice *should* mean, no matter how strongly this conflicts with people's evolved understanding of what is their due in pursuit of their well-being as individuals. And if the acceptance of any particular alternative among many such ideals cannot be shown to be required by reason, proposing such an ideal as an account of *justice* represents mere dogmatism.

But my account differs from both unfettered social relativism and inflexible conservatism in two main ways. First, although I consider *bona fide* principles of justice to be indeed a complex function of existing patterns of expectations and sentiment about justice, the three conditions for what is to count as a "reasonable compromise" ensure that such principles do not necessarily reflect the expectations and sentiment embodied in existing laws or held by some politically dominant group. Second, even though *bona fide* principles of justice may embody a high degree of cultural and domain-dependent relativity, they are morally justified only if they cannot be shown to conflict with the ultimate consequentialist test.

Injustices can thus be morally justified at times. People must be educated to recognize more clearly why having determined an injustice relative to some more or less well-defined domain is not necessarily the end of all moral debate. It only creates a very strong presumption that a moral wrong has been committed and creates a heavy burden of moral justification. If we are using the term "injustice" in the way that seems to be advisable from a consequentialist perspective, then some things that are quite properly termed injustices can only be rectified by sacrificing

values of equal or greater moral importance or by committing similar or even greater injustices in other domains.

But admittedly, this account makes *legitimate* charges of injustice conceptually fuzzier than many people's facile and confident use of this term reflects. What makes a distribution unjust, on this account, is that it violates all principles which propose a reasonable compromise between justice-based claims in some viable domain. But what seems like an injustice if viewed within some narrowly restricted domain may, of course, no longer be so clearly an injustice if viewed from a more expansive domain. Moreover, overall moral justification has to look to more than one domain; it has to look to the best consequences for distributing resources among all distributive domains. Therefore, on the moderately restrictive (domain-relative) account of injustice, morally justified injustices may arguably occur in just those cases when every single one of the *bona fide* principles of justice can be be plausibly argued to fail the ultimate consequentialist test.

3.3.4 Rational disagreement about consequentialist justifications

§1 But how do we apply this test in practice? Ideally, of course, we would bring our entire system of already existing personal and political norms to bear on a proposed principle of justice and reach a decision in reflective equilibrium about where any necessary adjustments should be made in case of conflicts. But for many purposes we clearly need a more manageable decision-making process. Are there perhaps plausible approximations to our ideal consequentialist goal which we could use as criteria to decide which particular *bona fide* principle of justice among all those principles striking a reasonable compromise among conflicting justice-based claims in a given domain should actually be followed?

I can think of no better criteria for choosing between conflicting *bona fide* principles than the idea of using certain traditional moral standards *in combination*. These standards accord with the considered judgments of many thoughtful people

and direct our attention to plausible aspects of the ideal consequentialist goal, such as the promotion of happiness, moral consistency, respect for persons, or voluntary agreement. The following table states what seem to me plausible interpretations of such moral standards, applied to the problem at hand:

The greatest happiness test	"Will including this principle as a principle of justice in our system of moral norms foreseeably create more unhappiness than some viable alternative?"
The moral consistency test	"Will recognizing this principle as a principle of justice foreseeably set an example which could not be adopted in other domains without disastrous consequences?
The respect-for-persons test	"Will recognizing this principle as a principle of justice foreseeably (and avoidably) treat some people merely as a means to other people's ends?"
The voluntary agreement test	"Could informed people of good will voluntarily agree to recognize this principle as a principle of justice for our mutual benefit (at least if we judge from a reasonably impartial standpoint)?"

Table III: Consequentialist formulations of traditional moral standards used as criteria for proposed principles of justice

These standards cannot, of course, serve in the fashion of mechanical decision-making devices, but only point in the general direction for looking beyond a restricted domain in our attempt to pick the most defensible option among conflicting *bona fide* principles.

Giving confident answers to the tests of our basic moral standards will often be difficult. They may therefore leave ample, though not unlimited, scope for rational disagreement. But many (in some cases perhaps all) *bona fide* principles of justice will clearly fail to meet the test of these standards. In this way, these standards *constrain* what can permissibly be done in the name of justice. Insofar as disagreement about *what can justifiably be done in the name of justice* arises from one party's unwillingness to accept these constraints, I do not consider it to be a form of *rational* disagreement.

Of course, there may often be several *bona fide* principles of justice relative to

a given distributive domain which pass the test of these moral standards. So there will, at this final stage in the process of (first-order) moral justification, often be considerable room for rational disagreement about which principles of justice a morally responsible person is obliged to comply with. This result is no surprise. Several rival principles of justice will often be equally morally justified *at the first-order level* by any plausible consequentialist account of moral justification (and the particular principle which actually requires compliance can only be singled out by a second-order legitimation process).

What creates so much room for rational disagreement about principles of justice? These principles must be morally justified by applying the test of our basic moral standards to various alternative ways of achieving a reasonable compromise between the conflicting justice-based claims of different individuals. But our basic moral standards often leave a wide range of strongly conflicting ways of achieving such a compromise equally permissible, not only because of our epistemic limitations, but also because different types of social distributions will realize different but equally plausible forms of well-being. The best we can reasonably expect, in most important distributive domains, is that our rational standards, together with our conditions for what counts as a *bona fide* principle of justice, will select a relatively narrow range of competing principles from the immense range of logically possible principles.

§2 Although this consequentialist account of justice seems to me to have interesting normative implications for many contemporary moral issues, I want to steer clear of applied ethics here. Doing even rough justice to all the consequentialist pros and cons of any interesting real-life moral issue would consume far too much space. Nevertheless, some illustration of my consequentialist account of justice is in order. Suppose we want to find a principle which provides a sufficient condition for judging justice within some distributive domain, say distributing opportunities for

higher education. Applying first the greatest happiness standard, we might ask: "Which principle specifying a reasonable compromise among justice-based claims in this domain is least likely to create unhappiness?"

Fortunately, *bona fide* principles of justice will generally have some inbuilt tendency to be compatible with promoting happiness, because each just-making criterion evolved in the human understanding as a response to individuals' deep aversion to types of distributions causing misery. Plausible answers to the question, "Which policy *both* achieves a reasonable compromise between justice-based claims in the domain of opportunities for higher education *and* is not obviously incompatible with promoting happiness (and will also satisfy our other moral standards)?" will yield our justified principles of justice in this particular domain. Since there will probably remain a range of such principles, the question, "Which particular principles of justice must an informed person of good will comply with when it comes to distributing opportunities for higher education?" has no completely definite answer at the first-order level of moral reasoning, and the final answer must be determined by purely procedural principles at the level of second-order accommodation.

What justice-based claims are there in the domain of opportunities for higher education? Many claims are advanced within this domain on the basis of desert, need, equality, rights and good-faith expectations, in a way strongly backed by people's sense of justice. For example, those who worked hard to meet high academic standards argue that they *deserve* to be admitted to university. Those who come from socially disadvantaged ethnic groups which need doctors or lawyers or academics as role models argue that they especially *need* a university education and should perhaps enjoy relaxed admissions standards. Those with relevantly equal abilities argue that they should be treated *equally*, regardless of socioeconomic background or other characteristics, when it comes to access to university spaces. Each of these claims might be couched in the language of *rights* and some may even

be justifiable as rights. Certainly, a lot of people have *good-faith expectations* of being admitted and often feel unjustly treated whenever admissions standards are suddenly raised to keep them out.

Now, finding a principle or policy which specifies a reasonable compromise between the conflicting justice-based claims in a way that will not conflict with the ultimate consequentialist test or its approximations obviously requires all the judgment, wisdom, experience and intelligence which we can mobilize for the task. It may be impossible ever to be absolutely sure that one has hit upon the best solution, but the line of effort required by a rational conception of morality is fairly clear. Most *possible* principles or policies fall clearly outside the range of what is plausibly justifiable as morally permitted. One such possible principle that is outside this range is: "Distribute university spaces preferentially to those with low intellectual capacity." This principle is outside the range of justifiable principles of justice, both because it does not specify a reasonable compromise between people's actual justice-based claims and because distributing university spaces preferentially to those with low intellectual capacity will clearly not promote happiness, not even the happiness of those who get to occupy these spaces (without fundamentally changing the character of the institutions we now call "universities"). Much more happiness will be created by giving spaces preferentially to those most able and motivated to benefit themselves and society from a university education.

One principle of justice which might fall inside the range of the rationally permitted is: "Make university spaces available to all those who achieved at least an 80% grade point average in the final year of high school and applied before June 11th, except for applicants from certain disadvantaged backgrounds." Of course, literally thousands of different variations of principles of justice can be imagined in this situation which are all rationally permitted by first-order moral reasoning.

We can try to narrow down the number of competing principles somewhat by applying, in combination, the greatest happiness test, the moral consistency test, the

respect-for-persons test and the voluntary agreement test. The remaining principles must then be submitted to the judgment of legitimate decision-makers elected or appointed for this purpose. Only after one such principle has been selected by this second-order process can a *bona fide* principle of justice be regarded as morally justified, in the sense that compliance with it would generally be required.

§2 Does our consequentialist account of first-order moral justification actually succeed in meaningfully narrowing, in various domains, the range of principles of justice that any informed person of good will must regard as morally justified (at least subject to second-order legitimation)? Or can everyone fit their particular normative prejudices into this framework? The general answer I will defend in the next chapter is this. In complex, pluralist societies with at least moderately favorable economic prospects, a consequentialist framework requires us to embrace directly responsive principles of justice which are broadly liberal-egalitarian in character. Within these limits, however, lawmakers (and indirectly the people who elect them) have morally considerable room for context-dependent judging in their choice of principles of justice as well as for first-order rational disagreement. But a consequentialist framework could usually, I believe, ensure enough common ground for fruitful debate. If someone is dissatisfied with the degree of normative indeterminacy which a consequentialist framework entails, in the abstract, for the project of spelling out the demands of justice, I can only say: "Give me a theory of justice which is more determinate at the abstract level, and I will give you reasons why informed people of good will need not find its acceptance rationally required (or even permitted)."

There are some things of which a theory can be had, and others of which perhaps only an account can be given. A *theory* of justice could plausibly be thought to consist in showing how rational agreement on a highly comprehensive set of normative principles for the distribution of the benefits and burdens of social

cooperation can be reached which is sufficiently responsive to patterns of expectations and sentiment about justice to enlist enough people's sense of justice in support of these principles to make them the basis for a stable social consensus about justice. An *account* of justice consists in explaining why no such agreement has been reached or is to be expected in pluralist liberal democracies, except within very broad limits, and stating how we can nevertheless implement measures to minimize the destructive force of people's sense of injustice while, at the same time, trying to foster gradual convergence on such principles over time by building a stable structure of expectations. I have only attempted to accomplish a fraction of the first part of such an account here and hardly touched the second part, namely the question of what kind of second-order accommodation constrained by which procedural rules, is most promising in the context of a particular society and within various domains.

We cannot predict, with great precision, which principles of justice will turn out to be preferable in various domains, independent of information about existing patterns of people's expectations and sentiment about which just-making criteria are appropriate in this domain. On this account, finding a principled response to people's justice-based claims is no longer a matter of erecting a grand structure of timeless abstractions on the bedrock of pure rationality, but will involve the applied ethicist in sociological field studies or at least require an awareness of the contributions of social scientists. What people already *feel* to be just, and how negatively they would react to a principle that ignored such sentiments, will strongly influence (but not by itself determine) which principles promote happiness, which principles minimize the need to treat individuals as a means to others' goals, and which principles it is to our mutual benefit to voluntarily adopt.

The recognition of this fact may, of course, itself motivate the pressing of extravagant claims in the name of justice. The proliferation of blatantly self-serving claims in the language of justice is a common practical problem for consequentialist

constructivism or any other theory of justice whose principles emerge partly from contextual and pragmatic considerations (i.e. all remotely plausible theories). But the illegitimacy of such claims is often evident by their lack of having credible backing from just-making criteria. Satisfying brazenly phony claims put forward in the name of justice would generally have strongly negative consequences and will, therefore, be ruled out by the ultimate consequentialist test. In a political community, extravagant claims and sentiments cloaked in the language of justice will be held in check by counter-claims and counter-sentiments, although at a cost of rising tensions. It is precisely in such situations that a clearer appreciation of the possibility of rational disagreement about justice may be helpful to motivate as well as constrain compromise-seeking.

3.4 THE CONFLICT BETWEEN LIBERALS AND NON-LIBERALS

This chapter offers a brief assessment of the scope for rational disagreement about the conceptions of justice offered by liberals and their opponents.

3.4.1 Liberal and non-liberal conceptions of justice

§1 Two value tensions at the heart of the concept of justice have created such profound divisions between substantive conceptions of justice that no credible attempt to assess the scope for rational disagreement about justice can pass them by. I mean the divisions between liberal and non-liberal theories of justice, and between the theories of libertarian and egalitarian liberals. The literature generated by these value tensions is so vast that I cannot possibly hope to do more than wrestle with some fundamental issues and to create some presumption in favor of my conclusions.

Liberals share the core belief that grown-up, mentally competent individuals are best left to judge for themselves how to live their lives, and that they should enjoy as much freedom to live according to their own preferences as is compatible with a similar freedom for all. This belief is *not* the same as believing that some lives are not lived better than others, or that individuals cannot make major mistakes in their quest for the good life. The liberal core belief has been given a number of formulations, and they are not obviously equivalent in every respect. One can say, with Mill, that people or governments can justifiably interfere with the actions of any person *only* to prevent harm to others:

> ...the sole end for which mankind are warranted...in interfering...is self-protection. That the only purpose for which power can rightfully be exercised over any member...is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled...because it will be better for him to do so, because it will make him happier...These are good reasons for remonstrating with him...but not for compelling him. (13)

Or one can say that there is a private sphere of action for each individual into which the state must not step, and which the state must protect against interference by others.[1] Or, to use the formulation that has served as the starting point for much contemporary debate, one can say that the state should be neutral between diverse ideas of the good.[2]

In a helpful attempt to clarify the complexity of the standard liberal creed, McDonald has taken care to distinguish three levels of liberal political thinking: the justificatory level, the political-thesis level and the contextual level (1991, 220f). These levels of *political* thinking are roughly analogous to the meta-ethical, normative and applied levels of *moral* thinking. According to McDonald, the two-part core of the standard liberal creed, at the political-thesis level, is (i) the recognition of rights which "establish normative limits on the use of both state and private power for the sake of members of the society in question"(225), and (ii) "the attribution of rights to *individuals*"(226).

While granting that liberals may, in principle, attribute rights to groups as well as to individuals, McDonald stresses the basic ways in which liberal politics tends to be strongly individualistic. For liberals, the primary role of the state is to protect the civil and political rights of individuals; the state is to protect these rights without enforcing some ideal of what makes a life worthy or a person virtuous; the choice of how to live is vested with individuals as the locus of intrinsic value and importance; and individuals must be given a significant range of choices (226).

All capsule characterizations of the standard liberal creed are afflicted with much the same vexing problems of interpretation. Which freedoms are compatible with the same freedoms for all? What constitutes sufficient harm to others to justify interference with an individual's actions? What are the boundaries of the "private sphere of action" or "the normative limits on the use of both state and private

---

[1] Cf. e.g. Narveson (1988).
[2] Cf. Dworkin, "Liberalism" (1985)

power"? What exactly does state neutrality among diverse ideas of the good amount to? How are conflicts between the rights of different individuals to be mediated? Does our consequentialist moral framework permit us, within the usual range of indeterminacy, to answer such questions in principled ways, or must we fall back on purely ad hoc pragmatism? To ask an even more fundamental question: must consequentialists necessarily be liberals in pluralist societies? Or are there perhaps forms of non-liberalism with greater promise for helping us progress toward the goal of achieving human well-being?

This chapter can accomplish only a small part of the task of defending what seem to me reasonable answers to a few of these immensely complex questions.

§2 I want to begin by clarifying the relation between liberal politics and views about what constitutes a good or worthwhile life. Liberals are frequently contrasted with perfectionists. Kymlicka seems to draw such a contrast when he says:

> Such theories [as Marxist perfectionism] are 'perfectionist' because they claim that certain ways of life constitute human 'perfection' (or 'excellence'), and that such ways of life should be promoted, while less worthy ways of life should be penalized. This is unlike liberal or libertarian theories, which do not try to encourage any particular way of life, but rather leave individuals free to use their resources in whatever ways they themselves find most valuable. (1990: 187)

But contrasting liberals with perfectionists, as if "perfectionist" were equivalent to "non-liberal", seems to conflate two cross-cutting distinctions. A term is clearly needed for a certain view within ethics, the view that some ways of life are known to be *better* or *worthier* than others (I shall use the terms "good" and "worthy" interchangeably here). And another term is needed for a certain view of politics, the view that the state should protect a private sphere of individual action within which individuals can pursue whatever ways of life they prefer, whether worthy in anyone's opinion or not. If we adopt "perfectionism" as the term to refer to the first view, and "liberalism" as the term to refer to the second view, then the category of

perfectionists and that of liberals are not mutually exclusive. It is an open question whether or not someone who believes she knows which ways of life are worthier than others will espouse the politics of protecting a private sphere of action.

There is, however, a problem with using the term "perfectionists" simply for all those who believe that we can judge some forms of life to be better or worthier than others while reserving the term "non-perfectionists" only for those who deny that any such assessment can in principle be made. Used in this way, the distinction between "non-perfectionist" and "perfectionist" would simply duplicate the distinction between "nihilist" and "non-nihilist". Even if there are people who can remain sincerely nihilistic or even agnostic about whether some lives go better than others, they are of no interest to me here. So if we do not want to define "perfectionism" in such a way that all non-perfectionists become essentially nihilists, I think we must clarify the distinction between perfectionists and non-perfectionists roughly as follows:

Perfectionists are those who think that there are *objective* standards of the good human life - i.e. standards which are completely *indifferent* to facts about what ways of living people actually prefer. As I shall use this term, perfectionists also believe that these objective standards usually allow them to determine or predict the goodness of a person's life with sufficient confidence and precision so that coercing people into leading good lives becomes viable. Perfectionists who think that there exist objective standards of the good human life, but that they are so vague, epistemically elusive and difficult to apply that we can't generally tell whether a person is or isn't leading a good life are, for practical purposes, hardly distinguishable from non-perfectionists. Different types of perfectionists will, of course, be more or less pluralistic in what sorts of values they include among constituents of a good human life.

Non-perfectionists, as I shall use the term, are those who think that there are no objective standards of the good human life (in the sense defined above), and that

at least one *necessary* condition for what sorts of values make a mature human life good always is that the person living it *likes* or *chooses* to live according to these values. Although non-perfectionists don't deny that the goodness of a person's life can generally be determined by observation, they do not believe that currently (in the absence of a highly accurate predictive psychology) we can tell in advance with great confidence and precision whether a particular way of living will constitute the good life for a given person.

Again in the interest of avoiding empty categories, I must further distinguish *radical* from *moderate* non-perfectionists: those who think that knowing a person's preferences are a necessary and sufficient condition for determining what makes that person's life good, and those who think that such knowledge is only a necessary condition. For radical non-perfectionists, the question of what is good or worthy for a person reduces exclusively to an empirical question about what this person likes or prefers. For moderate non-perfectionists, *some* general values whose presence generally tends to make a human life better can in fact be known, and persons who like or prefer the opposite of these values can, in fact, be mistaken about their own good. I feel compelled to make this somewhat awkward distinction because, again, I doubt that there are in fact radical non-perfectionists among sane people who have thought the issue through. A radical non-perfectionist would have to *deny* that anyone's life could ever go better than anyone else's, no matter how uninformed or self-destructive the choice of someone's values, because they are after all the values he prefers. Radical non-perfectionism would again seem to collapse into a sort of nihilism or at least a value subjectivism which, in the face of our knowledge of at least some of the factors which determine people's preferences, strikes me as highly implausible. In what follows I shall, therefore, concern myself *only* with moderate non-perfectionists: those who think that the goodness of a person's life depends crucially on what is liked or preferred by that person and cannot be predicted with great confidence and precision for any given person, even though some very

general values whose presence tends to make human life better can in fact be known.

What is the relationship between liberals and perfectionists, in the senses defined? The most that can be said, I think, is (1) that some who espouse liberal politics are perfectionists and some are not, (2) that non-perfectionists need not *necessarily* espouse liberal politics, and (3) that perfectionists who espouse liberal politics can do so for very different reasons.

(1) A non-perfectionist who espouses liberal politics may do so for one of three basic types of reasons: because it provides the most pragmatic response to the brute fact of pluralism about the human good; because it is the most "neutral" ground in the face of a perfectly rational diversity of ideas about the good; or because it is empirically most likely to lead to social conditions which enable individuals to discover their own good. (I shall examine these reasons below.)

(2) But non-perfectionists need not necessarily espouse liberal politics. There is nothing strictly inconsistent about holding that what gives a life its value is highly individuated and at the same time believing that the state need *not* protect a private sphere of action within which individuals can do whatever they prefer. The non-perfectionist thesis is only one of many compatible theses *within* ethics a person may hold. A non-perfectionist might, quite consistently, also hold the thesis that there is nothing more to morality than verbal posturing or venting of emotions, and that anything whatever is permitted. And such a view clearly need not lead to liberal politics.

(3) A perfectionist who espouses liberal politics may do so for at least two reasons. All perfectionists must by definition think that there are objective standards of the good life by which they can determine the worthiness of someone's life with considerable precision and confidence. What I shall call *type-A perfectionists* espouse liberal politics because they think that only the self-determined life is worthy, and that a liberal social order is most likely to permit self-

determined ways of life. Any perfectionist who affirms self-determination as the central ingredient of a worthy life must also affirm a whole web of political values that can plausibly be considered conducive to creating favorable conditions for self-determination such as, for example, a liberal education, access to economic means for the realization of one's choices, or the creation of valuable opportunities for the exercise of self-determination. Type-A perfectionists, by the very nature of their particular idea of what makes a human life worthy, will necessarily recognize a much broader diversity of equally worthy ways of living than other types of perfectionists. But self-determination is clearly a value that presupposes the use of state power in order to create favorable conditions both for the cultivation of self-determining character types and for the effective exercise of self-determination. (Examples of contemporary type-A perfectionists are Joseph Raz or Thomas Hurka.)

Type-B perfectionists do not regard self-determination as necessary or even important for a worthy life, but nevertheless espouse liberal politics (in pluralistic societies) because they think *either* that any attempt to use state power would simply unleash a destructive struggle for control of the state among opposing groups *or* that using state power to enforce the worthiest way of life would be downright counterproductive to converting people to such a way of life.[3]

Both type-A and type-B perfectionists could intelligibly be called liberal perfectionists or perfectionist liberals. *Type-C perfectionists* are a residual category of non-liberal perfectionists which includes all and only those who think (i) that they are in possession of objective standards of the good life by which they can determine the worthiness of almost any person's life with great precision and confidence; (ii) that self-determination is not necessary for a worthy life; and (iii) that state power should be used to enforce worthy ways of living for everyone. It may be upsetting to those who want to see a simple contrast between liberals and

---

[3] These two arguments are forcefully advanced by Locke in defense of freedom of religion (see *A Letter Concerning Toleration*). See Susan Mendus for an assessment of the comparative merits of Locke's and Mill's arguments for liberal politics (*Toleration and the Limits of Liberalism*, 1989).

perfectionists, and who define liberalism in terms of state neutrality between diverse ideas of the good, that by no means *all* perfectionists advocate the use of state power to enforce favorable conditions for their idea of the good, while those liberals who justify liberal politics in terms of the value of self-determination clearly *do* advocate the use of state power to enforce favorable conditions for their central idea of the good.

But even most hardcore type-C perfectionists who want the state to impose their particular idea of the good on everyone would probably agree that there should be some *a priori* restrictions on what the state can do, such as a prohibition on concentration camps for the large-scale extermination of dissenters. Moreover, many type-B perfectionists may also value self-determination to a degree, but only as one value among others, so that it may at times become a difficult judgment for them what it is more important - to respect someone's right to self-determination or to prevent her from making a major mistake about the good life.

The contrast between those who espouse liberal politics and those who claim to be able to determine objective standards of a good life for each person is further complicated by the fact that even the most non-perfectionist liberal would certainly want to keep certain antisocial tendencies in check, while even the most fervent type-C perfectionist would hardly want to control every trivial detail of individuals' lives. Both those who espouse liberal politics and and those who don't will have to make judgments about which behaviors are antisocial enough to be made illegal or too trivial to be bothered with at all. Kymlicka seems again surprisingly sweeping when he asks,

> Should we therefore be perfectionists, supporting state policies which discourage trivial activities to which people are mistakenly attached?"
> (1990: 203)

It may well turn out that many behaviors which liberals tolerate are trivial enough in the eyes of many non-liberal perfectionists to be tolerated by them as well, while behaviors which many non-liberal perfectionists want to make illegal are judged to

be antisocial enough by most liberals to be declared illegal by them as well. Neither liberals nor perfectionists are united among themselves about what behaviors can be tolerated without causing unacceptable harm to identifiable others or to society's vital long-term interests in civil peace, security, prosperity, cultural flourishing and a livable natural environment.

§3 Contrary to what is sometimes implied, then, the two categories of "liberalism" and "perfectionism" are best regarded as intersecting in complex ways rather than as sharply disjoint. Only because of an unfortunate conflation of cross-cutting distinctions has it been possible for some perfectionists to accuse liberals in general of being skeptics about the good and for some liberals to accuse perfectionists in general of being totalitarians by definition. Drawing clear distinctions about the different reasons for espousing liberal politics, including perfectionist reasons, can avoid a great deal of sterile disagreement. This is not to deny that many perfectionists who see a realistic chance for achieving control of state power will, as a rule, want to use state power to get the obstinate to lead worthy lives. But liberals, whether they be type-A perfectionists or non-perfectionists, are certainly no exception to the rule: they, too, propose to use state power to enforce their ideal of the best social order. The obvious question is whether liberals can, without inconsistency, consider the choice of *their* favored value – i.e. self-determination or letting people live by their own preferences – as sufficiently special to entitle *them* to use state power while denying type-C perfectionists the same right. I shall address this question below. The issue I want to address next is how ideas about the good life and about the best social order are connected with conceptions of justice.

§4 Having distinguished perfectionist and non-perfectionist views about the good life, and liberal and non-liberal views about the best social order, we now face a

choice: on which of these two distinctions should we base our fundamental distinction between different substantive conceptions of justice? This choice is non-trivial because differences marked by our fundamental distinctions will channel our thoughts and disagreements in profound ways. The distinctions we draw between substantive conceptions of justice can either mark the difference between various types of views about the human good that can ultimately be advanced to justify different principles of justice. Or they can mark the difference between various types of views about the best social order which can also be advanced to justify different principles of justice.[4]

If we choose the first alternative, our distinctions between different basic conceptions of justice will mark such differences as, for example, that between non-perfectionists, type-A perfectionists and type-B perfectionists - all of whom *may* in practice, at least in a pluralist democracy, converge on a broadly liberal social order and on similar *substantive* ideas about justice. But not only will very different basic types of views about the good life sometimes be consistent with the same substantive principle of justice, it will also be the case that subtypes of the same basic type of view about the good life will often justify very different substantive principles of justice. The relation between views about the good life and principles of justice would seem to be far more intractable than the relation between views about the best social order and such principles. We could, for example, distinguish between an essentially liberal conception of justice and various types of non-liberal conceptions supported by those perfectionists who want to use state power to advance some ideal of the good incompatible with self-determination. This taxonomy of different conceptions of justice assumes, of course, that there is

---

[4] Or they can mark some other interesting difference between proposed principles of justice, such as whether they are directly or indirectly responsive, intuitionistic or non-intuitionistic. Which distinctions we draw depends, of course, on our purposes. But sometimes we seem none too clear about our purposes. It is probably an understatement to say that confusion reigns in political philosophy about how to group different principles of justice into conceptions of essentially the same type.

something sufficiently special and cohesive about liberal politics to justify our distinguishing liberal conceptions of justice from the immensely wide and diverse range of all the others. That there is something that makes liberalism special in this way is vigorously disputed by some non-liberals like Alistair MacIntyre (cf. 345).

I will argue below that, indeed, liberal politics is special in a way that justifies us in drawing a fundamental distinction between substantively liberal conceptions of justice and the rest. But first I want to draw my overview of basic distinctions between different conceptions of justice, and draw them rather starkly. Understanding distinctions in their unqualified starkness is often necessary, I believe, before one can intelligibly blur them - especially as we enter the conceptual maze of liberalism, its alternatives, and its internal divisions. My categories are not, of course, meant to be exhaustive but simply useful for entering contemporary debates.

I. *Liberal conceptions of justice:* "Social goods should be distributed in a way that allows individuals the freedom to live as they prefer, while promoting the widest range of valuable opportunities for self-determination."

    A. Libertarian liberalism

    B. Egalitarian liberalism (often called "welfare" liberalism)

II. *Non-liberal conceptions of justice:* "Social goods should be distributed in a way that eliminates unworthy ways of living and promotes the greatest possible flourishing of worthy ways, regardless of the extent to which doing so conflicts with people's preferences."

    A. Communitarian non-liberalism

    B. Utopian non-liberalism

    C. Religious non-liberalism

Each of these generic principles characterizing liberal and non-liberal conceptions of justice does, admittedly, gloss over a multitude of substantive

differences within each camp.[5] Liberals can be usefully grouped into a number of opposing factions (I shall briefly consider the conflict between libertarian and egalitarian liberals). Non-liberals can also be grouped into many different types such as, for example, into communitarian, utopian and religious non-liberals (which will further subdivide into countless antagonistic factions).

Of course, communitarians, utopians or religious believers may consistently support or oppose liberal politics. Broadly speaking, communitarians believe that community is an essential constitutent of the good life, that a person's self is through and through the product of a particular cultural community and can realize itself only through participation in the life of that community, and that our social obligations, commitments and attachments are are not simply a matter of voluntary choice, but deeply part of who we are. MacIntrye, Sandel and Charles Taylor are usually cited as leading contemporary communitarians, even though a great deal of what they say seems to me to have been well articulated in F.H.Bradley's *Ethical Studies*. Those communitarians who oppose liberal politics seem to be concerned with strengthening individuals' identification with their social roles and with maintaining whatever received traditions are most likely to sustain a close-knit network of warm human attachments and a sense of belonging within a community, *even at the expense of denying individuals such core liberal rights as the right to freedom of religion, expression, or association.*[6] A prominent sub-type

---

[5] Examples of egalitarian liberals are John Rawls, Kai Nielsen, Derek Phillips, Brian Barry, David Braybrooke, Ronald Dworkin, James Sterba, as well as R.H. Tawney and L.T. Hobhouse. Examples of libertarian liberals are Robert Nozick, David Gauthier, Jan Narveson, and Friedrich Hajek. A possible example of a communitarian perfectionists is Alistair MacIntyre. Michael Walzer, William Galston, Michael Sandel or Will Kymlicka are best seen as liberals who try to make room for communitarian values within liberalism. Examples of utopian perfectionists are Plato, Marx, Thomas Moore or B.F. Skinner.

[6] There can be communitarian liberals, or liberal communitarians. Cf. Will Kymlicka, *Liberalism, Community, and Culture* (1989), for an outstanding attempt to accommodate communitarian concerns in a principled way within a basically liberal vision of the best social order. Allen Buchanan has also argued that liberalism can make room for the good elements of communitarianism while avoiding its totalitarian elements (1989). An excellent discussion of the complexities associated with the "collective rights" solution to accommodating the value of community within liberalism is to be found *The Canadian Journal of Law and Jurisprudence*, Vol IV, No 2, July 1991, which is devoted entirely to the topic of collective rights.

of non-liberal communitarians are classical conservatives, such as Roger Scruton. For non-liberal communitarians, principles of justice will have this general character: "Distribute social goods according to whatever distributions will strengthen community bonds and make individuals identify themselves as members of a shared culture with a common good - no matter whether such policies interfere with most individuals' preferred life choices."

Utopian non-liberals believe that promoting human excellence consists in revamping society according to some blueprint of ideal social arrangements, if necessary without the informed consent of the majority of the population. Stated bluntly, utopian principles of justice will have this character: "Distribute social goods according to whatever my particular blueprint of the ideal society specifies." Religious non-liberals believe that promoting human excellence consists in following the moral code associated with what each considers the true religion. Their principles of justice will have this character: "Distribute social goods according to whatever principles are laid down in the Holy Scripture of our particular religion." Whereas, for communitarian non-liberals, religion may play an important role as one means of preserving community bonds, for religious non-liberals, spreading the true religion is the ultimate value, even where it destroys community bonds. It needs stressing, however, that many religious believers, just like many communitarians and utopians, consistently can, and in fact do, support liberal politics rather than the use of state power to enforce their particular ideal of the good.

This, then, is a starkly drawn matrix of how I see the basic divisions between various substantive conceptions of justice. The variety of ideas about justice appealed to in real-world debates is bewildering; the number and complexity of arguments offered in support of conflicting ideas is often overwhelming. We now have enough of a synoptic perspective, I think, to sketch the outlines of the strong

consequentialist case that can be made for a substantively liberal conception of social justice.

3.4.2 On the nature of human well-being

§1 If those perfectionists are correct who hold that we can know with confidence *both* what is good for each person *and* that this good has little or nothing to do with self-determination, then obviously the case for liberalism is weak. In that case the only two reasons I can see for espousing liberal politics are that attempts to enforce the worthiest way of living will unleash a destructive struggle for control of the state among opposing groups in pluralistic societies or will often be downright self-defeating as attempts to convert people to such a way of living.

But do perfectionists who hold that we can know with confidence what is good for each person *and* that this good has little or nothing to do with self-determination have a plausible case at all? Is there perhaps a possibility of looking objectively at some basic psychological facts about human nature and the human condition and drawing *some* conclusions about what makes a human life good or worthy? I think there is.

What could possibly count as making a human life good or worthy? At the most fundamental level, only two interestingly different *prima facie* plausible answers seem to me possible. Either we accept one of many competing religious or mystical accounts - the view, for example, that we are God's creatures, here on earth to earn our place in heaven through obedience to God's will; or the view that we live in a morally ordered universe whose moral ordering principles can be discovered in much the same way as its natural ordering principles. Or we are in a morally indifferent, uncaring universe governed by blind material forces, thrown into existence by the mechanistic process of evolution, and all of us must reflect on our own experience and try to discover values that seem most likely to promote whatever well-being the constraints of our situation permit. I think, though I shall

not try to argue this here, that a sober look at the evidence, as opposed to wishful thinking, shows belief in the second alternative far more plausible than belief in the first - at least insofar as belief in the first alternative involves specific claims about the nature of God or the nature of this independent moral order.

A number of questions arise. Is at least some degree of rational consensus on the *range* of plausible conceptions of personal well-being possible? Assuming the answer is yes, what can we say about this range of plausible conceptions of well-being for different individuals? What defines the limits of this range? Are there some properties which all plausible forms of personal well-being have in common? Such questions, I believe, must at least make sense to any grown-up, mentally competent individuals who approaches them with an open mind.

After due reflection on the answers offered by others, each of us will have to introspect his or her own deepest understanding to answer these questions. What ultimately matters for a person's well-being? Having abundant *resources* of a certain kind, such as Rawls' primary goods? Or being in that kind of *hedonic state* which individuals in this state identify with well-being? I think the most plausible conception of well-being is not captured by either alternative. Well-being cannot consist in having abundant resources, because no matter how abundant one's resources, they are not sufficient for well-being unless a happy state of mind is actually achieved through them. Beyond a certain minimum, additional resources are not necessary for well-being; an abundance of possessions may, in fact, detract from a happy state of mind. Nor can well-being *simply* be a pleasant hedonic state. A happy state of mind caused by drug-induced delusions or by the life-long stimulation of one's cerebral pleasure spots, while one lies strapped down on a bed, doesn't plausibly capture what ultimately matters to a person's well-being.[7]

---

[7] Nozick's example of the "experience machine" vividly makes this point: cf. *Anarchy, State, and Utopia*, 42-45. It is, of course, conceivable that some thoroughgoing hedonists will opt for a life hooked up to the Experience Machine, and most of us would probably like to plug into it intermittently.

What, ultimately, matters to a person? Let me sketch my own rough idea. Others, in the light of their own introspective understanding of what *they* fundamentally care about, may well find reasons to disagree with it: I am a friend of rational disagreements. Apart from satisfying one's biologically-determined needs, what constitutes well-being, I think, is working toward and achieving the goals one has chosen from a range of meaningful options, *and* experiencing a happy state of mind as a result of doing so. I believe that almost always, for a happy state to be experienced as a result of doing what one values, one must expend some degree of real effort, even though the happiness achieved is not proportional to effort expended. Not everything a person values doing and actually does will result in her experiencing a happy state of mind. A happy state of mind is certainly not the only thing a person can intelligibly value; in fact, a happy state of mind is generally a byproduct of sincerely valuing something else for its own sake and making progress in promoting this value. It is quite possible for people not to value their own happiness, or to sacrifice it for the sake of other people or other values, or to knowingly act contrary to their well-being out of weakness of will, and all too familiar for them to value their long-term well-being less highly than intense immediate enjoyments.

The most important point is this. What is conducive to a person's well-being is something which that person can be mistaken about, but at the same time something which may differ considerably from one individual to the next. The basic types of things that human beings, such as they are, can actually value in this world, such as it is, with the result of experiencing a happy state of mind which is not outweighed by subsequent unhappiness, add up to a lengthy, varied, but quite finite list. Although this list will get less universal and more controversial the more specific it gets, it will almost certainly, for most people, include items like securing such things as food, shelter, and other requirements for basic physical and social functioning, instinctual enjoyment, affectionate human relationships, self-esteem, a

meaningful social role or other ways of getting a sense of social belonging, as well as personal freedoms. Out of varying amounts of these and other values, each individual fashions his own mix of values to live by, although there will of course be a heavy cultural ingredient in any such mix.

What gets something onto this list of values which are constitutive of human well-being? Nothing more mysterious, than that we can recognize it as frequently or universally necessary or conducive to people's experiencing their lives as good. Some degree of convergence is possible, I believe, between the reflective equilibria of informed people of good will on what should go onto this list. Such a list of constitutents of human well-being, I think, can furnish the basis for an appreciation of the diversity of possible forms of personal well-being as well as an understanding of plausible limits to such forms. In cashing out the important notion of "plausible limits on the range of values which can conduce to personal well-being", a broad consensus is possible that certain values will clearly fall inside the range (e.g. friendship) and certain values will clearly fall outside (e.g. heroin addiction), while others may well be matters for rational disagreement.

But I take it to be a fundamental fact about the human condition, obvious to any grown-up, intellectually competent person with an open mind (and perhaps a certain minimum of education), that there exists a diversity of values for human beings which no one life can encompass and which can be permutated in almost infinitely many ways. And this fact, I believe, furnishes a powerful consequentialist argument for liberal politics - an argument which is neither identical to the pure pragmatist's or political realist's argument based on the avantage of mutual accommodation nor to the perfectionist argument based on self-determination (see below).

3.4.3 On the nature and justification of liberal politics

§1 Let me try to characterize the nature of liberal politics more clearly before considerating justifications for it. I follow the current practice of taking Dworkin's definition of liberalism in terms of the principle of government neutrality among people's diverse ideas of the good life as my point of departure:

> ...government must be neutral on what might be called the question of the good life...The first theory of equality supposes that political decisions must be, so far as is possible, independent of any particular conception of the good life, or of what gives value to life. (1985: 191).

In the years since Dworkin's 1978 paper "Liberalism" tried to rally liberals around this "liberal neutrality" principle, competing interpretations of liberal neutrality have emerged. What does it mean for liberals to claim that they are neutral between different ideals of the good - and how can liberal neutrality be justified?

There seem to be at least three basic ways in which liberal neutrality might be understood. The terminology here is far from standardized, and I want to adopt reasonably descriptive labels that make sense to me. I want to distinguish "neutrality of aim" from "neutrality of effect". Neutrality of effect can be further subdivided into what I shall call "non-differential" and "differential" neutrality. Which kinds of neutrality can we have?

Neutrality of aim requires that the official reasons for justifying the state's policies should never include the promotion of some group's ideas of the good life over that of other (legitimate) groups, even though most policies will, of course, impact quite differently on different groups. *Non-differential neutrality of effect* requires that the policies of the state should be neutral in their impact on different legitimate groups. All legitimate groups, that is, should be helped or hindered equally, in some plausible sense, in allowing each individual member of such a group to pursue her idea of the good life or in allowing the group as a whole to gain adherents to its values. (These twin objectives may well diverge.) *Differential neutrality of effect* requires that no legitimate groups be effectively prevented from

pursuing their ideas of a good life, although non-coercive means of persuasion and affirmative action programs may be used to correct specific undesirable inequalities by giving differential encouragement to disadvantaged or underrepresented groups over others. Difficult judgment calls will obviously be required to balance affirmative action for the disadvantaged or underrepresented without denying or seriously impairing other groups' effective pursuit of *their* ideas of a good life.

Obviously, all liberals must define some standard of legitimacy for groups: groups which value murder, rape, theft and the like are ruled out. And all liberals must draw a workable distinction between purely *personal* values (between which the state should be neutral), such as someone's religion, career, sexual preferences, and a multitude of lifestyle choices; and *political* values (between which the state cannot be neutral), such as liberal neutrality itself, peace, civility, tolerance, political freedoms and civil liberties, a range of social institutions like the market and distributive goods which serve as flexible means to diverse ends, and some principles of justice according to which these things may be distributed.[8]

Assuming that the distinctions between legitimate and illegitimate groups and between personal and political values can be drawn without begging crucial questions, which of the above three basic interpretation of liberal neutrality should liberals be interested in justifying? As a useful criterion for government action, neutrality of aim would seem to be too weak and too strong at the same time, at least for consequentialist liberals. Too weak, because it should often count decisively against a policy that it is foreseeably disastrous for some group, even if the state's official justification make no reference whatever to its effect on any particular group. Too strong, because it is hard to see how most policies could avoid having the promotion of the good of some group (disadvantaged or otherwise) as their goal. Endorsing neutrality of aim would clearly open liberals up to the charge of hypocrisy.

---

[8] See Rawls (1988) for a lucid summary of political values liberals must accept.

Non-differential neutrality of effect would also seem to be too strong, because it is impossible to live up to and because it would effectively paralyze the government at every step. Rawls has this to say about non-differential neutrality of effect, which he calls "neutrality of effect or influence" and defines as the position "that the state is not to do anything that makes it more likely that individuals will accept any particular conception rather than another unless steps are taken to cancel, or to compensate for, the effects of policies which do this" (262):

> ...it is surely impossible for the basic structure of a just constitutional regime not to have important effects and influences on which comprehensive doctrines endure and gain adherents over time, and it is futile to try to counteract these effects and influences, or even to ascertain for political purposes how deep and pervasive they are. We must accept the facts of common-sense political sociology. (1988: 262)

These and similar considerations seem to me decisive against non-differential neutrality of effect. Of our three *prima facie* options for interpreting liberal neutrality, differential neutrality of effect seems to come closest to a viable political value. Rawls defines this position as meaning

> that the state is to ensure for all citizens equal opportunity to advance any [permissible] conception of the good they freely affirm. (1988: 262)

But this definition, too, is in obvious need of both clarification and qualification. And how a liberal clarifies and qualifies it will depend very much on the reasons that made him espouse liberal politics in the first place. Raz, for example, who bases his liberalism on the perfectionist value of self-determination, is quite explicit in advocating policies with substantially differential effects on different groups:

> ...a government whose responsibility is to promote the autonomy of its citizens is entitled to redistribute resources, to provide public goods and to engage in the provision of other services on a compulsory basis, provided its laws merely reflect and make concrete autonomy-based duties of its citizens....Autonomous life is valuable only if it is spent in the pursuit of acceptable and valuable projects and relationships. The autonomy principle permits and even requires governments to create morally valuable opportunities, and to eliminate repugnant ones....Perfectionist goals need not be pursued by the use of coercion. A government which subsidizes certain

activities, rewards their pursuit, and advertises their availability encourages those activities without use of coercion. (417)

A host of questions now confront liberals who espouse differential neutrality of effect. Is the idea of "equal opportunity" to be given a libertarian or some form of egalitarian interpretation? How is the line between persuasion and coercion to be drawn? When does affirmative action for some group become unjust discrimination against another? Which policies should liberals prefer concerning cultural minorities, melting-pot policies or multi-cultural policies? How can liberals draw a principled line between legitimate and non-legitimate groups without being accused of *ex*cluding socially benign ways of life and *in*cluding socially destructive ones?[9] I cannot pursue these complex questions here, but I consider it fairly obvious that informed people of good will, even if they share a broadly consequentialist moral framework, can be in immense disagreement about how exactly to answer them.

§2 If one is committed to some plausible form of liberal neutrality, how can one justify such a commitment? The consequentialist framework I have outlined would seem to allow at least four connected lines of justification to converge on liberal neutrality. Liberal neutrality can be defended (1) as providing the political pragmatist with the most promising way to avert civil strife in the face of pluralism about the good; (2) as being the most coherent position or most "neutral ground" in the face of a perfectly rational diversity of ideas of the good; (3) as being empirically most likely to lead to the sort of social conditions which enable the greatest number of individuals to achieve their interest in personal well-being; and (4) as being necessary for individuals to achieve one particularly important good, namely self-

---

[9] Discussions of the harm principle are numerous in the literature on liberalism. For recent examples, see Susan Mendus, *Toleration and the Limits of Liberalism* (1989); Joseph Raz, *The Morality of Freedom* (1986); and Lloyd Weinreb, *Natural Law and Justice* (1987). For a discussion of different effects of state intervention on the affairs of groups - and a defense of permitting certain "self-constituting" groups a large measure of autonomy even at the expense of the autonomy of individuals within such groups - see McDonald, "Should Communities Have Rights? Reflections on Liberal Individualism"(1991).

determination (although appealing to the personal value of self-determination to justify the political value of liberal *neutrality* seems, on the face of it, blatantly inconsistent). Let me briefly consider these lines of justification.

(1) Since some degree of political accommodation of diverse ideas of the good is essential to securing peace in strongly pluralist societies, since social peace is important to the pursuit of almost any idea of the good, and since liberal neutrality can be plausibly construed as an attractive form of political accommodation, liberal politics would clearly seem to be a pragmatically rational response to the fact of pluralism. But it has at least three limitations. First, this pragmatic defense of liberalism in strongly pluralistic societies will, of course, turn into an equally pragmatic defense of non-liberalism in less pluralistic societies and possibly justify repression of minorities. Second, this defense fails to give non-liberals a deep rationale for why they should not try to topple liberal institutions when they see an opportunity to do so. And third, it is hard to see what should make hardnosed political pragmatists endorse liberal principles rather than naked power politics.

Raz illustrates the limitations of any attempt to defend liberalism from the point of view of the pure pragmatist. He describes an interesting liberal-sounding "perfectionist" approach to establishing basic social arrangements that promises to circumvent the objection that non-liberal perfectionism must necessarily suffer from excessive intolerance. Such a perfectionist "doctrine of justice", he says,

> will fulfill the social role of justice, but it will be agreed upon by a process of reasoning quite different from [Rawls']. *First*, different ideals of the good far from being excluded will form the starting points of the argument about a doctrine of justice. *Second*, and as a result, supporters of different conceptions of the good will follow different routes in arguing for the doctrine of justice. There will be unanimity in the conclusion but (given the different starting points) no unanimity on the route to it. *Third*, the common feature of most routes will be the reliance on a rational reconstruction of a process of bargaining by which the common overriding goal of reaching an agreement leads the parties to compromise by accepting a less than perfect doctrine as the optimally realizable second best. (129)

Raz imagines supporters of different forms of perfectionism to come together in a contractarian bargaining session, and to negotiate agreement on a set of principles which are the most favorable "second-best" principles possible from the perspective of each group. Whereas Rawls imagines his bargainers to be free and equal persons behind a thick veil of ignorance, Raz imposes no constraint at all on his perfectionist bargainers except the "overriding goal of reaching an agreement". (He assumes that the bargainers know that there is sufficiently widespread disagreement about ideas of the good in their society that no one such idea can be implemented. It is not clear whether he also assumes equal bargaining power or threat advantage, but presumably not.)

The outcome of such negotiations would then seem to be a complex function of the particular perfectionist starting points, the range of diversity of perfectionist beliefs, and relative threat advantages. If the *social* role of a doctrine of justice is conceived purely in the political pragmatist's terms, Raz concludes that one does not require Rawls' "veil-of-ignorance". Individuals will reach agreement on a set of distributive principles, even if they know how they differ in their ideas of the good, "provided they are united in ranking the need for a public conception of justice very highly in their order or priorities."(128) Nevertheless, mere agreement on principles that fulfill the social role of justice cannot be enough to make them morally acceptable. "For reasons which are fairly obvious", Raz says, "this procedure of bargaining may in some societies lead to the endorsement of highly wicked principles as that society's doctrine of justice"(129). What he must have in mind here is that justice cannot simply be equated with whatever agreement on principles of social coexistence can be reached, presumably because of unequal bargaining power may allow some fairly pernicious types of perfectionism to impose their ideas of the good on the rest of society. He is right, I think, that we cannot hope to support liberal politics very strongly in terms of a purely pragmatic *modus vivendi*.

(2) But the crucial point is not simply that there exists pluralism about the good, but that there is no one correct way to live for everyone. Even for any given person there is always a broad range of possible ways of life that would have been equally good. As we concluded above, it seems to be a fundamental part of the human condition that there is a diversity of values for human beings which no one life can encompass and which can be permutated in almost infinitely many ways. Although we cannot create favorable social conditions for all valuable ideas of the good to flourish simultaneously, our social arrangements can at least enable individuals to explore a wide diversity of values and to seek out their own versions of the good life. If this value diversity is widely appreciated as arising from the fact that the constituents of human well-being are naturally diverse and is not equated with confusion about the good, and if social institutions can be so designed to make room for a wide diversity of values to flourish, then those institutions should enjoy widespread support from people with many diverse ideas of the good.

This, I think, is essentially the defense of liberalism adopted in Rawls' more recent writings, as well as in those of Stout and Larmore, although they call themselves (misleadingly) pragmatists.[10] Stout writes,

> We need not agree on all matters of moral importance to agree on many, and where our judgments happen to coincide, we need not reach them for the same reasons. The judgments we share about the achievable good are ample enough to justify our practice of ascribing rights. They allow us to offer a pragmatic defense of certain liberal arrangements and concepts as justified, here and now, given that we can not presuppose, nor bring about by acceptable means, a more nearly perfect meeting of minds. (226)

Larmore has vigorously argued that liberal neutrality can and must be justified in a way that is itself "neutral" in some sense, rather than based on any such perfectionist value as autonomy, although liberals may quite consistently hold such values in their private lives (50f, 73f).

---

[10] Cf. Jeffry Stout, *Ethics after Babel* (1988); Charles Larmore, *Patterns of Moral Complexity*(1987); and Rawls,"The Priority of Right and Ideas of the Good"(1988).

In order to take up a neutral position by abstracting from our substantial conception of the good life, we are not...repudiating that conception or lessening our attachment to it....But we are declining to appeal to it for a particular purpose - that of devising common political principles. Neutrality is simply a means of accommodation. It is a stance that we adopt in order to solve a specific problem to which our various commitments give rise, and so it is not a stance that expresses our full understanding of our purposes. It establishes a modus vivendi between persons whose ultimate ideals do not coincide.

A number of different considerations move us to take up such a stance. There is, first of all, a norm of rational conversation....that requires that if we want to solve some problem but encounter disagreement about how to do so, we should retreat to neutral ground. Then, there are the reasons we have for wanting to converse with others about what political principles to adopt. Besides a desire for civil peace and sympathy for those whose ideals are similar to ours, the norm of equal respect demands that we explain our proposals to those whom they will affect. To these reasons may be added the likelihood that our pursuit of some substantial ideal of the good life will be protected if others, with different ideals, also agree that common political principles must be sought from neutral ground. (75)

Dispute has centered on whether appeals to norms of "rational conversation" and "equal respect" are, in fact, relevantly value-neutral; whether a retreat to shared values can, by itself, provide sufficient ground for workable social policies; whether seeking such shared values will only lead to widespread but not unanimous agreement and therefore leave certain groups out; and whether policies based on whatever shared values there are will differentially affect people who attach different degrees of importance to these values and will, therefore, in a sense be non-neutral.[11] Assessing these objections raises complex matters. In spite of all these complications, I think the fact remains that a political system that strives to accommodate a wide diversity of personal ideals can reasonably claim to approximate neutrality, in the relevant sense, more closely than a political system that does not.

(3) The third cluster of justifications for liberal neutrality are justifications which consider liberal neutrality as most likely to further the interest of individuals

---

[11] Cf. Richard Arneson, "Neutrality and Utility"(1990); Will Kymlicka, "Liberal Individualism and Liberal Neutrality"(1989); and Robert E. Goodin and Andrew Reeve, eds, *Liberal Neutrality* (1989).

in their own well-being. To be sure, individuals make mistakes, but history shows that entrusting corruptible governments with the authority to coerce people for their own good harbors far greater dangers. Allowing individuals to engage in "experiments in living" is our best promise for creating a rich cultural life, economic prosperity and other social conditions conducive to reasonably good lives for most.

On the other hand, it has been claimed that liberalism sacrifices many of the most important forms of human excellence for a chaotic diversity of inferior forms.[12] A marketplace of ideas of the good will no doubt allow some inherently offensive or oppressive values to win a large following, which necessarily makes the social climate inhospitable for others, destroys community bonds, and promotes the kind of mindless greed, self-absorbed individualism, alienating working conditions, anomie, exploitation, and cultural decadence which are allegedly so evident in contemporary liberal societies.

There is some empirical evidence for both sides in this debate. I shall not discuss it here, but simply state that, on balance, the evidence in the contemporary wold seems to me convincing that societies which strive for liberal neutrality tend to be no less successful in maintaining a rich cultural life and relatively prosperous economic conditions than societies that do not.

(4) The fourth defense for liberal policies stresses the special importance of the value of individual self-determination or autonomy. What exactly "autonomy" is, and what forms of it consequentialists should value, cannot easily be spelled out clearly, but a degree of liberal neutrality seems necessary to promote autonomy. Many individuals, if allowed to make their own life choices, make major mistakes. But the value of letting people make such mistakes, from a consequentialist

---

[12] Cf. e.g. MacIntyre's chapter "Liberalism Transformed into a Tradition" in *Whose Justice? Which Rationality?* or Sandel's *Liberalism and the Limits of Justice.*

perspective, is that experiencing the consequences of such mistakes constitutes part of controlling one's own destiny in an authentic way and also constitutes a valuable opportunity to learn life's lessons and to reach a more conscious, enlightened state. Life, it is sometimes said, can only be lived from the inside; without freedom to discover our own truth about what the good life involves, and to live by this truth, we cannot realize our *selves*, but become little more than robots controlled by others. In Ackerman's words,

> What could be better for a person than his own development of a plan of life that seems to him good? (368)

We are beings who, by nature, care about social living, but we also care about a sense of determining our own course of life, especially after we have been educated to a sufficient level to discover exciting opportunities for weaving the strands of our life into a creative pattern. Moreover, since the constituents of human well-being are so diverse, even the most intelligent attempts at large-scale government paternalism are likely to be far less successful in promoting individuals' well-being than educating our children to a level of responsible autonomy.

It seems to me that there is a great deal of truth in all the above claims, and that an appeal to the widely (but perhaps not universally) shared value of autonomy goes a long way to providing a rational basis for affirming a basically liberal social order.

§3 But now we confront again the question raised earlier: What is so special about self-determination to justify liberals in using state power to enforce favorable conditions for *this* value without thereby also justifying attempts to use state power to enforce non-liberal values? Defending liberal politics both in terms of neutrality among diverse ideas of the good life and in terms of promoting the non-neutral value of self-determination seems obviously inconsistent. MacIntyre, for example, charges liberalism with this inconsistency:

> Liberalism provides a distinctive conception of a just order....[T]he theory and practice of justice within such a polity are not neutral with respect to rival

and conflicting theories of the human good. Where they are in force they impose a particular conception of the good life, of practical reasoning, and of justice upon those who willingly or unwillingly accept the liberal procedures and the liberal terms of debate. The overriding good of liberalism is no more and no less than the continued sustenance of the liberal social and political order. Thus liberalism, while initially rejecting the claims of any overriding theory of the good, does in fact come to embody just such a theory. (345)

MacIntrye's argument seems to be that the liberal belief that society should allow the greatest possible variety of differing conceptions of the good *contradicts* the other liberal belief that society should not promote one particular conception of the good, because the former belief itself necessarily embodies one particular conception of the good. How can liberals respond? They would seem to face the following dilemma. Either they claim that liberal politics is designed to achieve substantive neutrality among the personal values that may plausibly make up individuals' diverse ideas of the good life or they freely admit that liberal politics is indeed designed to create the most favorable conditions for personal autonomy. In the former case, liberals would have to show convincingly that their politics is not biased in favor of the particular non-neutral value of autonomy. In the latter case, it would seem, liberals could no longer claim to be neutral among the personal values that make up individuals' diverse ideas of the good life, and cannot consistently deny the use of state power to non-liberals to attempt to enforce *their* values.

Larmore choses the first horn of this dilemma and defends the peculiar "neutrality" of liberal politics, in the sense that such politics constitute a retreat to the most "neutral ground" possible *and* necessary in the face of pluralism about the good. *Publicly*, Larmore maintains, liberals must refrain from appealing to the value of self-determination in their defense of liberal politics and stick with "neutrality" as the peculiar virtue which justifies a liberal social order. *Privately* they may of course indulge in justifications for liberal politics which appeal to self-determination (50f; 73-76).

But this response seems to me to involve precisely the sort of hypocrisy against which MacIntrye inveighs. Unlike Larmore, I think the only credible

response to the dilemma posed for liberals by MacIntyre is to choose the second horn and to admit freely that liberal politics is indeed designed to create the most favorable conditions for personal autonomy. But then, how could liberals consistently deny that the use of state power by liberals to enforce favorable conditions for autonomy justifies, by parity of reasoning, the use of state power by non-liberals to enforce *their* idea of the good? Here is what seems to me to be the honest and sensible answer. Liberal politics (at its best) tries to be as unbiased between people's personal values as any political system can possibly be. The liberal conception of justice, as Rawls puts it,

> does indeed, presuppose a theory of the good, but within wide limits this does not prejudge the choice of the sort of person that men want to be (1988: 260)....a just liberal society may have far more space than other social worlds, but it can never be without loss (1988: 266).

It is not, of course, possible for such a political order to be neutral between the values of autonomy and non-autonomy. Nevertheless, there is something sufficiently special about autonomy to entitle those who value autonomy to use state power to enforce favorable conditions for it, without thereby justifying the use of state power by those whose idea of the good does not include personal autonomy. Allowing non-liberals to enforce perfectionist values other than autonomy would deprive those who value autonomy of the opportunity to lead autonomous lives, whereas liberal politics does not deprive people who do *not* value autonomy of the opportunity to practise a non-autonomous way of life by joining, for example, whatever religious sect they please. So liberalism is indeed special or "neutral" in a way no other political creed can be.

There can be no question, however, that the dynamics of a liberal society will profoundly shape or reshape a person's self. Brian Barry makes this point well:

> ...a liberal must take his stand on the proposition that some ways of life, some types of character, are more admirable than others...Liberalism rests on a vision of life: a Faustian vision. It exalts self-expression, self-mastery and control over the environment, natural and social; the active pursuit of knowledge and the clash of ideas; the acceptance of personal responsibility for

the decisions that shape one's life. For those who cannot take the freedom it provides alcohol, tranquilizers, wrestling on the television, astrology, psychoanalysis, and so on, endlessly, but it cannot by its nature provide certain kinds of psychological security. (1973: 126-7).

The liberal self is very different from a communitarian or tribal self. If this is the main point of MacIntrye's criticism, it can be readily granted without undermining the above argument that liberalism is special in a sense which is *relevant* to justifying its monopoly on the use of state power.[13] It is not liberals, but non-liberals, who would seem to face the more serious dilemma. They claim that there are standards of the good life by which they can determine worthy ways of life with great confidence and precision. But which are those? Either non-liberals are willing to offer a fairly specific list of goals and values which, they claim, all human beings should be educated to pursue; or they offer a broad range of vague generalities. If they do the first, they will immediately encounter rival proposals from other non-liberals, with no compelling rationale for choosing among them.[14] If they do the second, liberals will argue that these are personal values which a liberal society readily allows or encourages, or political values which it *constrains* its members to embrace as well. For these reasons, liberal politics is sufficiently special to justify the use of state power to create a favorable climate for self-determination *without*, by parity of reasoning, legitimating the use of state power by non-liberals.

---

[13] The above argument is, however, *sufficient* to justify liberalism's monopoly on the use of state power only under reasonably favorable economic and social conditions. This is so for two reasons. Only under minimally favorable economic and social conditions can the means to the effective exercise of personal autonomy be made widely available, and only under such conditions of relative economic security and a high level of education is it reasonable to have confidence in the constructive use of this autonomy on the part of most individuals.

[14] Non-liberal communitarians may not be persuaded by this dilemma. Rather than claiming that some fairly specific list of goals and values constitutes the good of persons under all conditions, they will argue that the good of persons consists in living by the traditions of their community, whatever they happen to be, as long as they provide a sense of belonging. The communitarian claim in this general form seems implausible to me, because members of communities with strong ethnic identities often purchase the good of ethnic belonging at a tremendous cost in opportunities for human betterment, and because there is often a great deal of ethnic chauvinism or fanaticism about the way in which many such communities have shaped the identity of their members. This is not to say that the well-being argument in favor of liberalism necessarily justifies the sort of radical uprooting of indigenous groups that took place, for example, in the case of native Indian communities.

Few political distinctions go deeper, in theory and practice, than the distinction between a non-liberal and a basically liberal state - between a state which enforces someone's preferred idea of what makes a human life worthy, as opposed to a state which strives, in good faith, to give individuals wide scope to live as they prefer, in an atmosphere of free expression, meaningful opportunity and mutual tolerance. This is not to say that liberal practice always works out as it should, far from it. But the same can be said about any conception of the best social order.

3.5 THE CONFLICT BETWEEN LIBERTARIANS AND EGALITARIANS

This chapter considers the possibility of rational disagreement about the strongly conflicting conceptions of justice held by libertarian and egalitarian liberals.

3.5.1 The presuppositions of freedom

§1 All liberals prize the freedom of individuals to seek their own good in their own way. In commonsense terms, *individual* freedom is perhaps best spelled out as "not being prevented by others from doing what, given adequate information and deliberation, one desires to do and could actually do, unless prevented by others." A liberal conception of *social* freedom, by contrast, can be understood as a system of constraints on the exercise of individual freedom so as to avert conflict, while maintaining individual freedom to the greatest possible extent.

But what should those constraints be? There are bitter divisions among liberals about the answer to this question. Perhaps the deepest rift exists between two families of principles which can be usefully labelled "libertarian" and "egalitarian". In an attempt to make vivid the difference between them, I want to begin with two arguments which lead from the same liberal core idea via plausible premises to strongly incompatible conclusions. First, a basic libertarian argument:

1. The state should be as impartial as possible between diverse ideas of the good.

2. If the state should be as impartial as possible between diverse ideas of the good, then it may interfere with any individual's activities only when they cause harm to others, and not to impose someone else's idea of the good.

3. If an individual's activities make that individual better off without thereby making others worse off, then any interference with these activities would impose someone else's idea of the good rather than prevent harm to others.

Therefore, the state must not interfere with any of an individual's activities which make that individual better off without thereby making others worse off.

Compare that to a basic argument for what I shall call "egalitarian liberalism":

1. The state should be as impartial as possible between diverse ideas of the good.

2. If the state should be as impartial as possible between diverse ideas of the good, then it must strive to create at least minimally favorable conditions for everyone to pursue his or her own idea of the good (from within a wide range of meaningful options), regardless of whether or not a person was born wealthy, smart, healthy or otherwise fortunate.

3. If the state does not ensure at least a minimally adequate degree of universal access to a stock of goods which serve as important means to a successful pursuit of a wide range of ideas of the good life, then those who happen to be born less fortunate in their endowments will be able to pursue their idea of the good far less effectively as those born rich, smart, healthy or otherwise lucky.

Therefore, the state should ensure at least a minimally adequate degree of universal access to a stock of goods which serve as important means to a successful pursuit of a wide range of ideas of the good life (such as, for example, access to proper nutrition, education, medical care, child care, employment, housing, cultural events, physical and economic security, and others).

Both arguments can be suitably elaborated to acquire great rational appeal, and yet their conclusions seem to remain irreconcilable. I defined social freedom as *a system of constraints* on individual freedom which seeks to avert conflict while at the same time maintaining individual freedom to the greatest possible extent. The above arguments make clear why libertarians and egalitarians will be sharply divided over the nature of this system of constraints.

For libertarians, social freedom requires the state to enforce principles which guarantee non-interference with an especially important and fairly sharply delimitable subset of individual actions (namely those actions which make an agent better off without thereby making others worse off). For egalitarians, social freedom requires the state to enforce principles which promote desirable new freedoms or perhaps even the most freedom overall, even though that will mean frequent

interference with that subset of actions of each individual which the libertarian considers especially important. The egalitarian's idea of social freedom requires governments to create a complex system of duties of mutual assistance to provide a framework of substantive equality of opportunity to enable all individuals to pursue their goals effectively – an idea which the libertarian opposes because forcing some to render assistance onto others would interfere with some individuals' freedom to use their own minds and bodies as they choose while not thereby making others worse off.

Let me label the idea of freedom cherished by libertarians "*basic* freedom" and define it, in very preliminary way, as "any person's right to non-interference with any actions of that person which do not make others worse off." It would completely miss the point of the libertarian position to justify some constraint on an individual's basic freedom by arguing that some plausible index of aggregate freedoms as a result of interference will sometimes clearly exceed the index of aggregate freedoms that would have existed without such interference. Libertarians do not deny that (although they may well deny that any such index is possible even in principle). They deny that trade-offs between an individual's *basic* freedom and other values, including maximum freedom overall, are ever permitted.

Exactly what sorts of actions are those which make an individual better off without thereby making others worse off? And *why* should such actions be regarded as entailing an absolute or near-absolute right to non-interference? I explore these concerns below. But first I must address a pressing question. It may seem an all too obvious empirical fact that the libertarian right to non-interference cannot pass the well-being test required for rights by a consequentialist moral perspective. If libertarian ideas about social justice are grounded on little more than a bare assertion of natural rights, can we not exclude them offhand from the realm of rational disagreement about justice?

The answer is no. First, some libertarians, for example Hayek (and apparently Nozick, Narveson and Gauthier as well), *deny* that libertarian rights-claims would in fact fail every plausible version of the consequentialist's well-being test. They argue that the libertarian's favored social institution, the free market, has either fortuitously or necessarily the socially most beneficial effects of all known alternatives.[1] Second, there has been a recent, prominent and sustained attempt by Narveson to defend a *contractarian* form of libertarian liberalism whose practical implications (though not its justification) are claimed to be largely compatible with a utilitarian value perspective. If this striking claim is really defensible – if moral practices grounded in Gauthier's mutual-benefit individualism can plausibly promise greater social benefits than more egalitarian-liberal practices built on consequentialist justifications – , then of course such contractarianism could be reinterpreted as a form of indirect consequentialism.   Consequentialists would then have to regard mutual-benefit individualist methods of moral reasoning as justifiable on *consequentialist* grounds. They would have to support the dismantling of most redistributive social institutions along with the demise of practically any other institutions, values or relationships which cannot survive the "harsh justice" of free markets.

§2 Differences between libertarian and egalitarian conceptions of freedom produce vastly different conceptions of social justice. For liberals generally, the very point of social justice is to ensure the greatest equal liberty for all. For egalitarians liberals, the greatest equal liberty for all cashes out as guaranteeing everyone an adequate basic scheme of equal political rights and freedoms compatible with the same rights and freedoms for all, together with a more or less extensive list of social and economic rights and opportunities. For libertarian liberals, however, social

---

[1] Cf. Friedrich Hayek, *The Mirage of Social Justice* (1976); Robert Nozick, *Anarchy, State, and Utopia* (1974); Jan Narveson, *The Libertarian Idea* (1988); David Gauthier, *Morals by Agreement* (1986).

justice is a matter of respecting the right to non-interference with each other's *basic* freedom to use our individual minds and bodies, and the products of our minds and bodies, as each of us sees fit as long as he or she does not thereby make others worse off. This cashes out as, essentially, being able to exercise one's talents in a more or less *laissez faire* market and letting others do the same. This particular type of freedom is what Narveson has in mind with is striking claim that

> constraining the liberty of some in order that others may prosper (or live at all) is contrary to justice. (101)[2]

Libertarians typically regard social distributions as just if and only if there is a free market in goods which allows individuals to acquire, through voluntary transactions, whatever private property they desire, and to dispose of this property through whatever voluntary transactions they choose, without any regard whatever to such commonly invoked just-making considerations as need or equality. Egalitarians, on the other hand, consider social distributions to be just only if individuals enjoy substantive equality of opportunity in their ability to develop their natural endowments, to make effective use of political rights, and to acquire the most important means to well-being. The spectrum of egalitarianism is, however, very broad, ranging from those who want to ensure for every individual a minimally favorable starting point in life's competition to all those who want not only to eradicate all *socially caused* disadvantages which some now suffer, but as far as possible compensate for *natural* disadvantages as well. (Some may even want to regard all disadvantages as ultimately socially caused.)

To further clarify the depth of the division between the egalitarian and libertarian outlook, I want to distinguish between two types of social distributions: intended and unintended. Distributions may be intended, as when government enacts a social welfare policy, or unintended, as when distributions result

---

[2] Bare page numbers in brackets following quotes from Narveson and Gauthier refer to *The Libertarian Idea* (1988) and *Morals by Agreement* respectively (1986).

inadvertently from many people's uncoordinated decisions. Many distributive outcomes which were not intended by anyone's conscious decisions can, however, be modified by deliberate political choice. According to egalitarians, distributive outcomes resulting from people's uncoordinated market transactions are unjust if they put some groups at a systematic disadvantage on the basis of irrelevant characteristics, such as race, sex, ethnic origin, and others. But according to libertarians, if distributive outcomes resulting from people's uncoordinated market transactions put some groups at a systematic disadvantage, this may be bad luck for them, but is not unjust. Libertarians often reason that anything that is not consciously intended by someone merely happens, and what merely happens to a person cannot be unjust.[3]

The core intuition of egalitarians about justice is perhaps best expressed by saying that there is a presumption of social injustice if some people, or members of some group, end up significantly worse off than others through no fault of their own. The corresponding core intuition of libertarians about justice could be expressed by saying that there is no presumption of social injustice if some people, or members of some group, end up considerably worse off than others through no fault of those others. Egalitarians will say that even if the better-off are not directly or intentionally causing others to be worse off, such undeserved inequalities are nevertheless often correctable faults of the system. Libertarians will reply that these alleged faults of the system could only be corrected at the expense of ongoing, intolerable interferences with the outcome of people's creative use of their own body and mind and with their voluntary transactions.[4] In any case, egalitarians cannot agree among themselves on a "correct" standard of equality. They have far

---

[3] Cf. e.g. J.P.Day, *Liberty and Justice* (1987). For him, the fact that the offspring of the poor may have considerably less access to quality secondary and tertiary education, for example, and have therefore no chance at all to qualify for many of society's most desirable positions, is simply "bad luck, like being female. And, as we have seen, there is no right to compensation for misfortune."(94)

[4] Cf. Nozick's Wilt Chamberlain example, now the classic argument of this kind (1974: 161-163).

greater trouble than libertarians when it comes to agreeing, in concrete terms, about how much inequality social justice permits. No doubt, the promise of escaping chronic infighting among egalitarians on this issue is, for many, part of the allure of libertarianism.

§3 The libertarian idea, according to Narveson,

> is to maximize individual freedom by accounting each person's *person* as that person's own *property*, that is, by giving each person the maximal level of rights to the disposition of that particular bit of 'property' that is compatible with everyone else's having such rights. (175)

Libertarians will typically identify a wide range of activities as coming under this property right to one's own person, such as, for example, engaging in any "capitalist acts among consenting adults", bequeathing an inheritance to one's offspring, or refusing to hire particular individuals as employees of private firms for whatever reasons one pleases. Libertarians also identify a whole range of political practices as interference with the property right to one's own person, such as, for example, land use restrictions or building codes, compulsory universal health or unemployment insurance and pension schemes, and most forms of taxation, most notably the progressive income tax. I shall call the disputed right to non-interference in this distinctive sense "libertarian property rights".

When reasonable, informed people of good will are locked into bitter disagreement, as they certainly are over the extent of property rights, each side is likely to have a strong defense.[5] Although the dispute between libertarians and egalitarians is an example of a disagreement between informed people with impeccable rational credentials, we can nevertheless not simply conclude, I think, that it must therefore be a rational disagreement which currently available

---

[5] Of the master principles listed in Sec 3.2.1, Narveson and Nozick's principles exemplify the libertarian approach to social justice. Nielsen and Derek Phillips' principles exemplify the strongly egalitarian approach while Rawls and Sterba's principles exemplify a more moderate egalitarian approach.

arguments, information or methods have no promise of resolving. I obviously cannot discuss all the constructivist justifications which libertarians have offered in defense of libertarian property rights, but to establish at least a *very* tentative presumption about the scope for rational disagreement, I want to briefly consider two of the most prominent contemporary justifications: Narveson's mutual-benefit individualist defense and the more straightforwardly utilitarian defense.

3.5.2 The contractarian justification of libertarianism

§1 In *The Libertarian Idea*, Narveson proceeds from seemingly innocuous assumptions to libertarian property rights. I shall summarize what I take to be the structure of his argument, relying heavily (as does Narveson) on concepts pioneered by Gauthier.

Narveson begins his "central argument" with a carefully worded question:

> The question is whether I am motivated, by virtue of my possession of almost any set of values you can imagine, to acknowledge, when I deal with others (other more or less autonomous beings, that is) any principles restricting my freedom of action in any way, and if so, which. (175)

"The general form of the answer," Narveson continues, is

> that we shall all adopt principles restricting the performance of actions by others that would make ourselves worse off than we would be if we didn't have that restriction. We must, by definition, favor the nonperformance of such actions, since by hypothesis we are assessing worse-offness in terms of our *own* values, whatever they may be. (175)

Narveson does not specify any hypothetical position for contractarian bargaining. Doing so would, of course, immediately invite the charge of question-begging. But can he plausibly claim that he is referring to *any* conceivable situation whatever in which rational persons may find themselves - such as, most importantly, the here and now in the actual word? If "rational person" is understood to mean "a person who cares only about the maximal fulfillment of his own preferences", then it would certainly *seem* that, starting from our situation in the actual world, he could never make out a case that it must be rational for each of us, regardless of our

differences in power and values, to adopt identical principles restricting the performance of actions "by others" in certain ways. (After all, "we" are interchangeable, for the purposes of Narveson's argument, with those "others".)

Is there a hypothetical bargaining position which we - assuming we are rational persons (individual utility maximizers caring only about maximal fulfillment of their own preferences) - would accept? If it is not rational for each of us, given our present situation, to agree to a certain principle restricting our behavior in the actual word, how could it be rational for each of us, given our present situation, to accept a certain hypothetical bargaining position in which each of us *would have found* it rational to agree to such a principle? From the point of view of a rational person in the actual world, is there any interesting choice to be made between agreeing to a certain principle and accepting a hypothetical position in which he would have agreed to this principle? If someone wouldn't rationally do the first, why would she do the second?

The answer, of course, is that she would not. Each of us would find it rational to accept some hypothetical position only if the principles she agrees to in this position would also be rational for her to agree to in the actual world. Nevertheless, postulating a hypothetical position may be a useful expository device. It may model our rationale for adopting principles in a vivid enough way to actually suggest promising principles to us. As long as we can make out a case that a certain principle which each of us*would* agree to in a hypothetical position is such that it is also rational for each of us to agree to it in the actual world, we have begged no question.

§2 I shall now address four questions in the following order: (1) What sort of hypothetical position should we postulate for expository purposes? (2) What principle would rational persons agree to in this position? (3) Is it indeed rational for us in the actual world to agree to this principle as well? And (4) Is it rational for us in the actual world to comply with this principle - or if not, is it perhaps still

rational for us to do something about the fact that it is not rational in the actual world to comply with this principle?

(1) Which hypothetical position should we assume for expository purposes? The answer to *that* question is strongly suggested by Narveson's "general form of the answer" to the carefully worded question which he wants rational persons to ask themselves (see quotes above). It seems to be an obvious analytic truth that to determine whether I have become worse off *as a result of actions by some specific other*, I must determine how well off I *would have been* had that particular person never acted in such a way as to affect me at all. I am supposed to imagine a state prior to being affected, positively or negatively, by any actions of Y. And similarly, to determine whether I have become worse off *as a result of actions by some person or other or other people generally*, it would seem that I must determine how well off I *would have been* had other people never acted in such a way as to affect me at all. This situation in which I imagine myself to be when I imagine that other people never acted in such a way as to affect me at all will be called "the presocial position" (cf. Gauthier, 1991:25).

(2) What principle would rational persons agree to in the presocial position? Let us imagine rational persons in a presocial position preparing to negotiate the fundamental principle of a social contract: would it be rational for each of them to accept some principle restricting the behavior of each? The answer to that is obviously yes. Although *ex hypothesis* we imagine them not to have affected each other in any way yet, the following developments can be plausibly foreseen in the absence of any restrictions whatever on each other's behavior. Certain general facts about the human condition - scarcity of goods relative to people's desires for them, sufficient equality of power so that usually even the relatively weak can inflict harm on the relatively strong, and greater concern among most individuals for satisfying their own desires than those of others - will predictably tempt some individuals to take goods away from others. This will lead rapidly lead to high levels of mutual

resentment and general mistrust along with preemptive defense measures on the part of many individuals who want to protect their lives and possessions. This situation can plausibly be imagined to escalate into Hobbes' famous "war of every man against every man" which would make practically everyone's life "solitary, poor, nasty, brutish and short."

Rational people clearly would not want to lapse into a Hobbesian state of nature. But neither would rational people in the presocial position simply accept any principle restricting each other's behavior that will merely stave off the Hobbesian state of nature – perhaps making their lives a little bit less poor, nasty, brutish and short. Obviously, for *that* purpose a great many principles would do. Rational people want to do better than that. I take my clue from Narveson when he says,

> "Each rational agents wants, after all, to get the best possible deal, and not just to beat the state of nature."(155).

Although there is a sense in which it is clearly rational for each of us to accept any principle averting the state of nature to any principle which would not avert it, we are looking for something better: one particular principle which can be shown to be uniquely in the rational self-interest of *both* the relatively weak *and* the strong among us to agree to in the presocial position.

Which principle is that? Is there perhaps a principle such that a rational person would not voluntarily agree to anything less favorable to himself – while at the same time realizing that other rational persons would not voluntarily agree to anything more favorable to him? Well, take the principle permitting everyone to better his situation in life while not permitting others to worsen it at any time whatever. Let us call this principle "constraint X". Let us consider now whether, in plausible circumstances, both the relatively powerful and the relatively powerless could be expected to voluntarily agree on any constraint that is either stronger or weaker than constraint X. (By "plausible circumstances" we do not mean to import controversial assumptions like equality of power, which obviously do not hold in

the actual world. We are only assuming three things which do very often hold: that circumstances obtain in which a cooperative surplus is possible which can then be split to the mutual advantage of both the weak and the strong; that each person is "rational", in the sense of caring only about the fulfillment of his own preferences; and that no rational person would *willingly* accept a principle restricting each person's behavior that is clearly less advantageous to him than to the others.)

In these circumstances, then, would the relatively powerful and the relatively powerless agree on any constraint either stronger or weaker (from the point of view of any given individual) than constraint X? Stronger constraints than constraint X (i.e. more restrictive from the point of view of any given individual) would be, for example, a constraint permitting a given individual to better his situation *only if* he thereby also betters the situation of others; or a constraint permitting a given individual to better his situation while permitting others to worsen it. It seems quite plausible that stronger constraints than constraint X will be unattractive, both to the powerful and the powerless. After all, we are beings who want to be free to better ourselves. A principle that allows a rational person to better his situation while not allowing others to worsen it would seem to be clearly preferred over principles that allow him to better his situation while also allowing others to worsen it, or to better his situation only if he also betters the situation of others.

Very well, but what about weaker constraints than constraint X? A weaker constraint (i.e. less restrictive from the point of view of each individual) would be a constraint that permits that individual to better his situation and does not permit others to worsen it, but permits him to worsen the situation of others. Wouldn't *this* constraint seem extremely attractive from the point of view of a given individual?

Not so, as a little reflection shows. Every rational person who knows that others are also rational already knows beforehand that such a constraint would not be acceptable to them, and would therefore give rise to "costly interaction".

As Gauthier puts this point,

> Rational persons will voluntarily accept an agreement only insofar as they perceive it to be equally advantageous to each. To be sure, each would be happy to accept an agreement more advantageous to herself than to her fellows, but no one will accept an agreement perceived to be less advantageous. Agents whose rationality is a matter of common knowledge will recognize the futility of aiming at or holding out for more, and minimize their bargaining costs by coordinating at the point of equal advantage. (1991: 28; cf. also 1986: 227)

This crucial - and perhaps initially a bit counterintuitive - step in the argument is worth belaboring for a moment. It is clear why the relatively powerless would not demand a weaker constraint than constraint X on their own behavior: they know beforehand that it will not be acceptable to the relatively powerful and would give rise to conflict – in which the powerless would by definition lose out. But why would the relatively powerful hesitate to demand a weaker constraint than constraint X on their behavior vis-a-vis the relatively powerless? Isn't it commonsense that it would be rational for the powerful – fulfill more of their preferences – if they reserved the right to use the powerless any way they feel like, much the same way in which many people currently use animals? Why should the powerful settle for a constraint that doesn't permit them to better their own situation at the expense of the powerless, when they could clearly get away with it? As Narveson explains,

> Would the ones doing better voluntarily move to a situation in which their advantages were negated? In fact, the suggested answer is that they will do so, provided their situation after the agreement is better than it was before the agreement, *even if* the agreement negates their *relative* advantage over the weak. (177)

So there are highly persuasive reasons why, in plausible circumstances, the powerful might benefit far more from voluntary cooperation rather than from the oppression, enslavement or outright extermination of the powerless. But if no stronger nor weaker constraint than constraint X will be voluntarily accepted by rational persons, we would seem to have achieved a significant result: that rational

persons would converge on constraint X in plausible circumstances *in a presocial position*.

(3) *But is it rational for us to agree to constraint X in the actual world as well?* We said that it is rational for us to accept the presocial position *if and only if* the principles agreed to in this situation would also be agreed to by rational persons in the actual world. But why would it be rational for both relatively powerful and powerless persons, in the actual world, to agree to a principle constraining their behavior in such a way that they can better their own situation, but may never worsen the situation of others?

It would clearly be rational for them to do so if each could be sure of everyone's general compliance with this principle, and if general compliance with this principle would make greater cooperative surpluses possible in the actual word than are being realized at the present, and if each could be sure that he would not be left out when it comes to splitting this cooperative surplus. But these would seem to be three incredibly tough, counterfactual conditions to meet.

First, how could each rational person be sure of everyone's general compliance with constraint X? Making the plausible assumption that a rational person Y knows that other people are also rational, Y also knows that everyone would generally comply if each thought that doing so would make a cooperative surplus possible which would be split so as to benefit him relative to his present situation. For reasons to be spelled out below, it is extremely plausible that everyone's compliance with this principle would, in fact, make a cooperative surplus possible which could then be split to everyone's benefit. Now if everyone could be made to understand the reasons why it is extremely plausible that everyone's compliance with this principle would make a cooperative surplus possible which could be split to everyone's benefit, wouldn't *that* give each rational person in the actual world sufficient incentive to comply?

Unfortunately, no. In the actual world, a rational person cannot expect general compliance with constraint X, but at most feel more or less confident that, at least in many cases, other people will comply with *whatever principles reflect the actual moral beliefs prevailing in a given society.* If so, it is generally rational to comply *not* with the principle that everyone may better his own situation and no one may worsen it, but with whatever principles one is expected to follow in that society, on pain of moral condemnation, legal sanctions and various other forms of social ostracism. But this prevailing state of affairs is deeply unsatisfactory from a rational point of view, for at least five main reasons.

(a) First of all, it is unsatisfactory because different people in a given society tend to have conflicting moral beliefs which lead to social conflict and costs in terms of forgone cooperative surpluses.

(b) Second, it is unsatisfactory because the diversity of moral beliefs and the often transparently superstitious justifications given for them can only promote widespread moral skepticism among intelligent people. Rational persons will be motivated to violate moral principles in surreptitious ways, whenever they judge doing so to their personal benefit.

(c) Third, it is unsatisfactory because the system of sanctions necessary to enforce compliance with whatever moral beliefs prevail in a given society can be extremely costly to maintain (costly in terms of foregone cooperative surpluses).

(d) Fourth, it is unsatisfactory because *even if* all people agreed on the same moral beliefs, *and* each of them voluntarily displayed perfect compliance, the intrinsic nature of these principles would still be highly unlikely to lead anywhere near to an optimal allocation of social resources to their most productive uses. Universal compliance with the moral code of most societies would still be highly unlikely to produce the types and division of cooperative surpluses which would maximally fulfill the preferences of each or even of most individuals.

(e) Fifth, it is unsatisfactory, because to the extent that rational persons realize this last point, their realization will tend to reinforce their unwillingness to support the existing social order and will thus promote further social instability and foregone cooperative surpluses, over and beyond the instability produced by factors outlined under (a) and (b) above.

Nevertheless, unsatisfactory as this state of affairs is, the conclusion stands: in the present situation, it is generally not rational for a given individual (restricted economic contexts now governed by free markets are an exception) to comply with constraint X, rather than to make complex judgments, in the light of a host of considerations, to what extent compliance with prevailing moral beliefs may be to his benefit. But does that fact not deal a fatal blow to Narveson's and Gauthier's project? The answer, surprisingly, is no. From the fact that it is not generally rational to comply with constraint X in the actual world, it does not follow that it may not be rational, for many individuals, to do something about this very fact.

(4) *Is it rational for us to do something about the fact that it is not rational for most of us in the actual world to consistently comply with constraint X?* And to this question the answer would seem to be clearly yes for most rational people (if not necessarily for multimillionaires). If, as argued above, the present situation in the actual world indeed involves tremendous forgone opportunities for producing cooperative surpluses, then it would clearly be rational for us to move to a social order that *is* fundamentally structured so as to reliably produce cooperative surpluses as a result of practically every social interaction – one in which observing constraint X wouldn't only be actually (but unbeknownst to most people) rational, but would also be *known* to be rational by confirming its own rationality in transparent ways in almost every social interaction. But that would require that social interaction would have to consist almost entirely of transactions voluntarily engaged in for mutual benefit; in short, free market transactions. There are two important points to note about a social order permeated at every level by free

markets. (1) In such a market order, constraint X would be almost universally self-enforcing, because people wouldn't voluntarily choose to keep engaging in transactions with others (at least not more than once or twice) who worsen their situation. Because constraint X would be observed almost universally, this social order would be stable in way in which no order can be which routinely allows some people to make others worse off. And (2), in a market order, decisions about allocating resources are made in a decentralized fashion, by each individual using his or her special knowledge of opportunities for mutually beneficial cooperation with particular others. This sort of special knowledge is necessarily the property of individual persons. As Hayek never tires to remind us, no single mind can possibly encompass all this special knowledge of all individuals and thus plan the allocation of resources in a centralized way. Therefore, in a market order, individuals have incentive to use their special knowledge of opportunities for mutually beneficial cooperation so that resources will gravitate to their most productive uses – that is, to maximum fulfillment of each individual's preferences.

But if a market order embodies these two features – the greatest possible stability of any conceivable social order *and* a strong tendency for resources to gravitate toward their most productive uses – then it becomes extremely plausible to think that this order makes far larger cooperative surpluses possible than any other social order. Now as soon as rational people realize this fact, the existing social order will become even more unstable than it already is. In most situations, it will become rational for individuals to provisionally comply with prevailing moral principles while at the same time working toward a radical change in the existing social order. This will be so for a given individual when the expected costs to him of working toward such change are outweighed by the expected benefits. And, at least in democratic states which honor freedom of expression and association, the expected costs may be quite small. In dictatorships (such as the former Soviet Union), the risks of working toward a radical change in the social order are, of course, vastly

greater, but since the anticipated payoffs may also be vastly greater, it may well become rational for many individuals, even in dictatorial states, to work toward change by intelligent risk-minimizing means.

This, in brief, is what I understand to be the basic structure of Gauthier's and Narveson's contractarian argument for a libertarian social order.

§3 The force, coherence and sheer intellectual beauty of the mutual-benefit individualist's arguments have certainly not eluded me. Their tight coherence as a self-contained conceptual system makes it tempting, however, to overlook some serious limitations.

Before I express my worries about these limitations, I must stress three fundamental respects in which I take contractarians like Gauthier to be fellow travellers, and in fact intellectual leaders, toward goals shared by consequentialist constructivists. Both consequentialist constructivists and mutual-benefit individualists regard "morality" as a humanly contrived system of constraints on behavior which is worth having because such a system, at least if intelligently designed and sufficiently internalized by most people, will vastly improve almost everyone's chances for a good life by overcoming coordination and freeloading problems (and by offering intrinsic emotional satisfactions to many). Both consider the ultimate goal of a constructivist morality to be essentially the maximal satisfaction of people's informed preferences, although they disagree on the means to promote this goal most effectively. And both will want to employ contractarian reasoning or game-theoretic modelling of moral decisions in an attempt conceptualize moral issues and to identify "invisible-hand" processes which may achieve far better outcomes than other methods of promoting consequentialist goals (although here again they disagree on the extent of their reliance on such processes).

But eventually consequentialists and contractarians like Gauthier must part company. While Gauthier's work has inspired a number of promising research

projects (c.f. e.g. Danielson's work), I now want to state some fundamental worries about his mutual-benefit individualism as well as the radical libertarianism Narveson wants to base on it.

(1) Even if fully rational persons would agree to a principle which allows them to better their situation while not permitting others to worsen it, it is far from clear that this abstract principle can simply be mapped onto real life by the idea of all-pervasive markets replacing political decision-making or morally cemented community practices. In theory it might seem that a social world permeated by markets at every level of cooperative interaction – instead of such things as redistributive social programs – is a world where there is optimum resource allocation, no coercion, no need for a clumsy and corrupt bureaucracy, utter bliss, as close as we can humanly come to creating heaven on earth. But our empirical knowledge of markets must surely give us mixed feelings. Markets have their place. But they also have their own coercive dynamics, their horrendous externalities, their monopolistic momentum, their distinctive failures. So when Narveson claims that Gauthier's arguments imply the libertarian principle of justice, which says (among other normatively potent things), that

> We [i.e. via our government] have no fundamental general duty to provide others with such goods as the necessities of life, let alone some particular proportion of all the socially distributable goods there are;.... (166)

I cannot help wondering how many of us would find ourselves falling by the wayside if this principle were implemented. (It is only fair to point out that Gauthier himself is far more cautious about the concrete normative implications of his work.)

Nozick once saw markets as a panacea too and, in fact, saw utopia as a veritable orgy of markets, including markets in types of communities which individuals would freely join or leave according to their preferences (1974: 312f). I found it most sobering to reflect on what Nozick has to say fifteen years later:

"The libertarian position I once propounded now seems to me seriously inadequate, in part because it did not fully knit the humane considerations and joint cooperative activities it left room for more closely into its fabric. It neglected the symbolic importance of an official political concern with issues or problems, as a way of marking their importance or urgency, and hence of expressing, intensifying, channeling, encouraging, and validating our private actions and concerns toward them. Joint goals that the government ignores completely - it is different with private or family goals - tend to appear unworthy of our joint attention and hence to receive little. *There are some things we choose to do together through government in solemn marking of our human solidarity, served by the fact that we do them together in this official fashion and often also by the content of the action itself.*...A concern for the expression and symbolization of values that can best and most pointedly, not to mention most efficiently, be expressed jointly and officially - that is, politically - is continuous with a concern for individual self-expression. There are many sides of ourselves that seek symbolic self-expression, and even if the personal side were to be given priority, there is no reason to grant it sole sway. If symbolically expressing something is a way of intensifying its reality, we will not want to truncate the political realm so as to truncate the reality of our social solidarity and humane concern for others. ...The libertarian view looked solely at the purpose of government, not at its meaning; hence, it took an unduly narrow view of purpose too....If a democratic majority desires to jointly and symbolically express its most solemn ties of concern and solidarity, the minority who prefer differently will have to participate sufficiently to be spoken for." (1989: 286-290; emphasis added)

Obviously, the scope and limits for markets in the best social order is an immensely complex topic. I shall briefly consider some of the pros and cons of markets below, when I discuss the utilitarian defense of libertarianism.

(2) But if the connection between Narveson's extreme libertarianism and Gauthier's contractarianism is severed, it becomes deeply problematic what exactly mutual-benefit individualist arguments prove for disputes about justice in the real world. Gauthier certainly seems to have made a case that rational people in plausible circumstances might converge on the principle that each should be permitted to better his own situation while no one else is permitted to worsen it. But how precisely can this be operationalized in the context of contemporary real-world applications?

Gauthier maintains that the intuitive and ambiguous principle must be spelled out in terms of what he calls "the Lockean proviso":

...just as I better your situation in so far as you prefer the outcome of interaction with me to what you would have expected otherwise, in my absence, so I better my own in so far as I prefer the outcome of interaction with you to what I should have expected otherwise, in your absence or unavailability for interaction. And I worsen my own situation in interaction with you in so far as I prefer what I should have expected in your absence to the outcome that I actually bring about...We interpret the Lockean proviso so that it prohibits worsening the situation of another person, except to avoid worsening one's own *through interaction* with that person. (205; emphasis added)

What exactly counts as "interaction" in this context? Let me unpack Gauthier's crucial idea. We interact with someone only if the other person's presence causally affects the process of bettering ourselves in any way. If, in the other person's absence, this process *would* have unfolded in exactly the same way, no interaction in the relevant sense *actually* took place (209). But this account will cause alarm, and the more so after reading Gauthier's own provocative example:

Suppose that we live as fisherfolk along the banks of a river....But if you, living upstream from me, merely use the river for the disposal of your wastes, then even though you thereby kill many of the fish in my part of the stream, you do not violate the proviso. For although you worsen my situation in relation to what I should expect in your absence, you do not better your own situation through interaction with me. You are no better off than you would be were no one to live downstream from you. (211-2)

So if I own a factory upstream from you and better myself by polluting the river and killing all the fish on which you depend for your livelihood, I have not interacted with you because even if you had never existed, this would not in any way have causally affected the process by which I better myself. Since no interaction took place between you and me, I have not, in this case, made myself better off by making you worse off *through interaction*. Therefore, Gauthier concludes, the Lockean proviso has been fully met.

Any acceptable formulation of the Lockean proviso must contain the idea that in bettering oneself one must not make others worse off or take advantage of them. But I see no good reason why we should accept Gauthier's particular account of "interaction" as a necessary feature of the commonsense idea of not taking

advantage or not making others worse off. In fact, Gauthier's own example is a paradigm case of one person's taking advantage of her position to exploit resources in a way which makes someone else worse off, even though the causal interaction goes all one way. It would simply seem to beg the question in favor of libertarianism to spell out "taking advantage" in the particular causal way Gauthier proposes. If the Lockean proviso is interpreted this way, then it would not have been the principle that every individual utility maximizer who wants to form a society would already be prepared to accept as her or she approaches the bargaining table.

If the Lockean proviso is interpreted so that it clearly *is* to everyone's advantage to accept, it will only allow a person to better himself in ways that do not have negative effects, however indirect, on others, in which case its scope of application will be rather narrowly restricted. But, Gauthier may reply, surely the Lockean proviso must mean *something more.* How should it be interpreted, if not the way he proposes?

If the Lockean proviso were interpreted as allowing a person to better himself only in ways that do not have *any* (non-trivial) negative effects on others, then it would become unrealistically restrictive because, in practice, individuals' attempts at bettering themselves almost always causally impact on the prospects of others in many subtle and not-so-subtle ways. Any relative or absolute increase in the wealth of some will, for example, worsen the situation of others with respect to inherently scarce goods in ways that are certainly not trivial. No one lives in a vacuum; and so the Lockean proviso is almost never uncontroversially satisfied. Within hard-to-define limits, bettering oneself, even if it is sometimes at the expense of indirectly or marginally worsening the situation of others, must under some conditions be perfectly compatible with social freedom and justice. In practice, a clean causal separation between some people's wealth-acquiring activities and other people's poverty-stricken situation is rarely possible, even though the causal connections are often far too complex for anyone to trace exactly. But this opens up all the questions

surrounding Mill's harm principle, a principle of which the Lockean proviso is clearly only a variant. The question arises what, if anything, we are actually free to do if we are only free to do what will not harm others. The best we can do in approaching this question, it seems to me, is to weigh the consequentialist pros and cons of different types of actions or policies with as much foresight as we can humanly muster, making intelligent judgments about which action-types should qualify for consequentialist rights-status while recognizing that what does and does not harm, in a way serious enough to be a concern of justice, will change with ever-changing circumstances. Certainly, given the complex causal web which interconnects the actions of individuals in a society, Mill's harm principle is far too simplistic by itself to adjudicate between permissible and impermissible types of individual freedom or permissible and impermissible types of egalitarianism. I cannot address these complexities here. What I have argued is simply that the Lockean proviso does nothing whatever to settle the dispute between egalitarians and libertarians in favor of libertarianism. If the Lockean proviso is interpreted so as to be in everyone's self-interest to accept prior to bargaining, it does not have libertarian implications. If it is interpreted so as to have libertarian implications, it is not in everyone's self-interest to accept prior to bargaining.

(3) Gauthier goes on to give a mathematically precise interpretation of what it means for co-operators be be constrained by his version of the Lockean proviso. According to Gauthier, a division of the cooperative surplus among fully rational cooperators requires them to determine, for each cooperative interaction, what sort of utility outcomes would have resulted in the complete absence of both cooperative and non-cooperative interaction and then to minimize the "maximum relative concession". But the counterfactual information required to establish the non-agreement point in typical real-life situations in which disputes about justice arise seems elusive. Gauthier seems to me to abstract from the particulars of real-life situations to such an extent, and would require such enormous informational

input, as to leave completely unclear how the relative concessions required of the different parties should be calculated. In fact, many of those conversant with bargaining theory find it unclear why Gauthier's minimax relative concession principle should be regarded as authoritatively defining the agreement which rational persons must reach when they bargain over the division of a cooperative surplus. This point would, of course, require detailed technical discussion.[6] But it is hard to see what follows from the mathematical precision of Gauthier's approach for real-life moral decisions. The fact is that most real-life situations in which disputes about justice arise seem to involve far too complex variables for ex ante generalizations about the most rational outcome of such bargaining among self-interested utility maximizers.

(4) When criticizing alternative moral theories, Gauthier seems to promise us a *robust* mutual-benefit individualism but in the end, I think, he delivers only an *abstract* one (while equivocating between the two). Let me make this important distinction clear. Either mutual-benefit individualists conceive of rational self-interests *abstractly*, in terms of the maximum gain to be realized by each member of a society of mutually disinterested utility maximizers who are highly successful in avoiding costly lapses into non-cooperation due Prisoner's Dilemmas (freeloading) and Assurance (coordination) Problems. Or mutual-benefit individualists conceive of individual interests *robustly*, in terms of what sorts of interests mutually disinterested utility maximizers, given the epistemic limitations of ordinary human beings, will plausibly find motivationally potent for themselves.

In the case of "abstract" mutual-benefit individualism, I cannot see what, in specific distributive contexts, it actually amounts to, because it would seem to

---

[6] This objection to Gauthier's contractarian libertarianism is most lucidly elaborated by Brian Barry (1989). Gauthier's theory presupposes the possibility of a general solution to real-life bargaining situations to give principled guidance to decision-makers. Barry argues persuasively that there are many possible types of "splitting-the-difference" solutions, none of them having any rational advantage over others (cf. 388-92). Others also make a forceful case that Gauthier's proposed solution is, at any rate, faulty (cf. e.g. Danielson, in Vallentyne, 1991; or Gibbard, 1991).

require an uncanny form of prescience on the part of individuals. While relatively simple and highly artificial situations are conceivable where Gauthier's strategy may indeed work to maximize everyone's preference satisfaction, how can this strategy be applied to the dynamic rough-and-tumble of interaction in a complex society? I can make no sense of the idea of keeping track of counterfactual Prisoner's Dilemmas which have been avoided, and keeping track of the actual payoffs which have been realized as a result, thanks to compliance with moral constraints on the part of a multitude of identifiable individuals. But if the contribution of specific individuals to the avoidance of (counterfactual) Prisoner's dilemmas cannot be credited, then how is the cooperative surplus gained from such avoidance to be calculated and its division arranged? And why should those who are relatively poorly off at an earlier point in time expect that giving the already well-off a disproportionate share of any cooperative surplus will serve as a special incentive for the well-off to cooperate in ways which will make the relatively poor better off in the long run than they would have been in the absence of such a skewed division? It would seem clear that the "abstract" mutual-benefit individualist can offer a self-interested utility-maximizer neither specific normative guidance about how to maximize his rational self-interest nor, *a fortiori*, motivationally potent reasons for not maximizing his self-interest by worsening the situation of others.

"Robust" mutual-benefit individualism, on the other hand, would seem to sanction the sort of predatory actions which will vitiate it as a plausible candidate for carrying out morality's main job. According to this theory, only when the relatively strong and well-off judge, according to their own epistemic lights, that they have more to gain from cooperating with the weak and poorly-off rather than excluding, oppressing or even exterminating them are they rationally (and thus morally) required to cooperate. Robust mutual-benefit individualism would seem to rationally require individual utility maximizers to cultivate dispositions committing them to a mixed strategy of intelligently combining constrained (but not

too constrained) maximization with straightforward maximization whenever a major scoop can be made with impunity. Since morality, for Gauthier, arises only between parties whose cooperation will lead to a cooperative surplus from which both benefit, he cannot consistently say that there is anything whatever morally wrong whenever those with vastly superior technology have nothing to gain from cooperation with their technological inferiors and brutally exterminate them: in fact, Gauthier explicitly affirms the "rationality" (and therefore morality) of doing so (231-2). The "robust" mutual-benefit individualist has a morality which simply isn't conducive to creating the sort of mutual trust which would seem to provide the right climate for optimum social cooperation.

"Abstract" mutual-benefit individualism may seem to save individual utility-maximizers the price of straightforward maximization due to nearly ubiquitous Prisoner's Dilemmas in social interaction. But it presupposes a completely unrealistic prescience on the part of self-interested individuals of how specific Prisoner's Dilemmas will arise in the rough-and-tumble of social interaction and therefore fails to be motivating for actual individuals. "Robust" mutual-benefit individualism, on the other hand, succeeds in motivating actual individuals to act according to their best judgment of their own self-interest, but only at the price of allowing collective outcomes to reflect the price of straightforward maximization in many Prisoner's Dilemma situations.

While morality is indeed justifiable only as a cooperative scheme for mutual advantage, I do not see how the mutual-benefit individualist can, in the end, launch a scheme of social cooperation without also having to engage in cost-benefit calculations which separate individuals' (non-measurable) contributions to the success of any such scheme from the share of benefits each individual gets out of the total cooperative surplus resulting from this scheme. There is no credible way in which the avoidance of non-cooperative behavior in potential Prisoner's and Assurance Dilemma situations could get systematically credited to specific

individuals, and the cooperative payoff as a result of such avoidance could get apportioned among particular cooperators according to how well off each one of them had been if *neither cooperative nor non-cooperative* interaction had taken place. But this means two things. It means giving up on the idea that contractarianism could somehow ensure moral compliance by harnessing individual utility-maximizers' self-interest and could thus make up for deficits in individuals' moral motivations *without* external sanctions. And it also means that there is really no way around ranking the desirability of alternative systems of distributing the payoffs from social cooperation according to whether they maximize some plausible aggregate satisfaction of individuals' interests.

3.5.3 Limits to rational disagreement about libertarianism

§1 In the debate between egalitarian and libertarian liberals, the key issue - for consequentialists - is whether almost exclusive reliance on the market for social and political decision-making will, ultimately, meet the consequentialist well-being test better than an attempt to justify such decisions by an appeal to a hierarchy of moral principles. The core thesis defended by such *utilitarian* libertarians as, for example, Hayek, Friedman or Hospers, is captured well by N.P. Barry:

> [T]he co-ordinating mechanisms of the market and the incentives structure provided by the institutions of the private property system always produce, accidentally, more beneficial outcomes than all known alternatives. (20)

In other words, *utilitarian* libertarians defend capitalism, the widely dispersed ownership of the means of economic production under competitive market conditions. Contractarian libertarians do so as well, but they seek to first justify the right of individuals to use their own minds and bodies as they choose as long as they do not thereby make others worse off, and then defend capitalist economic arrangements as flowing from this right. Utilitarian libertarians, in contrast, defend capitalist economic arrangements by a direct appeal to their beneficial consequences as compared to any alternative economic arrangements. They argue that capitalism,

because it tends to reward work with socially beneficial effects, provides maximum incentives for such work and therefore maximizes socially beneficial effects. It makes sense to equate justice with market outcomes, they argue, both because ultimately only markets can solve large-scale commensuration problems, and also because rewarding creative or conscientious work rewards desert in some plausible form. When such libertarians argue that the harsh "justice of the market" should be left undisturbed, they imply, in effect, that capitalism, or at least some essentially capitalist form of economic arrangements, has the potential to capture enough of the pattern of people's expectations and sentiment about justice to become acceptable to them as just arrangements and to enlist their sense of justice in support of these arrangements.

How plausible is this thesis? I cannot delve into the details of the utilitarian arguments offered by Narveson, Hayek and other libertarians. Nevertheless, I cannot neglect their claims entirely, because if they were true, then a consequentialist moral framework would require us to equate social justice largely or even exclusively with free market outcomes and whatever political arrangements are required to uphold the freedom of the market. To cast some doubt on the claims of utilitarian libertarians, I can only suggest here the complex nature of the pros and cons that would have to be weighed to decide which precise economic arrangement will best promote well-being. But it is fairly clear that leaving all distributive matters up to real-life (as opposed to highly idealized) markets will not achieve this goal.

Capitalism has many consequentialist advantages. Countless economic decisions are coordinated without any need for an expensive, corruptible central bureaucracy. It saves us from having to balance the demands of conflicting just-making criteria or to commensurate the value of different goods for different people. When all goes well, it provides incentives to creativity, risk-taking and socially beneficial effort of many kinds, and generally there is at least a very rough,

partial correlation between people's contributions and rewards. Capitalism harnesses some of the the strongest human motives, notably greed and fear, in the service of economic goals and thereby liberates staggering productive energies. In many countries, capitalism (accompanied by social programs) has proven compatible with a high degree of political freedom combined with relatively widespread prosperity.

But capitalism seems clearly guilty of major evils as well. It is true that capitalist markets spare us complex balancing problems, but only at the expense of making *effective preference* the sole criterion of who gets what - which means any preference backed up by money. Capitalism tends to overproduce luxury items for the wealthy, often without providing some of life's essentials for others. It creates chronic tensions as extreme wealth and poverty exist side by side. It has led to the predation of third-world countries by multinational corporations. There is the overemphasis on often rather joyless material consumption together with waste and discontent resulting from artificially created desires. There is the concentration of economic power which frequently translates into political power and privilege. There are the alienating conditions of work for many, the reinforcement of some of the worst traits of human personality and the erosion of community bonds. There is the emotional upheaval and pervasive climate of insecurity associated with the relentless pace of change under capitalism. There is reason to fear that the same productive energies whose liberation makes capitalism seem so beneficial may, before long, overwhelm the earth's ecosystem.

Obviously, all these points would need extensive empirical support in order to allow us to decide on the best economic system from a consequentialist point of view. But it is surely clear that the choice of the best economic system is far more complex than many utilitarian libertarians would have us believe. Economic experience has not shown, despite Hayek's slippery-slope "road to serfdom" arguments, that judicious redistributive taxation undermines people's incentive to

risk their capital in pursuit of wealth. Even if, at some point, taxation undermines the incentive to invest and stifles the initiative of the enterprising to work hard, it can also quite plausibly be argued, purely on grounds of economic efficiency, that a great deal of inequality gives a massive headstart to the offspring of the well-to-do while demoralizing the downtrodden and depriving them of both motivation and opportunity to contribute their best to society. And there are other complex problems faced by the utilitarian libertarian's defense of capitalism. Keeping competitive markets intact over time is not simply a matter of allowing capitalist incentives to work their "invisible hand" magic, but requires a powerful enough state apparatus to prevent price collusions, monopolies and environmental spillover effects, to promote wide dissemination of accurate product information, and to maintain an infrastructure to keep transaction costs low. Large corporations may amass such wealth and power that it becomes often difficult for government officials to preserve their integrity, rather than promoting the interests of capitalists in exchange for various sorts of direct and indirect forms of support. I will not pursue these difficult issues further here. It is a fixed point among my considered judgments, however, that something far short of libertarian property rights will provide the optimum consequentialist mix of private and public control of the means to production and the profits flowing from them.

§2 Although the case for libertarian liberalism strikes me us ultimately not persuasive, this is not to deny the importance of the libertarian's cherished freedom to claim ownership over one's own body and mind and to exercise whatever talents or effort one can marshal in a market. There seems to be no convincingly principled way for liberals to delimit a private sphere of action for individuals once the libertarian's way of delineating this sphere is rejected. When asked what degree of interference with a person's actions is justified for the sake of maximizing some index of aggregate freedom or of some other important value, egalitarian liberals

will propose a variety of domain-restricted principles in an effort to strike a reasonable balance in the light of weighing consequentialist pros and cons in the widest possible reflective equilibrium. This method of generating political principles is, of course, afflicted with enormous imprecision, presupposes large amounts of information and goodwill for its proper application, and usually permits a wide range of equally rational alternatives. At the second-order level of moral reasoning, actual decision-makers in legitimate positions of power will have to select the most politically opportune principle from among all those which pass the first-order tests. Apart from giving us the standard liberal core of political and civil rights, employing this method will often have an air of sheer arbitrariness which is likely to give rise to controversy, especially among those who understand political morality in much simpler terms.

Nevertheless, I think the libertarian way of delineating a private sphere of action for individuals is rationally untenable. It seems clear that Narveson's contractarian libertarianism has no credible potential to capture enough of the pattern of people's expectations and sentiment about *justice* to become acceptable to them as defining a just social arrangement and to enlist their sense of justice in support of it. Utilitarian defense of libertarianism consists essentially in extolling the consequentialist virtues of capitalist economic arrangements over all alternatives, but the consequentialist case for putting *some* restraints on the unfettered workings of capitalist markets seems to me overwhelming, especially in a world threatened by ecological disasters. The unfettered workings of free markets could lead to a situation where the effective exercise of the freedom which libertarians so treasure, the freedom of individuals to use their own minds and bodies as they choose as long as they do not thereby make others worse off, could become restricted to a relative minority of wealthy individuals.

Freedom to use our own minds and bodies as we choose as long as we do not thereby make others worse off presupposes conditions which allow us to develop

our minds and bodies in the first place, and a social framework within which this freedom can be effectively exercised. Providing these conditions and this framework in a substantive way for the greatest number of individuals will require restrictions on the very freedom for whose sake they are to be provided. This paradox, perhaps appropriately called "the libertarian's paradox", seems to me to be the crux of the whole dispute between libertarians and egalitarians. Since both egalitarians and libertarians agree that freedom to exercise one's talents is valuable, and since exercising one's talents presupposes conditions which permit them to be developed in the first place, some moderate egalitarianism combining an economic safety net with competitive markets would seem to me rationally superior, from a consequentialist perspective, to any pure libertarianism or a radical egalitarian extreme. But there are many conflicting compromise principles of this sort. Where exactly the proper balance lies must be judged for a given society at a given time and will surely leave immense room for rational disagreement.

SUMMARY AND CONCLUSIONS

What accounts for the immense disagreement about social justice among informed people of moral goodwill - and how could at least some of it be resolved?

To approach these questions intelligently, in the face of the recent proliferation of conflicting theories of justice, I proposed to explore the possibility of "rational" disagreement and identified some plausible sources and limits of rational disagreement about justice. Let me now sum up what I think has been accomplished.

I defined rational disagreement as disagreement about conceptual, normative or empirical issues which are either *currently irresolvable* (there are empirical or non-basic conceptual or normative questions at issue which additional information, clearer reasoning or conceivable methods may settle eventually) or *fundamentally irresolvable* (there are basic conceptual, basic normative or intractable empirical questions at issue which no amount of accessible information, clearer reasoning or innovative methods will ever settle).

Before this idea of rational disagreement could be applied to the problem of social justice, parameters of relevance had to be established. I analyzed "justice" as fundamentally a dual-purpose concept. The first purpose of this concept, at the level of *feeling injustice*, is to give voice to the claims which individuals or groups, backed by their sense of justice, want to make to a share of social resources. The second purpose, at the level of *doing justice*, is to articulate an appropriate political response to the pattern of expectations and sentiment about justice in a society.

The sense of justice has an emotive and a cognitive component. The cognitive component of the sense of justice allows individuals to assume a perspective beyond mere self-interest in pressing their claims and to recognize a (culturally variable) set of criteria such as desert, need, equality, rights and good-faith expectations as justifying other people's analogous claims to a share of social

resources. There is generally a vast range of logically possible options for designing social institutions to respond to the pattern of expectations and sentiment about justice in a society. An important task of justice theorizing consists in conceptualizing and rationally ranking these options in a way that, ultimately, can guide political decision-makers.

I insisted on a somewhat neglected distinction between theories of justice and social ideals. Just as there can be many equally rational conceptions of the good life, so there can be many equally rational visions of the ideal society. Claims about what *justice* requires are not mere claims about what this or that person's vision of the ideal society requires. Trying to pass off one of many rationally permitted visions of the ideal society as a theory of justice may well contribute to mischief-causing expectations and exacerbate the clash of dogmatisms.

What is a theory of social justice? I argued that a proposal for the distribution of valued things in society must satisfy two criteria to count as a theory of justice: (1) it must propose a politically feasible response to the conflicting claims to social resources which different people (as individuals or groups) make in pursuit of their well-being and which they back up with their sense of justice; and (2) this proposed response to the pattern of justice-based claims must be explicitly and credibly - though not, of course, necessarily with success - argued as having a sufficient degree of rational superiority over rival proposals so that informed people of good will (insofar as they are interested in finding publicly defensible rules of social cooperation) cannot reasonably reject it. Without meeting this second condition, a proposal for the distribution of valued things can immediately be consigned to the scrap heap of social ideals.

I then proposed two necessary, though admittedly quite vague, conditions for the *adequacy* of a theory of justice: (1) the theory must have sufficient normative determinacy to provide guidance to political decision-makers, at least in the form of recognizable second-best approximations; and (2) it must respond to the pattern of

expectations and sentiment about justice in such a way as to have credible potential to enlist enough people's sense of justice in support of this theory to make it the eventual basis for a reasonably stable social consensus about the just distribution of goods. But offering necessary *and* sufficient conditions for an adequate theory of justice requires yet another condition: (3) An adequate theory of justice must, at least in the minds of informed people of good will, uncontroversially meet the first two necessary conditions *to a greater degree* than rival theories, or else it must have additional innovative features which will make it superior to its alternatives.

We currently have many theories of justice, but none, I think, which plausibly satisfies the above criteria of adequacy. Political decision-makers who seek normative guidance from contemporary theories of social justice will discover a perplexing array of competing ideas, a quagmire of abstractions, a continuation of real-world conflicts about justice carried on by other means.

It is clear that a theory of *justice* must meet stricter conditions than someone's pipe dream of the ideal society. But why the *above* criteria? Aren't we begging the question against those who want to identify the 'correct' theory of justice in radically different ways - for example, by intuiting a realm in which everyone's just distributive share exists in its perfect Platonic form?

Obviously, claims about the possibility of rational disagreement about justice are non-trivial only if reason can define limits to it. Unless there can be some rationally compelling way to rule out many widespread ideas about the ultimate nature of morality, the judgments of philosophers about justice are really no more compelling than those of people shouting from the barricades.

On the one hand, I felt strong pressure to endorse a moral framework capable of setting non-trivial limits to rational disagreement about justice; but on the other hand, I was sensitive to the objection that making strong enough moral commitments to set such limits, in the face of intractable metaethical disputes among people with impeccable rational credentials, would seem dogmatic in the

context of my basic project of identifying plausible sources of rational disagreement about justice. In the face of this dilemma I decided to accept a form of consequentialism and to use four lines of defense against the above objection: (1) by making an (admittedly very truncated) effort to suggest how the widest possible reflective equilibrium of relevant considerations is ultimately likely to converge on some broadly consequentialist framework among all *prima facie* plausible meta-ethical options; (2) by admitting the tentativeness of my claims about the likelihood of such eventual convergence and allowing a reading of my findings about the *limits* of rational disagreement as presupposing a broadly consequentialist moral framework; (3) by pointing out that a number of plausible *sources* of rational disagreement within consequentialism are likely to emerge in analogous form in many other moral theories as well; and (4) by suggesting that the context-dependent flexibility of the form of consequentialism I adopt offers greater prospects for realizing Rawls' hope of devising "principles of accommodation between moralities" than plausible alternatives.

Canvassing *prima facie* plausible meta-ethical options in an attempt to establish the widest possible reflective equilibrium of relevant considerations, I was led to adopt a constructivist, non-foundational, distribution-sensitive form of consequentialism. This moral framework is constructivist because it conceives of morality as a social pressure system constructed by human beings for the purpose of creating sufficient uniformity and thus predictability in people's behavior to establish the trust and security required for mutually beneficial interaction.[1] It is non-foundational because it conceives of the justification of moral principles as showing them to be supported by a web of cohering considerations rather than deduced in a non-circular fashion from independently given foundations. It is a form of (indirect) consequentialism because it regards moral reasoning as involving

---

[1] Cf. Haslett (1987), p. 77. I paraphrase his felicitous formulation here of the basic rationale for having a morality at all.

the task of making it plausible that following certain principles will actually have the consequence of promoting well-being. And it is distribution-sensitive, because it regards the goal of promoting well-being to involve both aggregative and distributive values. For a consequentialism of this kind,[2] promoting the goal of well-being means ideally aiming at *that* combination of satisfied interests of individuals in their own well-being that all of us would converge on from an *enriched perspective* which involves giving every individual's interest in his or her well-being exactly the same consideration which each of us would give it if each of us had this individual's life to live and judged from this individual's perspective.

A consequentialist understanding of justice arises from the recognition that *taking seriously* those claims to social resources which individuals or groups make in pursuit of their well-being and back up with their sense of justice is likely to promote the goal of well-being both directly and indirectly: directly, because people's justice-based claims are a good though fallible source of information about what will promote their well-being; and indirectly, because ignoring these claims is likely to unleash destructive sentiments. But at least in complex pluralist societies, no particularly smooth fit will generally be possible between the function of the concept of justice at the level of feeling injustice and its function at the level of doing justice.

At the level of doing justice, then, what is inside the scope of rational disagreement about justice and what is outside? I found reason to be hopeful that many issues dividing different conceptions of justice can indeed be settled at the level of fundamental theory with the rational means currently at our disposal for resolving the conceptual, normative and empirical points at issue. Such issues are, therefore, outside the scope of rational disagreement, as I defined this concept. It needs to be remembered, however, that by excluding some issue from the scope of rational disagreement I do *not* imply that there is no longer any dispute about it among reasonable people of good will. All I mean to imply is that, from the

---

[2] To be labelled simply "consequentialism" in what follows.

perspective of the widest possible reflective equilibrium I have been able to achieve, I tentatively judge the issue to be resolvable by close attention to currently available information, arguments and methods. I cannot claim finality for my judgments. An issue which I optimistically classify as decidable by rational means currently at our disposal may, from the perspective of an even wider reflective equilibrium, turn out to be rationally undecidable after all, or decidable in a very different way than my limited understanding allowed me to see.

On the basis of my inquiry, which issues promise to be rationally decidable, and which do not, when it comes to social justice? In fact, this question raises two distinguishable concerns: first, how to define non-trivial limits to rational disagreement to keep my position from collapsing into skepticism; and second, how to identify plausible sources of rational disagreements about justice.

Plausible *limits* to rational disagreement about justice (within consequentialism)

There are several major issues about which I tried to establish at least a some presumption of rational decidability so as to justifiably exclude them from the scope of rational disagreement. The following summary is meant to be at best suggestive of the general lines of reasoning that seem to me cogent.

1. The issue between basically liberal and non-liberal conceptions of justice: There are two fundamental views about the best social order. The liberal view is that social institutions do not exist to enforce worthy ways of living, but primarily to promote the freedom of individuals to live as they prefer, within a system of constraints designed to avert conflict while maintaining individual freedom to the greatest possible extent. The non-liberal view is that social institutions should constrain individuals to pursue a fairly narrow spectrum of ways of living known to be worthy in advance, regardless of the preferences of those individuals. The problem of justice will take different forms for liberals and non-liberals. If the state exists to enforce a limited spectrum of ways of living which can be known, with

sufficient confidence and precision, to be worthy, then we have available a more or less well-defined blueprint for determining just distributions. If, however, the state should not coerce individuals to pursue worthy ways of living, then the problem of justice will involve the complexities discussed below.

I tried to establish some presumption that, within at least moderately prosperous pluralist societies, the dispute between basically liberal and non-liberal views about the best social order is *not* a disagreement between equally rational conceptions of the ideal society. Broadly egalitarian forms of liberalism can be justified as the best social order on consequentialist grounds, provided that we are looking for a basic social arrangement which can be *publicly defended as reasonable to* individuals with different ideas of the good. This justification depends on the fact that a liberal order leaves greater social space for a wide variety of ideas of the good than any other alternative. It does *not*, however, depend on denying that non-liberal orders can embody cultural virtues which are likely to succumb to competitive pressures in an egalitarian-liberal order, nor on denying that, from the narrow perspective of the *self- or group interest* of a member of a non-liberal culture, egalitarian liberalism may well *not* be the best social order.

The defense of a liberal order I have in mind proceeds as follows: (1) We can know that there is a diversity of possible values for human beings which no one life can encompass, and therefore that there are many equally good ways of living rather than one particular way which is best for everyone. (2) Every non-liberal proposal for the best way of living which the state should enforce encounters rival proposals from other non-liberals, and there exists no non-partisan rationale for choosing among an immense array of non-liberal options. (3) In the face of perfectly rational diversity about what is of value in life, one basic political scheme stands out in the chaos of competing schemes as offering a way of mutual accommodation while minimizing mutual coercion. This is the proposal to set up social institutions around the idea of permitting individuals to seek their own good in their own way

(and equipping them through education to do so wisely), within a system of constraints designed to avert conflict while maintaining individual freedom to the greatest possible extent. There are many different systems of constraints that arguably satisfy this admittedly very general idea equally well. Although a basically liberal social order can take on a variety of specific shapes, a commitment to such an order takes us a long way from skepticism about justice and makes talk about "rational disagreement" meaningful.

Within a consequentialist framework, this justification for a basically liberal social order will, of course, be most persuasive given already fairly pluralist societies enjoying at least moderately favorable economic prospects.[3] Some additional consequentialist reasons for a basically liberal social order are: the desire to avoid a civil war between competing non-liberal forms of perfectionism - liberalism offering perfectionists of many different stripes probably the most viable second-best terms of accommodation; the notorious fallibility of government paternalism; the character-building effects of making one's own choices and mistakes; and the innovations arising out of allowing experiments in living. Therefore, there is a sufficiently strong presumption *against* non-liberal conceptions of justice to exclude them from the scope of rational disagreement about justice, at least in already fairly pluralist societies enjoying moderately favorable economic prospects.

2. *The issue between directly and indirectly responsive conceptions of justice in liberal societies:* Directly responsive conceptions respond to people's justice-based claims by interpreting justice as requiring redistributions of certain important goods

---

[3] For societies facing massive starvation and illiteracy problems, I would want to raise very different considerations, given the quite likely case that liberal-democratic institutions would further add to the political and economic chaos. For culturally homogenous societies which embody a non-liberal but "benign" form of perfectionism that has deeply shaped the self-understanding of individuals in that society (e.g. certain indigenous tribal communities), the consequentialist case *against* promoting radical social change is also very strong, and the argument I make for liberalism would have be qualified considerably. Moreover there is a real possibility that, within a few decades, population pressures will not leave space for liberal democracies at all, as ten billion people lead a cramped existence on an ecologically ravaged planet. Liberal justice theorizing may become a lot simpler in this situation, or altogether irrelevant. But I cannot pursue these matters here.

from those who have too much to those who have too little, according to a standard of ideally just individual holdings of these goods. Indirectly responsive conceptions respond to people's justice-based claims by interpreting justice as requiring a scheme of basic social institutions organized in a particular way, according to a standard of an ideally just pattern of overall distributive outcomes or procedures.

The question of which of these two basic conceptions of justice is superior can, I think, be rationally decided in favor of directly responsive conceptions. I looked at two prominent examples of indirectly responsive conceptions: John Rawls' theory of justice and the libertarian theory developed by Narveson on Gauthier's contractarian foundations. These two conceptions, although clearly not the only possible indirectly responsive conceptions, are probably the initially most plausible current examples of their kind. Rawls' theory specifies a distributive *outcome*, libertarian theories a distributive *process*, which society's basic structure must be designed to uphold. I argued that Rawls' theory, and probably any other outcome-specific indirectly responsive theories of justice that might be proposed, can be ruled out for three main reasons:

(1) The method of wide reflective equilibrium, if directly applied to the question of which overall distributive outcome society's basic structure should achieve, is too indeterminate to support more than a broad range of strongly conflicting principles. (2) A distributive blueprint for society's basic structure must incorporate massive simplifications of commonly invoked just-making criteria - most notably the idea of judging justice by maximizing the expectations of one particular representative social position - so as to end up having little or no potential to harness enough people's sense of justice in its support to make this blueprint the basis for a stable social consensus about justice. (3) Despite massive simplifications such a distributive blueprint is not likely to carry much normative payload in the end, when it comes to giving useful guidance to would-be reformers of society's basic structure because it cannot specify, with any precision, a process to

decide the division of responsibilities among particular institutions for achieving the required outcome.

Libertarian theories can also be ruled out, for reasons to be outlined below. Therefore, only directly responsive conceptions of justice would seem to fall within the scope for rational disagreement.

3. *The issue between libertarian and egalitarian conceptions of justice:* Libertarian conceptions of justice require the state to uphold a free market and to enforce a right to non-interference with the property of individuals which they acquire by using their own minds and bodies as they choose without thereby making others worse off. In contrast, egalitarian conceptions of justice require the state to provide a social framework of universal access to a more or less extensive stock of goods which constitute essential means to a successful pursuit of almost any idea of the good life, such as access to proper nutrition, education, medical care, employment, housing, legal aid, child care and many other possible goods.

Currently available information and arguments are probably sufficient to settle the issue between libertarian or egalitarian conceptions of justice in favor of broadly egalitarian conceptions. Ruling out non-constructivist moral appeals to natural rights, we looked at two *prima facie* credible constructivist attempts to ground libertarian property rights: Gauthier's contractarian argument and the consequentialist arguments of utilitarian libertarians like Hayek. Although these seem to me the initially most plausible versions of libertarianism, I do not deny, of course, that other possible versions would have to be examined as well.

Narveson's attempt to ground the libertarian right to non-interference in a principle of rational choice proposed by Gauthier (a principle that everyone is supposed to be equally ready to accept, whatever the relative strength - within wide limits - of his or her bargaining position), succeeds in evading the contractarian's dilemma only at the expense of becoming indefensible on at least three related grounds:

(1) There is reason to believe that no general principle exists which would be equally acceptable to strong and weak bargainers prior to bargaining. Even if the normatively rather toothless idea behind the Lockean proviso (that no rational person would object to others making themselves better off if doing so had absolutely no detrimental effect on them) *is* a principle of rational choice, it is clear that Gauthier's particular spelling-out of this proviso cannot be regarded as a sufficiently uncontroversial interpretation of it. Narveson's attempt to connect Gauthier's interpretation of the Lockean proviso up with strong libertarian non-interference rights failed precisely because there is every reason to think that neither the strong nor the weak would find it in their self-interest to agree to this principle regardless of the specifics of a particular bargaining situation.

(2) Despite the seeming precision of Gauthier's approach at the abstract level, he abstracts from the particulars of real life situations to such an extent that his theory gives no useful normative guidance. Even if justice is equated to the outcome of bargaining among individual utility-maximizers, most real-life disputes about justice involve far too complex variables for ex ante generalizations about the most rational agreements from the point of view of self-interested bargainers. (On this question of normative guidance, Gauthier seems to equivocate between what I have called "abstract" and "robust" mutual-benefit individualism.)

(3) Even if Gauthier's contractarian theory of "justice" had a useful normative payload, it would have no credible potential to enlist enough people's sense of justice in support of this theory to make it the eventual basis for a widespread social consensus about justice. It fails to respond to the pattern of expectation and sentiment about justice in three ways: it systematically splits the cooperative surplus arising from each cooperative interaction in such a way that existing inequalities will be maintained or even exacerbated; it allows the strong to better themselves by undermining the livelihood of the weak whenever the benefits of doing so are greater than cooperative interaction; and it explicitly affirms

the rationality for those with "superior technology" to exterminate the weak when their doing so outweighs the potential gains of cooperation (231-2). Because the self-interested calculations of Gauthier's individual utility-maximizers may well make such actions rational and therefore morally required, Gauthier's theory seems spectacularly unsuited for encouraging a real sense of security and trust among people and thus holds no promise for accomplishing morality's main job.

In contrast to contractarian libertarians, utilitarian libertarians rely on a defense of the libertarian right to non-interference which tries to show that this right passes the consequentialist well-being test required for rights after all. They equate the exercise of this right with exercising one's talents and ambition in a free market, and extol the ability of a fairly unbridled capitalist economic system to promote human well-being more effectively than any alternative. I have briefly sketched my reasons for thinking that their arguments cannot justify the libertarian right to non-interference, but can establish at best a strong presumption for a market economy supplemented by redistributive social programs. But of course, such corrective redistributions tailored to the situation of particular individuals will open the floodgates to competing claims once more (which can only be resolved in domain-restricted ways). Although consequentialists must acknowledge that the freedom of individuals to use their own minds and bodies to make themselves better off is indeed an important value, exercising this freedom presupposes means and conditions which allow individuals to develop their minds and bodies in the first place, and therefore requires a social framework of *substantive* equality, or at least *limits* to inequality, of access to these means and conditions. Hence protecting the universal exercise of the very freedom that utilitarian libertarians appeal to (as well as realizing other important values) requires restrictions on this freedom. I therefore conclude very provisionally - not having examined *all* libertarian arguments in sufficient detail - that the scope of rational disagreement can probably

be effectively narrowed to liberal-egalitarian conceptions of social justice which propose directly responsive principles.

4. The issue between directly responsive conceptions of justice that are of society-wide scope and those that are of domain-restricted scope: Conceptions of society-wide scope propose a master principle of justice which (1) is intended to provide necessary and sufficient conditions for determining every individual's just share of every type of good at any given moment in time throughout society; and (2) individuates a fairly static list of goods whose distribution it regards as the concern of justice, along with a fairly static list of criteria a person must meet to qualify for certain amounts of certain goods. In contrast, conceptions of domain-restricted scope propose a plurality of different (and more or less easily modifiable) master principles of justice of widely varying characteristics for a more or less large plurality of restricted and often changing domains. Conceptions of justice whose principles are of domain-restricted scope may vary greatly in the degree of their ambition to pull up people's conflicting intuitions about justice in particular cases to the level of principles. A complete collection of all such domain-restricted master principles, if they could ever actually be collected, would of course add up to a sort of master principle of society-wide scope, even though its internal complexity and lack of priority ordering would not allow us to determine every individual's just share at any given time without resorting to extremes of intuitive balancing which would render it effectively indeterminate.

I think that the issue between directly responsive conceptions of justice of society-wide and of domain-restricted scope is decidable, by currently available information and arguments, in favor of domain-restricted conceptions. To understand the reasons why, however, we need to draw upon a further layer of distinctions. I will briefly recapitulate these distinctions, since the reasoning for my conclusion depends crucially on keeping them clear.

Conceptions of justice, we said, are either intuitionistic or non-intuitionistic. Intuitionistic conceptions of justice propose a plurality of relevant criteria for identifying a just distribution of a particular bundle of goods, without specifying an algorithm for assessing the relative importance of the various criteria when they give rise to conflicting claims. Non-intuitionistic conceptions propose *either* a single criterion as ultimately relevant for identifying a just distribution of a particular good, *or* else they specify a lexical ordering as a method for avoiding the need for intuitive balancing of criteria in cases of conflict.

Single-criterion (directly responsive) principles of justice of *society-wide scope* specify only one criterion which individuals must meet to qualify for a particular distributive share. But ultimately there are no plausible criteria of this kind. Criteria which might *seem* to meet this condition, like needs or (non-libertarian) rights, are internally complex and require intuitive balancing after all, as well as being unacceptably lopsided as a credible response to individuals' justice-based claims to a variety of different goods in different social contexts. Nor can we go very far in aligning a plurality of different just-making criteria with this variety of goods and then finding a plausible lexical ordering for different criteria, because such a lexical ordering will appear unacceptably rigid in the face of trade-off opportunities which are clearly required on consequentialist grounds. So all *prima facie* plausible master principles of society-wide scope must either be directly responsive and intuitionistic (or if non-intuitionistic, then highly domain-restricted). Now there exists an overabundance of conflicting proposals for society-wide principles of the intuitionistic kind, and they are easy to make up, according to one's favorite social utopia. Each would seem to face a standard set of next-to-insuperable obstacles: the justification problem, the priority problem, the aggregation-distribution problem, the index problem as well as additional commensuration problems. Such obstacles can be overcome, if at all, only in very small, non-pluralist societies.

There is a strong presumption, therefore, that *only* domain-restricted liberal-egalitarian principles of justice (which in wide domains will be intuitionistic and in fairly narrow domains may be relatively non-intuitionistic) fall within the scope of rational disagreement in complex pluralist societies. Only domain-restricted principles can factor the balancing problems into small enough parts to keep them manageable while having a hope of responding to the pattern of domain-relative expectations and sentiments about justice in a way to enlist enough people's sense of justice in support of such principles.

Under conditions currently prevailing in many societies, then, our consequentialist framework narrows the scope for rational disagreement about justice to a basically liberal-egalitarian social order. It is clear that such a social order will pass the consequentialist well-being test only if it strikes a balance between two important values which are irreconcilably in tension. It must respect the value of providing secure opportunities for everyone to gain reasonable access to the necessary means for realizing almost any legitimate idea of the good (including, most importantly, sufficient means to meet course-of-life needs by instituting an economic safety net). And it must also respect in large measure the freedom of individuals to use their own minds and bodies as they see fit and to pursue their own intentional tracks through life. But within such admittedly broad parameters, attempts to find principled responses to people's justice-based claims must consist in seeking support for domain-restricted principles by promoting stable expectations within restricted domains - rather than in multiplying ideals of everyone's just individual share of total social resources or in articulating ambitious ideals of the overall distributive outcome to which society's basic structure must conform.

Our consequentialist framework leaves considerable room, at the level of abstract theory, for context-dependent justice theorizing and for rational disagreement which must be resolved at the second-order level of political negotiation and compromise.

Plausible *sources* of rational disagreement about justice (within consequentialism)

I shall now recapitulate, in the briefest of terms, some plausible candidates for the status of rational disagreements about social justice. Deeper insights into what is, admittedly, a still rather unruly collection of problems will emerge only after a great deal of clarifying dialectic.[4] Although my account of these problems was developed against the conceptual background of a form of consequentialism, they may of course also arise, in recognizably analogous forms, for theories of justice embedded in alternative moral frameworks.

1. *The criteria definition and inclusion problems:* At the level of feeling injustice, there is an unruly family of very imprecise evaluative considerations which people invoke in making claims in pursuit of their well-being while being willing to allow others to make claims based on the same considerations. These considerations include many (culturally variable and even intraculturally diverse) interpretations of desert, need, equality, rights and good-faith expectations. At the level of doing justice, difficult problems will arise – solved by different conceptions of justice in different ways – about what considerations to include as just-making criteria in a principled response to the pattern of expectations and sentiment about justice.

2. *The problem of feasible comprehensiveness:* On the one hand, we want to pull judgments about justice up to the level of principles; on the other, we cannot pitch the level of generality of these principles very high without emptying them *both* of normative determinacy *and* of credible potential to enlist people's sense of justice. Individuals construct their selves in the context of restricted domains, and their justice-based claims are projections of such domain-embedded selves. Further work is clearly needed to link up my very general remarks about domain-restricted principles of justice with the viable domains in a particular

---

[4] Rawls named the priority problem and the index problem.

society. Doing so would obviously be a highly complex, society-specific inquiry into evolved and contrived institutions, practices, social roles, cultural groupings, economic development, geographic particulars and other contingent circumstances.

3. Balancing problems: These problems include the *priority problem* (what weights to assign to the various just-making criteria invoked by a theory whenever resources are insufficient to completely satisfy them all); the *aggregation-distribution problem* (how to decide between satisfying a greater amount of justice-based claims and satisfying at least some of the most urgent justice-based claims of all people); and various other *commensuration problems* (how to assess in practical detail, for all individuals in society, what redistributions are required from those whose holdings exceed their just share to those whose holdings fall short of it). Rawls' method of judging the justice of society's overall pattern of distribution from the perspective of a particular social position vastly simplifies commensuration problems. It reduces many of them essentially to the *index problem* (finding an index for different bundles of primary goods held by different people to compare their values). But even if this index problem could somehow be kept manageable for indirectly responsive theories, Rawls would still face difficult commensuration problems in deciding which elements of the basic structure to reform in which ways.

While breaking the problem of social justice into domain-restricted principles is, I argued, probably the only way to make the priority, aggregation-distribution and commensuration problems manageable while yet maintaining credible potential to harness enough people's sense of justice in support of these principles, such a domain-restricted, directly responsive approach to social justice will still face all these problems *to some extent,* in addition to two new problems: the domain demarcation problem (how to select viable domains which are sufficiently independent from neighboring domains for regulation by their own domain-specific principles); and the domain balancing problem (what to do when different domains compete for the same resources). These problems require contextual

perspective-seeking in reflective equilibrium, and will be rich sources of rational disagreement, at the first-order level of moral reasoning, about the demands of social justice and those of morality in general.

4. Various problems of first-order justification (within consequentialism): There can be rational disagreement about which *bona fide* principle of justice an informed person of good will is actually obligated to comply with. The concept of a *bona fide* principle of justice is central to my analysis. *Bona fide* principles of justice are, in essence, all *those* proposal for balancing justice-based claims in a domain that could be reached by decision-makers who are guided by a thorough understanding of the distributive domain at issue and the desire to satisfy people's justice-based claims in that domain to the greatest extent compatible with respect for the separateness of persons.

Bona fide principles are *not* equivalent to morally justified principles of justice - principles which are such that compliance with them is morally required. A *bona fide* principle can be regarded as morally justified only if it passes the test of those moral standards which are plausibly derivable from the form of consequentialism we have adopted. But usually a range of competing principles will pass this test relative to any given domain, and this range defines the scope for first-order rational disagreement about what fundamental moral theory requires in this domain. To identify a particular principle within this range, such that actual *compliance* with it is morally required for a person, will generally take us beyond the first-order level of moral reasoning, into the sphere of legitimation of a particular principle by appropriate political decision-makers.

Rational disagreements arise at several other points in the process of moral justification. I adduced consequentialist reasons for thinking that the most plausible way to interpret the demands of justice is as follows: "Justice requires that at least one among the *bona fide* principles in some viable domain be implemented." If justice requires only that at least one among the *bona fide* principles in a domain be

implemented, and if a range of different *bona fide* principles is equally permissible in most domains, it follows that first-order moral reasoning can identify the demands of justice in a domain only with that degree of precision which corresponds to this range of equally permissible *bona fide* principles. It also follows that the important concept of "injustice" (i.e. a violation of what justice requires) will be defined in a largely domain-relative way: as referring to a distribution that is *in*compatible with all *bona fide* principles within a given domain. To the extent that domains are not securely demarcated, therefore, something may be seen as an injustice if domains are demarcated one way but not if they are demarcated another. Moreover, what justice requires in one domain may conflict with what justice requires in another. And what justice requires may conflict with what morality requires, all things considered. Although the consequentialist rationale for this structure of moral justification is clear enough, in practice there will often be room for rational disagreement about whether an injustice has actually been committed, and if so, whether committing it is morally justifiable, all things considered.

In developing my account I mentioned additional sources of rational disagreement and no doubt passed over other potential sources. I am conscious of having unravelled only a modest part of the enormous complexities involved in any effort to overcome the current impasse in justice theorizing.

The upsurge of justice theorizing since the publication of *A Theory of Justice* two decades ago has given us interesting social ideals. None of them is sufficiently compelling, it appears, to serve as a promising basis for a social consensus about the demands of justice within large, complex pluralist democracies. These theories have, however, taught us important lessons. A particularly important lesson seems to me to be this: *curbing conflicts about justice demands a vastly clearer understanding of the nature of rational disagreement than is currently available.* Only a systematic focus on the sources of rational disagreement about justice defines

a line of effort which, I believe, may yet fulfill the sort of hope eloquently expressed by John Rawls:

> The hope is that...existing differences between contending political views can at least be moderated, even if not entirely removed, so that social cooperation on the basis of mutual respect can be maintained. Or if this is expecting too much, this method may enable us to conceive how....a public understanding could arise consistent with the historical conditions and constraints of our social world. Until we bring ourselves to conceive how this could happen, it can't happen. (1985: 231)

Conflicts about justice incite flaming passions throughout the contemporary world. To make sense of justice-talk in our time, to avoid both dogmatism and skepticism as well as the pragmatism of Machiavellian power politics, and to invoke the concept of justice in ways that will help mediate rather than escalate conflict, all this requires a fuller recognition of the conceptual, normative and empirical sources of rational disagreement about justice. Such recognition will perhaps move human minds from a state of moral certainty to that state of doubt which is the beginning of wisdom. Such recognition can also fill the gap between opposing beliefs, producing an enriched perspective from which the need to negotiate compromises and to set new conventions is better appreciated. A philosophy of justice which promotes an understanding of the deepest sources of rational disagreement and thereby produces such an enriched perspective may well offer our best prospect for transcending many raging conflicts about justice, both within human societies and between them.

BIBLIOGRAPHY

Ackerman, Bruce A., *Social Justice in the Liberal State*, New Haven: Yale University Press, 1980.

Aristotle, *Nichomachean Ethics*, in: *The Complete Works of Aristotle*, Vol. Two, ed. Johnathan Barnes, Princeton: Princeton University Press, 1985.

Arneson, Richard J., "Introduction", *Symposium on Rawlsian Theory of Justice, Ethics*, Vol 99 No 4, Julyu 1989, pp. 695-711.

Attig, Thomas, Donald Callen and R.G. Frey, ed., *Social Policy and Conflict Resolution*, Bowling Green Studies in Applied Philosophy, Vol VI - 1984.

Baier, Kurt, "Justice and the Aims of Political Philosophy", *Ethics*, Vol 99 No 4, July 1989, pp. 771-790.

Baldwin, R.W., *Social Justice*, New York: Pergamon, 1968.

Barrow, Robin, *Injustice, Inequality, and Ethics: A Philosophical Introduction to Moral Problems*, Totowa, N.J.: Barnes & Noble, 1982.

Barry, Brian, *Political Argument*, London: Routledge & Kegan Paul, 1980.

Barry, Brian, *The Liberal Theory of Justice*, London: Oxford University Press, 1973.

Barry, Brian, *Theories of Justice*, Vol I of *A Treatise on Social Justice*, Berkeley: University of California Press, 1989.

Barry, Norman P., *On Classical Liberalism and Libertarianism*, London: Macmillan, 1986.

Benjamin, Martin, *Splitting the Difference, Compromise and Integrity in Ethics and Politics*, Lawrence, Kansas: University Press of Kansas, 1990.

Benson, Peter, "The Priority of Abstract Right, Constructivism, and the Possibility of Collective Rights in Hegel's Legal Philosophy", *The Canadian Journal of Law and Jurisprudence*, Vol IV, No 2, July 1991, pp. 257-91.

Berlin, Isaiah, *Four Essays on Liberty*, New York: Oxford University Press, 1969.

Bird, Otto A., *The Idea of Justice*, New York: Frederick A. Praeger, 1967.

Blackstone, W.T., and R.D. Heslep, eds., *Social Justice and Preferential Treatment: Women and Racial Minorities in Education and Business*, Athens, Georiga: University of Georgia Press, 1977.

Blocker, H. Gene, and Elizabeth H. Smith, ed., *John Rawls' Theory of Social Justice*, Athens: Ohio University Press, 1980.

BonJour, Lawrence, *The Structure of Empirical Knowledge*, Cambridge: Harvard University Press, 1985.

Bradie, Michael, and David Braybrooke, *Social Justice, Bowling Green Studies in Applied Philosophy*, Vol IV, 1982.

Bradley, F.H., *Ethical Studies*, 2nd ed., Oxford: Clarendon, 1959.

Brandt, Richard B., ed., *Social Justice*, Englewood Cliffs, N.J.: Prentice-Hall, 1962.

Brandt, Richard B., *A Theory of the Good and the Right*, Oxford: Clarendon Press, 1979.

Braybrooke, David, *Three Tests for Democracy: Personal Rights, Human Welfare, Collective Preference*, New York: Random House, 1968.

Braybrooke, David, *Meeting Needs*, Princeton, N.J.: Princeton University Press, 1987.

Braybrooke, David, "Michalos' Comments on *Meeting Needs* Disposed of", *Dialogue*, Vol XXVII No 3, Autumn 1988, pp. 517-522.

Brown, Grant, "Critical Notice of Jan Narveson, 'The Libertarian Idea', *Canadian Journal of Philosophy*, Vol 20, No 3, September 1990, pp. 417-447.

Brown, Harold I., *Rationality*, London: Routledge, 1988.

Buchanan, Allen, *Ethics, Efficiency, and the Market*, Totawa, N.J: Rowman & Allanheld, 1985.

Buchanan, Allen E., "Assessing the Communitarian Critique of Liberalism", *Ethics*, Vol 99, July 1989, pp. 852-882.

Buchanan, Allen., *Secession: The Morality of Political Divorce from Fort Sumter to Lithuania and Quebec*, Boulder: Westview Press, 1991.

Campbell, Tom, *Justice*, London: Macmillan, 1988.

The Canadian Journal of Law and Jurisprudence, Topic: Collective Rights, ed. Michael McDonald, Vol. IV, No 2, July 1991.

Cauthen, Kenneth, *The Passion for Equality*, Totowa, N.J.: Rowman & Littlefield, 1987.

Christian, James L., *Philosophy, An Introduction to the Art of Wondering*, 5th ed., Toronto: Hold, Rinehard and Winston, 1990.

Cohen, G.A., *Karl Marx' Theory of History: A Defence*, Princeton: Princeton University press, 1978.

Cohen, Ronald L., ed., *Justice: Views from the Social Sciences*, New York: Plenum Press, 1986.

Copp, David and David Zimmerman, ed., *Morality, Reason, and Truth: New Essays on the Foundations of Ethics*, Totowa, N.J.: Rowman & Allanheld, 1985.

Coval, S.C. and J.C. Smith, *Law and Its Presuppositions*, London: Routledge & Kegan Paul, 1986.

Coval, S., J.C. Smith and Simon Coval, "The Foundations of Property and Property Law", *Cambridge Law Journal*, 45(3), November 1987, pp. 457-475.

Cunningham, Frank, *Democratic Theory and Socialism*, Cambridge: Cambridge University Press, 1987.

Dahl, R.A., *Dilemmas of Pluralist Democracy*, New Haven: Yale University Press, 1982.

Daniels, Norman, ed., *Reading Rawls: Critical Studies of A Theory of Justice*, New York: Basic Books, 1974.

Daniels, Norman, "Wide Reflective Equilibrium and Theory Acceptance in Ethics", *Journal of Philosophy*, Vol 76, No 5, May 1979, pp. 256-282.

Danielson, Peter, "The Lockean Proviso", in: *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, ed. Peter Vallentyne, Cambridge: Cambridge University Press, 1991, pp. 99-111.

Danielson, Peter, "Closing the compliance dilemma: How it's rational to be moral in a Lamarckian world", in: *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, ed. Peter Vallentyne, Cambridge: Cambridge University Press, 1991, pp. 291-322.

Danley, John R., "Liberalism, Aboriginal Rights, and Cultural Minorities", *Philosophy and Public Affairs*, Vol 20 No 2, Spring 1991, 168-185.

Davidson, Donald, *Essays on Actions and Events*, Oxford: Clarendon Press, 1980.

Day, J.P., *Liberty and Justice*, London: Wolfboro, 1987.

Deutsch, Morton, *Distributive Justice: A Social-Psychological Perspective*, New Haven: Yale University Press, 1985.

Dworkin, Ronald, "The Original Position", in: *Reading Rawls: Critical Studies of A Theory of Justice*, ed. Norman Daniels, New York: Basic Books, 1974, pp. 16-53.

Dworkin, Ronald, *Taking Rights Seriously*, Cambridge: Harvard University Press, 1977.

Dworkin, Ronald, "What is Equality? Part 1: Equality of Welfare", in: *Philosophy & Public Affairs*, Vol 10 No 3, Summer 1981, pp. 185-246.

Dworkin, Ronald, "What is Equality? Part 2: Equality of Resources", in: *Philosophy & Public Affairs*, Vol 10 No 4, Fall 1981, pp. 283-345.

Dworkin, Ronald, 'Liberalism', *A Matter of Principle*, Cambridge, MA: Harvard University Press, 1985.

Elster, Jon, "Rationality, Morality, and Collective Action", *Ethics*, Vol 96, No 1, October 1985.

"Equal Opportunity", Special Issue of *Social Philosophy & Policy*, Vol 5, Issue 1, Autumn 1987.

Ethics, Special Symposium Issue on "Rationality and Morality", Vol 96 No 1, October 1985.

Ethics, Special Symposium Issue on "Rawlsian Theory of Justice: Recent Developments", Vol 99 No 4, July 1989.

Ewin, R.E., Liberty, *Community, and Justice*, Totowa, N.J: Rowman & Littlefield, 1987.

Feibleman, J.K., *Justice, Law, and Culture*, Dortrecht: Martinus Nijhoff, 1985.

Feinberg, Joel, *Social Justice*, Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Feinberg, Joel, *Rights, Justice, and the Bounds of Liberty*, Princeton, N.J.: Princeton University Press, 1980.

Fishkin, James, S., *Tyranny and Legitimacy: A Critique of Political Theories*, Baltimore: Johns Hopkins University Press, 1979.

Fishkin, James, S., *Justice, Equal Opportunity, and the Family*, New Haven: Yale University Press, 1983.

Frankel Paul, Ellen, Fred D. Miller Jr, Jeffrey Paul and John Ahrens, *Equal Opportunity*, New York: Basil Blackwell, 1987.

Friedman, Milton, *Free to Choose: A Personal Statement*, New York: Harcout Brace Jovanovich, 1980.

Galston, William A., *Justice and the Human Good*, Chicago: University of Chicago Press, 1980.

Gauthier, David, *Morals by Agreement*, Oxford: Clarendon Press, 1986.

Gauthier, David, "Why Contractarianism?", in: *Contractarianism and Rational Choice*, ed. Peter Vallentyne, Cambridge: Cambridge University Press, 1991, pp. 15-30.

Gauthier, David, "Rational Constraint: Some Last Words", in: *Contractarianism and Rational Choice*, ed. Peter Vallentyne, Cambridge: Cambridge University Press, 1991, pp. 323-330.

Gibbard, Allan, "Constructing Justice", *Philosophy and Public Affairs*, Vol 20, No 3, Summer 1991, 264-279.

Gibbard, Allan, *Wise Choices, Apt Feelings, A Theory of Normative Judgment*, Cambridge, Mass.: Harvard University Press, 1990.

Ginsberg, Morris, *On Justice in Society*, Ithaca, N.Y.: Cornell University Press, 1965.

Goldman, Alan H., "Justice and Hiring by Competence", *American Philosphical Quarterly*, Vol 14 No 1, January 1977, pp. 17-28.

Goldman, Alan, *Justice and Reverse Discrimination*, Princeton: Princeton University Press, 1979.

Goodin, Robert E., *Reasons For Welfare: The Political Theory of the Welfare State*, Princeton, N.J.: Princeton University Press, 1988.

Goodin, Robert E., and Andrew Reeve, eds., *Liberal Neutrality*, London: Routledge, 1989.

Gordon, Scott, *Welfare, Justice, and Freedom*, New York: Columbia University Press, 1980.

Graham, Gordon, *Contemporary Social Philosophy*, New York: Basil Blackwell, 1988.

Grant, George, *English-speaking Justice*, Toronto: Anansi, 1974, 1985.

Greene, Thomas H., *Comparative Revolutionary Movements: Search for Theory and Justice*, 2nd ed., Englewood Cliffs, N.J.: Prentice-Hall, 1984.

Griffin, James, "Some Problems of Fairness", in: *Ethics*, Vol 96 No 1, October 1985, pp. 100-118.

Gross, Barry, ed., *Reverse Discrimination*, Buffalo, N.Y.: Prometheus, 1977.

Gutman, Amy, *Liberal Equality*, Cambridge: Cambridge University Press, 1980.

Gutman, Amy, "Communitarian Critics of Liberalism", *Philosophy and Public Affairs*, Vol 14 No3, Summer 1985, pp. 308-322.

Gutman, Amy, and Dennis Thompson, "Moral Conflict and Political Consensus", *Ethics*, Vol 101 No 1, October 1990, pp. 64-88.

Hayek, Friedrich A., *Law, Legislation, and Liberty*, Vol 2: *The Mirage of Social Justice*, London: Routledge & Kegan Paul, 1976.

Haksar, Vinit, *Equality, Liberty, and Perfectionism*, New York: Oxford University Press, 1979.

Hardin, Russell, *Morality Within The Limits Of Reason*, Chicago: University of Chicago Press, 1988.

Hare, R.M., "Rawls' Theory of Justice", Part I and II, *Philosophical Quarterly*, Vol. 23, April 1973 (144-155) and July 1973 (241-252).

Hare, R.M., *Moral Thinking: Its Methods, Levels, and Point*, Oxford: Oxford University Press, 1981.

Harman, Gilbert, "Moral Relativism Defended", in *Relativism Cognitive and Moral*, Michael Krausz and Jack W. Meiland, eds., Notre Dame: University of Notre Dame Press, 1982.

Harsanyi, John C., "Morality and the theory of rational behavior", in *Utilitarianism and Beyond*, eds. Amartya Sen and Bernard Williams, Cambridge: Cambridge University Press, 1982.

Haslett, D.W., *Equal Consideration: A Theory of Moral Justification*, Newark: University of Delaware Press, 1987.

Heller, Agnes, *Beyond Justice*, Oxford: Basil Blackwell, 1987.

Hobbes, Thomas, *Leviathan*, Parts I and II, Indianapolis: Bobs-Merrill, 1958.

Hobhouse, L.T., *Liberalism*, New York: Oxford University Press, 1911, 1964.

Hobhouse, L.T., *The Elements of Social Justice*, London: Allen & Unwin, 1922.

Hochschild, Jennifer, *What's Fair?*, Cambridge: Harvard University Press, 1985.

Honderich, Ted, *Conservatism*, London: Hamish Hamilton, 1990.

Hume, David, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, 2nd ed., ed. P.H. Nidditch, Oxford: Clarendon Press, 1978.

Hume, David, *An Enquiry Concerning the Principles of Morals*, repr. from the edition of 1777, La Salle, Illinois: Open Court, 1966.

Hurka, Thomas, "Why Value Autonomy?", *Social Theory and Practice*, Vol 13, Fall 1987, pp. 361-382.

Jonsen, Albert R. and Stephen Toulmin, *The Abuse of Casuistry: A History of Moral Reasoning*, Berkeley: University of California Press, 1988.

Joseph, Keith and Jonathan Sumption, *Equality*, London: John Murray, 1979.

Jouvenal, Bertrand de, *The Ethics of Redistribution*, Cambridge: Cambridge University Press, 1952.

Kant, Immanuel, *Groundwork of the Metaphysics of Morals*, London: Hutchinson, 1956.

Kant, Immanuel, *The Metaphysical Elements of Justice*, New York: Bobbs-Merill, 1965.

Kamenka, Eugene, and Alice Erh-Soon Tay, ed., *Justice*, London: Edward Arnold, 1979.

Kaufmann, Walter, *Without Guilt and Justice: From Decidophobia to Autonomy*, New York: Delta, 1973.

Kelbley, Charles A., ed., *The Value of Justice*, New York: Fordham University Press, 1979.

Kluegel, James R., and Eliot R. Smith, *Beliefs About Inequality: Americans' Views Of What Is and What Ought To Be*, New York: Plenum Press, 1980.

Krausz, Michael, and Jack W. Meiland, eds., *Relativism, Cognitive and Moral*, Notre Dame: University of Notre Dame Press, 1982.

Kuhn, Thomas, *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press, 1970.

Kukathas, Chandran and Philip Pettit, *Rawls:* A Theory of Justice *and its Critics*, Cambridge: Polity Press, 1990.

Kymlicka, Will, "Rawls on Teleology and Deontology", in: *Philosophy and Public Affairs*, Vol 17, No 3, Summer 1988, pp. 173-190.

Kymlicka, Will, *Liberalism, Community, and Culture*, Oxford: Clarendon Press, 1989.

Kymlicka, Will, *Contemporary Political Philosophy*, Oxford: Clarendon Press, 1990.

Lerner, Melvin J., *The Belief in A Just World*, New York: Plenum Press, 1980.

Levine, Andrew, *Arguing for Socialism: Theoretical Considerations*, London: Routledge & Kegan Paul, 1984.

Locke, John, *The Second Treatise of Civil Government* and *A Letter Concerning Toleration*, ed. J.W. Gough, Oxford: Basil Blackwell, 1946.

Lomasky, Loren E., "Liberal Obituary?", *Ethics*, Vol 102, No 1, October 1991, 140-154.

Luban, David, "Bargaining and Compromise: Recent Work on Negotiation and Informal Justice", *Philosophy & Public Affairs*, Vol 14 No 4, Fall 1985, pp. 397-416.

Lucas, J.R., *On Justice*, Oxford: Clarendon, 1980.

MacIntrye, Alasdair, *After Virtue: A Study in Moral Theory*, London: Duckworth, 1981.

MacIntrye, Alasdair, *Whose Justice? Which Rationality?*, Notre Dame, Indiana: University of Notre Dame Press, 1988.

Mackie, J.L., *Ethics: Inventing Right and Wrong*, New York: Penguin, 1977.

Maguire, Daniel C., *A New Amercian Justice: Ending the White Male Monopolies*, New York: Doubleday, 1980.

Mapel, David, *Social Justice Reconsidered: The Problem of Appropriate Precision in a Theory of Justice*, Chicago: University of Illinois Press, 1989.

Martin, R., *Rawls and Rights*, Lawrence, Kansas: University Press of Kansas, 1985.

McClennen, Edward F, "Justice and the Problem of Stability", *Philosophy and Public Affairs*, Vol 18 No 1, Winter 1989, pp. 3-30.

McDonald, Michael, "Justice in Hard Times", in: *Social Justice*, ed. Michael Bradie and David Braybrooke, *Bowling Green Studies in Applied Philosophy*, Vol IV, 1982, pp. 34-43.

McDonald, Michael, "After Virtue, Taking Rights Seriously", *Journal of Business Ethics*, Vol 5, 1986, pp. 21-28.

McDonald, Michael, "The Personless Paradigm", University of Toronto Law Journal, Vol XXXVII, No 2, Spring 1987, pp. 212-226.

McDonald, Michael, "Culture, Political Obligation and Ideology", *Philosophy and Culture, Proceedings of the XVIIth World Congress of Philosophy*, Vol II, ed. Venant Cauchy, University of Montreal, 1988, pp. 319-323.

McDonald, Michael, "Should Communities Have Rights? Reflections on Liberal Individualism", *The Canadian Journal of Law and Jurisprudence*, Vol IV, No 2, July 1991, pp. 217-237.

Mendus, Susan, *Toleration and the Limits of Liberalism*, Atlantic Highlands, NJ: Humanities Press, 1989.

Mettrick, Alan, *Last in Line*, Toronto: Key Porter, 1985.

Michalos, Alex C., "Meeting Current Needs", *Dialogue*, Vol XXVII No 3, Autumn 1988, pp. 507-516.

Mill, J.S., *On Liberty*, New York: Crofts, 1947.

Mill, J.S., *Utilitarianism*, Indianapolis: Bobs-Merrill, 1971.

Mill, J.S., *Considerations On Representative Government*, Indianapolis, Bobs-Merrill, 1958.

Miller, David, *Social Justice*, Oxford: Clarendon Press, 1976.

Moon, J. Donald,ed., *Responsibility, Rights, and Welfare: The Theory of the Welfare State*, Boulder, Col.: Westview Press, 1988.

Moore, Margaret, "Justice for Our Times", *Canadian Journal of Political Science*, XXXIII:3, September 1990.

Moser, Paul K., *Empirical Justification*, Dordrecht: D. Reidel, 1985.

Mosovitch, Allan, and Glenn Drover, eds., *Inequality*, Toronto: University of Toronto Press, 1981.

Nagel, Thomas, *The View From Nowhere*, Oxford: Oxford University Press, 1986.

Narveson, Jan, *The Libertarian Idea*, Philadelphia: Temple University Press, 1988.

Narveson, Jan, "Collective Rights?", *The Canadian Journal of Law and Jurisprudence*, Vol. IV, No 2, July 1991.

Nielsen, Kai, "Arguing about Justice: Marxist Immoralism and Marxist Moralism", *Philosophy and Public Affairs*, Vol 17, No 3, Summer 1988, pp. 212-234.

Nielsen, Kai, *Equality and Liberty: A Defense of Radical Egalitarianism*, Totowa, N.J.: Rowman & Allanheld, 1985.

Nozick, Robert, *Anarchy, State, and Utopia*, New York: Basic Books, 1974.

Nozick, Robert, *The Examined Life*, Toronto: Simon and Schuster, 1989.

O'Neill, Onora, "Between Consenting Adults", *Philosophy and Public Affairs*, Vol 14 No 3, 1985, pp. 252-277.

O'Neill, Onora, *Faces of Hunger*, London: Allen & Unwin, 1986.

Pal, Leslie A., *Public Policy Analysis*, Toronto: Methuen, 1987.

Paul, Jeffrey, ed., *Reading Nozick*, Totowa, N.J.: Rowman and Littlefield, 1981.

Perelman, Chaim, *Justice*, New York: Random House, 1967.

Pettit, Philip, *Judging Justice: An Introduction to Contemporary Political Philosophy*, London: Routledge & Kegan Paul, 1980.

Phillips, Derek L., *Toward a Just Social Order*, Princeton, N.J.: Princeton University Press, 1986.

Plato, *The Republic*, in: *The Dialogues of Plato*, transl. B. Jowett, New York: Random House, 1937.

Pogge, Thomas W., *Realizing Rawls*, Ithaca: Cornell University Press, 1989.

Posner, Richard A., *The Economics of Justice*, Cambridge, Mass.: Harvard University Press, 1981.

Rachels, James, *The Elements of Moral Philosophy*, Philadelphia: Temple University Press, 1986.

Railton, Peter, "Moral Theory as a Moral Practice", *Nous*, Vol XXV, No 2, April 1991, pp. 185-191.

Ratner, Robert S. and John L. McMullen, eds., *State Control: Criminal Justice Politics in Canada*, Vancouver: University of British Columbia Press, 1987.

Rawls, John, *A Theory of Justice*, Cambridge, Mass.: Harvard University Press, 1971.

Rawls, John, "Fairness to Goodness", *Philosophical Review,* Vol 84, Oct. 1975, pp. 536-554.

Rawls, John, "The Basic Structure as Subject", *American Philosophical Quarterly",* Vol 14 No 2, April 1977, pp. 159-165.

Rawls, John, "Kantian Constructivism in Moral Theory", *Journal of Philosophy,* No 77, September 1980, pp. 515-572.

Rawls, John, "Social Unity and Primary Goods", *Utilitarianism and Beyond,* ed. Amartya Sen and Bernard Williams, Cambridge: Cambridge University Press, 1982.

Rawls, John, "Justice as Fairness: Political not Metaphysical", *Philosophy & Public Affairs,* Vol 14 No 3, Summer 1985, pp. 223-251.

Rawls, John, "The Basic Liberties and Their Priority", in: *Liberty, Equality, and Law, Selected Tanner Lectures on Moral Philosophy,* ed. Sterling M. McMurrin, Salt Lake City: University of Utah Press, 1987.

Rawls, John, "The Idea of an Overlapping Consensus", *Oxford Journal of Legal Studies,* 7, 1987, pp. 1-25.

Rawls, John, "The Priority of Right and Ideas of the Good", *Philosophy and Public Affairs,* Vol 17, No 4, Fall 1988, pp. 251-276.

Raz, Joseph, *The Morality of Freedom,* Oxford: Clarendon Press, 1986.

Reiman, Jeffrey, *Justice and Modern Moral Philosophy,* New Haven: Yale University Press, 1990.

Rescher, Nicholas, *Distributive Justice: A Constructive Critique of the Utilitarian Theory of Distribution,* New York: Bobbs-Merrill, 1966.

Resnick, Philip, *The Masks of Proteus: Canadian Reflections on the State,* Montreal & Kingston: McGill-Queens University Press, 1990.

Ripstein, Arthur, "Foundationalism in Political Theory", *Philosophy and Public Affairs* 16, No 2, Spring 1987, pp. 115-137.

Rousseau, Jean-Jacques, *Of the Social Contract and Discourse on Political Economy,* New York: Harper & Row, 1984.

Rule, James B., *Insight and Social Betterment,* New York: Oxford University Press, 1978.

Runciman, W.G., *Relative Deprivation and Social Justice*, London: Routledge & Kegan Paul, 1966.

Sandel, Michael, *Liberalism and the Limits of Justice*, Cambridge: Cambridge University Press, 1982.

Sandel, Michael, ed., *Liberalism and Its Critics*, Oxford: Basil Blackwell, 1984.

Sayre-McCord, Geoffrey, ed., *Essays on Moral Realism*, Ithaca: Cornell University Press, 1988.

Scanlon, T.M., "Contractualism and utilitarianism", in: *Utilitarianism and beyond*, eds. Amartya Sen and Bernard Williams, Cambridge: Cambridge University Press, 1980.

Scheffler, Samuel, *The Rejection of Consequentialism*, Oxford: Clarendon Press, 1982.

Scruton, Roger, *The Meaning of Conservatism*, Harmondsworth, Eng.: Penguin Boooks, 1980.

Segers, Mary C. et al., *Elusive Equality: Liberalism, Affirmative Action, and Social Change in America*, Port Washington, N.Y.: Kennikot Press, 1983.

Sen, Amartya, and Bernard Williams, eds., *Utitarianism and Beyond*, Cambridge: Cambridge University Press, 1982.

Shapiro, Ian, *The Evolution of Rights in Liberal Theory*, Cambridge: Cambridge University Press, 1986.

Sher, George, *Desert*, Princeton, N.J.: Princeton University Press, 1987.

Sher, George, "Three Grades of Social Involvement", *Philosophy and Public Affairs*, Vol 18, No 2, Spring 1989, pp. 133-157.

Shklar, Judith N., *The Faces of Injustice*, New Haven: Yale University Press, 1990.

Shue, Henry, *Basic Rights: Subsistence, Affluence, and U.S. Foreign Policy*, Princeton: Princeton University Press, 1980.

Sidgwick, Henry, *The Methods of Ethics*, Chicago: The University of Chicago Press, 1962.

Sikora, Richard I., "Six Viewpoints for Assessing Egalitarian Distribution Schemes", *Ethics*, Vol 99, No 3, April 1989, pp. 492-502.

Singer, Peter, ed., *A Companion to Ethics*, Oxford: Basil Blackwell, 1991.

Smart, J.J.C. and Bernard Williams, *Utilitarianism; For and Against*, Cambridge: Cambridge University Press, 1973.

Sobel, Jordan Howard, "Maximizing, Optimizing, and Prospering", *Dialogue*, Vol XXVII, No 2, Summer 1988, pp. 233-262.

Soifer, Eldon, "On Justifying Liberal Neutrality", Paper delivered at the Canadian Philosophical Association Congress, Kingston, May 1991.

Soltan, Karol Edward, *The Causal Theory of Justice*, Berkeley: University of California Press, 1987.

Spitz, David, *The Real World of Liberalism*, Chicago: The University of Chicago Press, 1982.

Sterba, James P., *The Demands of Justice*, Notre Dame, Ind.: University of Notre Dame Press, 1980.

Stout, Jeffrey, *Ethics After Babel: The Languages Of Morals And Their Discontents*, Boston: Beacon Press, 1988.

Sumner, L.W., *The Moral Foundation of Rights*, Oxford: Clarendon Press, 1987.

Sumner, L.W., "Justice Contracted", *Dialogue*, Vol XXVI No 3, Autumn 1987, pp. 523-548.

Sumner, L.W., "Positive Sexism", in *Equal Opportunity*, ed. by Ellen Frankel Paul et al., New York: Basil Blackwell, 1987.

Tawney, R.H., *Equality*, London: Unwin, 1964.

Tawney, R.H., *The Acquisitive Society*, London: G.Bell and sons, 1921.

Taylor, Charles, Hegel and Modern Society, *Cambridge: Cambridge University Press*, 1979.

Tempkin, Larry S., "Inequality", *Philosophy and Public Affairs*, Vol 15, No 2, Spring 1986, pp. 99-121.

Thakur, Shivesh C., "Social Justice: An Informal Analysis", in: *International Philosophical Quarterly*, September 86, pp. 213-221.

Thurow, Lester C., "A Surge in Inequality", *Scientific American*, Vol 254 No 5, Mary 1987, pp. 30-37.

Tucker, D.F.B., *Marxism and Individualism*, Oxford: Basil Blackwell, 1980.

Vallentyne, Peter, ed., *Contractarianism and Rational Choice: Essays on David Gauthier's Morals By Agreement*, Cambridge: Cambridge University Press, 1991.

Walzer, Michael, *Spheres of Justice: A Defense of Pluralism and Equality*, New York: Basic Books, 1983.

Wedeking, Gary, "Is Mandatory Retirement Unfair Age Discrimination?", *Canadian Journal of Philosophy*, Vol 20, No 3, September 1990, pp. 321-334.

Weinreb, Lloyd L., *Natural Law and Justice*, Cambridge, Mass: Harvard University Press, 1987.

Wilhoit, Francis M., *The Quest for Equality in Freedom*, New Brunswick, N.J.: Transaction Books, 1979.

Williams, Bernard, *Problems of the Self*, Cambridge: Cambridge University Press, 1973.

Wittgenstein, Ludwig, *Philosophical Investigations*, 3rd ed., New York: Macmillan, 1968.

Winfield, Richard Dien, *Reason and Justice*, New York: State University of New York Press, 1988.

Wolgast, Elizabeth, *The Grammar of Justice*, Ithaca: Cornell University Press, 1987.

Young, Michael, *The Rise of the Meritocracy*, New York: Random House, 1959.

Zinn, Howard, *Justice In Everyday Life: The Way It Really Works*, New York: William Morrow & Company, 1974.