

# **THE MORALITY OF GROUPS**

Donald Edwin Franklin

PhD Thesis

University College London

2000

ProQuest Number: U642997

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest U642997

Published by ProQuest LLC(2015). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code.  
Microform Edition © ProQuest LLC.

ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

## **THE MORALITY OF GROUPS**

### **ABSTRACT OF THESIS**

The principle of the equality of human worth that underlies the political doctrine of impartiality, seems as pertinent to personal as to political morality. Yet we commonly not only tolerate but commend moral practices that are partial to family, friends and countryfolk.

Definition of this problem focuses upon the demands of fairness and of the recognition of equality, against the value of personal projects and loyalties. I survey attempts by communitarians and by liberals to address the problem by isolating the realm of impartiality from that of personal morality, and conclude that such attempts are largely unsuccessful, as are efforts by consequentialists to subsume partial obligations under impartial rules. Rather, I argue that there is indeed independent value in community and in relationships, value in the flourishing of groups constituted by the connexions forged in acts of loyalty, but that these values are in irresolvable tension with the recognition of the equality of human worth. This tension should exercise us both in our personal lives (such that we must justify loyalty by the intrinsic values that it constitutes or to which it contributes), and in the design of political systems, wherein the equality of human worth is seen to represent but one defeasible moral consideration amongst many.

The source of the false appeal of impartiality as a meta-ethical principle in political philosophy may lie with the plurality of visions of the good. Recognition of this plurality has generated both scepticism (with respect to "the good") and an implicit denigration of the values and ends that may conflict with impartiality ("the right"), including the values of community. I argue that that scepticism is not justified. I show that a realist interpretation for ethical judgments (using criteria for objectivity derived from the theory of meaning) is consistent with the heterogeneity of visions of the good, being anchored in the extent of continuing consensus in particular ethical judgments, (though this claim should, I suggest, be susceptible to an empirical research programme).

Realism in ethics, thus established, not only belies the sceptical foundation of impartialism, it is also shown to underpin the obligations of individuals to large groups. And the epistemic dependence upon the group as the forum of the search for moral truth, itself constitutes much of the intrinsic value that justifies obligations of loyalty to the group.

Donald Franklin    December 1999

## THE MORALITY OF GROUPS

<b><u>CONTENTS</u></b>	<b><u>PAGE</u></b>
<b>INTRODUCTION and SUMMARY</b>	<b>6</b>
<b>1 OF FAIRNESS, EQUALITY AND THE DOCTRINE OF IMPARTIALITY</b>	<b>20</b>
1A Fairness and Equality	21
1B The Fairness of Punishment	25
1C Moral Luck	32
1D The Defeasibility of the Maximand of Fairness	36
1E The Doctrine of Impartiality	39
<b>2 FAIRNESS AND IMPARTIALITY: THEIR SCOPE</b>	<b>44</b>
2A Indulging Partiality	44
i The Demandingness of Impartialist Morality	
ii Partiality may Rightly be Lawful without being Moral	
iii Rights and Duties	
iv Epistemic Limits to the Demands of Impartiality	
2B Partialist Obligations?	66
i Spontaneity and the Benevolent Emotions	
ii Projects and Commitments	
2C Conclusion	100
<b>3 INDIRECT JUSTIFICATIONS OF PARTIALITY</b>	<b>103</b>
3A Eudaemonic Instrumentalist Justifications	104
3B Inadequacy of Conditional Loyalty	110
i Exceptional Cases Dilemma	
ii Hare's Double Think	
3C Aesthetic Instrumentalist Justifications	123
i Maximising Happiness and Beauty	
ii The Epistemic Argument for Partiality Revisited	

3D	Conclusion	132
<b>4</b>	<b>HISTORICAL TIES AS SOURCES OF PARTIAL OBLIGATIONS</b>	<b>134</b>
4A	Contract	135
4B	Debts of Reciprocity	139
4C	Groups as Goods in Themselves: Communitarian Obligations	147
	i Intrinsic Value of Groups and Relationships	
	ii Obligations of Loyalty to Groups and Relationships	
	iii Whether Loyalty to Groups or Relationships rather than to Particular Individuals is Genuine Loyalty	
	iv Conditionality of the Intrinsic Value of Groups/Relationships upon their Goodness	
	v Criteria for Obligations of Loyalty towards Groups/Relationships	
<b>5</b>	<b>QUESTIONING IMPARTIALITY AND EQUAL CONCERN &amp; RESPECT, AND COMPROMISING FAIRNESS</b>	<b>162</b>
5A	The Doctrines of Impartiality and of Equal Concern and Respect Compromised	163
5B	The Basis for the Doctrine of Equal Concern in Political Philosophy	177
	i Value-based Defences of Liberal Impartiality	
	ii Procedural Defences of Liberal Impartiality	
	iii Sceptical Defences of Liberal Impartiality	
5C	Conclusion	191
<b>6</b>	<b>THE SCEPTICAL FOUNDATIONS OF IMPARTIALISM UNDERMINED</b>	<b>193</b>
6A	Meaning-theoretic Considerations relating to the Attribution of Objectivity to a Domain of Discourse	196
6B	A Public Language Test of Objectivity	209
6C	Objectivity in Moral Discourse #1: Focus on the Particular	218
6D	Objectivity in Moral Discourse #2: a Research Programme for Testing Error Theories and for Establishing which Moral and Descriptive Judgments Warrant attribution of Objectivity	223
	i Hard Cases	

ii	Vagueness	
iii	Conclusion regarding Objectivity in the Presence of Hard Cases and Vague Penumbrae	
iv	Error Theory	
v	Global Error	
6E	Objectivity in Moral Discourse #3: How Charity in Interpretation should Limit attribution of Immoral Judgments to Other Cultures	241
6F	Moral Objectivity is not Reducible to Descriptive Content	245
6G	The Possibility of Convergence of Moral Theories	250
6H	Implications for Partialism	254
<b>7</b>	<b>PARTIALITY TOWARDS LARGE GROUPS</b>	<b>258</b>
7A	The Surprising Significance of Individual Acts of Loyalty to Large Groups	259
i	Assessing Single Acts in Collective Contexts: Smoking and Voting	
ii	The Non-Linearity of Risk Aversion	
iii	The Non-Linearity of the Social Gain Function	
7B	The Value of Large Groups	280
i	The Group as Moral Forum	
ii	The Group as Forum for Exercise of Relational Virtues	
7C	How Much Partiality Should the Group Mandate?	285
7D	Membership Criteria for the Group	294
7E	Concluding Remarks	300
	Bibliography	303

# THE MORALITY OF GROUPS<sup>1</sup>

## INTRODUCTION AND SUMMARY

Impartiality is a central doctrine in contemporary political philosophy; were a theory to claim "that certain kinds of people just do not matter as much as others, then most people in the modern world would reject that theory immediately" (Kymlicka (1990), pp 4-5, quoted in context p.41 below). Yet we commonly not only tolerate but commend moral practices that are partial to family, friends and countryfolk: we discriminate between our group and others, and between other groups. This thesis explores this apparent contradiction.

The tension between the impartial political perspective and the anti-impartialist local or personal perspective has become a commonplace of modern ethical discourse. It has spawned camps of communitarians and Aristotelian virtue theorists who emphasise the value of local loyalty and who give weight to what seems urgent to individuals in their social context; these camps are set against the universalist justice tradition (to use O'Neill (1996)'s nomenclature<sup>2</sup>). Barry (1995, section 31)

---

<sup>1</sup> I am grateful to members of the Golders Green Kreis (Oliver Black, Jonny Watling, Anthony Gottlieb, Donny Peterson, Edmund Fawcett) and to my supervisor (Jo Wolff) for helpful discussion of various drafts of this work over many years. Chapter Six was published in much abbreviated form and discussed at the Kirchberg Wittgenstein symposium in 1998.

<sup>2</sup> For a useful list of advocates in either camp see footnotes 3 and 5 in Chapter 1 of O'Neill (1996). O'Neill points out that the partialist interpretation of Aristotle adopted by communitarians such as MacIntyre is contentious. She also uses the term "particularist" to describe the anti-impartialist school. However, this involves sliding from discussion of the content of morality (some countenance partiality to particular groups, others do not) to discussion of epistemology (some grant epistemic priority to particular moral claims, others grant it to universal claims). Whilst epistemic particularism has been used as a ground for anti-impartialism, I will argue in Chapter Six for a form of epistemic particularism that does not have this consequence. Hence I prefer to keep these ideas distinct.

characterises the dispute as a "battle between defenders and critics of impartiality".

The anti-impartialists deny the relevance or the coherence of the perspective from which impartialism makes its claims - attempting thereby to justify partialist obligations whilst nonetheless denying that there is any implicit claim of superiority for the community within which they hold sway.

Amongst liberal impartialists, awareness of the local, often partialist perspective has encouraged the emergence of "neutralist liberals" such as Rawls and Barry, who make space within their theories for individual projects and commitments, whilst building the priority of impartialist constraints (the "priority of the right over the good") into the design of political institutions. These constraints are as far as possible neutral between competing visions of the good. Leaving space for each man and woman to create their own commitments is seen as one aspect of the equality of respect that is each person's due.

Underlying this tolerance, however, there is (I suggest) a certain disdain of the partialist commitments that subsist at the local level, a disdain that is manifest in the subordination of the commitments to the impartialist framework. Certainly the particularity of individuals' visions of the good are not allowed to determine the impartialist rules. Nor, more surprisingly, is there any obligation upon individuals to render their vision of the good consistent with the impartialist frame, so long as the rules are not transgressed. (This too is a sign of condescension: the content of individuals' visions of the good seems not to matter sufficiently to warrant critique.)

This neat separation does make moral life easier: on the one hand, individuals can pursue their local projects immune from impartialist critique; on the other, the system of justice can be created without



addressing the issues of ultimate value that motivate individuals. Yet it is doubtful whether the bifurcation is coherent. Much as neutralist liberals wish to limit the ethical commitments of their systems, it seems undeniable that they have as their foundation recognition of the equality of human worth; how then can this truth be forgotten in evaluating our individual commitments and loyalties to others? The principle of the equality of human worth seems as pertinent to personal as to political morality, generating both an obligation to maximise fairness in distribution (material and social), and a rigorous constraint upon other moral maxims.

The neglected requirement to integrate the truths of political and personal morality poses a challenge not only to the neutralist stance of Rawlsian liberals, it also challenges communitarians and virtue theorists. How can they sustain commitment to an authentic local vision and yet ignore the existence of rival local visions and accounts of virtue competing for scarce resources? Recognition of rivals forces the construction of a moral critique of others' moral claims, a critique that must also be applied at home. It also forces acknowledgement that communitarian virtues often implicitly discriminate against other communities. Awareness of alternative moralities has led some anti-impartialists to despair of any objective truth; others simply ignore the problem.

Conversely, liberal impartialists are also challenged in their political neutrality by the moral insights that lie behind communitarianism and virtue theories, particularly the value of community and of relationships. The state is composed of individuals each of whom is seized with the importance of their own projects and commitments. Political philosophy must take these values seriously: it must explain why (and to what extent) these individuals (who include the officers of state) should subordinate their projects to a framework of impartialism. Ethical theory and political philosophy must aid the individual officer or citizen beset with conflicting moral demands from the state and from her personal commitments.

My thesis is that it is not possible fully to resolve the tension between impartiality and the set of values that implicitly underpins our partialist practices, nor indeed to reconcile either of these values with various consequentialist ends (such as those of the utilitarian). But neither is it possible to compartmentalise our life into different spheres to be governed by different moral precepts. Rather the moral world is both holistically integrated and ridden with conflict: in particular, the maximand of fairness (embodied in the duty to be impartial) is set against other maximands like happiness and community, forcing upon us painful compromises. Trade-offs between different ends (and between ends and deontological constraints) are inevitable. There is no simple distinction to be made between the realms of political and personal morality, nor between the right and the good. Liberal prescriptions of impartiality, and communitarian personal projects and commitments have alike to be tested against the same complex set of ethical maximands and constraints.

This picture allows an understanding of how we can be committed to the equality of human worth, and yet acknowledge myriad partialist obligations, some of which implicitly rank some individuals and groups as more valuable than others.

Much of what follows involves substantiating the claim that there is a genuine tension to be understood through integration of perspectives: that impartialists, anti-impartialists and consequentialists (who claim a reconciliation of all perspectives) have alike been guilty of ignoring or brushing aside commonplace moral dilemmas posed by conflicting impartialist, consequentialist and communitarian claims.

To bring out these tensions, it is necessary to clarify the ground and the nature of our commitment to equality of human worth, and to impartiality and fairness. Chapter One argues that the maximand of fairness involves both procedural constraints and an egalitarian objective (and that there is

some tension between these). Both these concepts are founded upon a primitive commitment to the equality of human worth.

It is however immediately apparent from our everyday practice that we are forced to compromise fairness in the name of other ends, including utility maximisation. Similarly, we evaluate humans differentially according to a range of criteria, whilst still respecting their underlying moral equality. A simplistic doctrine of impartiality that demands fairness at all costs, is thus not viable. It is suggested (merely *arguendo* - see below and Chapter Five) that the principle of impartiality is not truly compromised in these cases so long as the end for which compromise is sanctioned can be defined impartially.

This limited compromise of the ideal of equal respect might be adequate were it possible to encompass partialist commitments within an appropriately restricted impartialist framework. Chapters Two and Three examine attempts to reconcile the ideal to partialist practices:

- the first, adopted both by most communitarians and virtue theorists, and by neutralist liberals, is to claim that the doctrine of impartiality is after all essentially a principle of political philosophy, leaving scope for partiality in private life;
- the second, adopted by utilitarians and other consequentialists, is to attempt to show that partiality can be justified instrumentally -- analysis might prove some partial practices to be derivable from impartial principles (perhaps utilitarian) that accord equal concern and respect to all.

Chapter two explores the first option: the position of the anti-impartialists and of the neutralist-impartialists, who share a belief that personal and

political moralities can properly be isolated. Whilst I will argue that this seal can and should be pierced from either side, it has been the invasion of personal space by impartiality that has been resisted, and it is this resistance that is explored and exploded in this chapter.

For example, many claim that any impartial duty of beneficence would intrude upon every decision a person makes, forcing intolerable asceticism at a personal level. I argue that the onerousness of asceticism implies only that we should indulge partiality; selfless impartiality as an ideal cannot be dismissed merely by claiming that it is hard to achieve.

More positive defences of personal discretion against the doctrine of impartiality claim that we are obliged to give more weight to our emotions or to our own projects than is consistent with an external impartialist critique. Hence, impartiality must be limited to a moral superstructure to which only governments and their officials must adhere. However, these arguments are not compelling: just as reasons for action need not be consciously articulated, so actions may spring spontaneously from altruistic emotions and yet be moderated by an impartialist critique; personal projects too can and should be chosen and appraised in the light of an impartialist critique.

The conclusion is that no moral claim can escape the impartialist critique. It is nevertheless conceded that both the demandingness of impartiality and consideration of the pleasure arising from partiality, can justify a degree of indulgence of partiality. Yet the existence of obligations of partiality cannot be explained in this way.

Chapter Three considers the second option open to the defender of partialist practices: instrumentalist justifications. The duty to maximise fairness, though deriving from the principle of impartiality, must be set against the other moral ends, for example, happiness. Partialist obligations might then be justified instrumentally when they are motivated

by these objectives - which at least do not explicitly favour particular groups, and hence can be seen as according equal concern and respect to all.

Instrumental justifications in the end prove unsatisfactory: they are ineradicably conditional upon the contribution to the end specified outweighing the loss of welfare and fairness involved in partiality; yet it is unconditional loyalty<sup>3</sup> to individuals that we applaud.

Having failed to accommodate partial practices either by carving out a space for individual discretion, or by justifying the practices instrumentally, we are forced to explore what moral theorems or moral ideals do underpin unconditional loyalty to individuals or to groups, and yet stand up to impartialist critique.

In Chapter Four, I suggest that unconditional loyalty can be justified where the circumstance that identifies that a particular individual or group is properly to be favoured is historic, and hence irrevocable. Recognition of the intrinsic value of (some) groups explains backward-looking obligations upon individuals to support each other, for groups are constituted of historic connexions between their individual members, and will flourish only if members recognise obligations towards their groups and towards each other. Rooting loyalty in the intrinsic value of good groups also renders it conditional upon the groups being good (we do not recognise all groups as good) and upon individual actions contributing to

---

<sup>3</sup> To clarify: unconditional loyalty need not be infeasible. Rather it is unconditional in the sense that it remains a motivating moral force upon someone in all circumstances, even when it is overridden by other moral claims. The contrast is with loyalty that is conditional upon some other end, such that if loyalty would not serve that end, the commitment would evaporate. Barry (1995) rails against those who claim that "you cannot really care for somebody unless your care is completely unconditional" (p.255). But his arguments demonstrate merely that care can be compromised - i.e. that the commitment is defeasible.

the welfare of the group. Nevertheless, this constitutive conditionality is consistent with partial practices in a way that instrumental conditionality is not.

This discussion establishes alternative sources of value that underpin partial practices, but these conflict with the maximand of fairness: membership of a group or relationship is seen to justify unequal treatment.

The implications for the doctrine of impartiality are considered in Chapter Five: the seal between the personal and the political sphere is here breached in the opposite direction; personal commitment challenges impartiality. The challenge presented to the liberal impartialists is the more potent in that the alternative maximands that are in competition with fairness are shown to include those of the consequentialists as well as those of the communitarians.

It turns out to be impossible to define a principle of non-discrimination that marks a line between objectionable and unobjectionable discrimination. If the doctrine of impartiality is taken to sanction practices if and only if they are instrumental to an end that can be stated impartially, it will exclude practices that are universally accepted on utilitarian grounds (as well as others that could be justified from a communitarian stance) whilst failing to exclude other practices that are repugnantly discriminatory.

An alternative interpretation of the doctrine of impartiality is suggested, that recognises our commonplace willingness to compromise fairness, and demands merely that we embody our conviction that all are of equal worth by avoiding practices that undermine self esteem. Yet here too our confirmed practice is to compromise even self-esteem for other ends.

I conclude that we have no reason to embody the goal of demonstrating an equal valuation of all humans into a doctrine to be respected regardless of consequential compromises with attainment of other goals. Rather it seems that what matters is both how much fairness is compromised, and to what extent practices undermine self esteem. These are important concerns, but not ones that trump other considerations. Thus the doctrine of equality of human worth is seen to underpin two important maximands but not, as many suppose, to provide an infeasible constraint upon moral practices.

In effect, in elevating the doctrine of impartiality to the status of a meta-principle liberal political philosophy values life and freedom irrespective of the uses to which they are put. I argue that although regarding the maximand of fairness, or regarding individuals' status as moral agents, as ends rather than means, each individual matters equally; yet we must also recognise that in their capacity to contribute to communitarian, aesthetic and even utilitarian ends some people actually matter more than others, justifying discrimination in their favour.

If fairness has had a stronger claim upon the political philosopher than other ends, it may be due merely to its congruence with the pragmatic benefits of tolerance. Tolerance may be of considerable importance, in underpinning personal autonomy and fostering social harmony, two independent ends; but, even combined, these ends do not justify tolerance irrespective of what flourishes beneath it. Intolerance and discrimination can in principle be justified in order to promote other aspects of the good.

What is the source of the doctrine of impartiality, of the claim that the right has priority over the good? Many contemporary moral philosophers take their cue from recognition of the plurality of visions of the good. Neutralist liberals have reacted by eschewing what Rawls calls comprehensive theories of the good, aiming rather to create the minimum just framework

that allows to each the freedom to develop their life according to their own theory of the good. One response to pluralism is scepticism, and a limited scepticism is indeed implicit or explicit in neutralist liberal political philosophy.

Communitarians and virtue theorists have paradoxically found value in the very diversity of visions of the good: that each live according to the value system of the community with whom they identify is a form of Aristotelian flourishing. However, the meta-ethics that seems to underpin this approach is a relativist one.

How can one avoid moral scepticism or relativism in the face of so many competing visions of the good?

Chapter Six attempts to address this question in order to force a confrontation between conflicting communitarian and impartialist values. If moral cognitivism is correct, then moral conflict must be faced as a cognitive challenge.

Moral objectivity is established by the existence of a group consensus as to the prescriptive use of moral terms, including for new cases, a consensus that cannot be explained by any reduction to terms with purely descriptive content. The plausibility of moral objectivism is greatly enhanced by noticing the strong parallels between the sources of objectivity in science and those in morality, and by attention to the consensus in both spheres that attends ascription of particular events to moral and scientific categories. The appearance of pervasive conflict is explained by value pluralism together with a distorting focus upon hard cases.

Ultimately, it is an empirical matter (worthy of research) whether the consensus attending ascription of moral terms is sufficient to justify moral cognitivism. Nonetheless, it is plausible (I suggest) that the extent of



consensus with respect to myriad particular moral judgments about core cases cannot be explained other than by the moral truth that the community severally and jointly perceives. There is, however, often conflict between different moral values, and these conflicts cannot often be resolved by reference to moral theory, because theory is underdetermined by the moral judgments that form the objective core. Conflict arises both when individuals find themselves pulled in different directions by different legitimate moral claims; and where different individuals and groups, working each in pursuit of a legitimate moral goal, find their projects mutually inconsistent.

This then is the nature of the conflict between impartialist claims of fairness and equal concern and communitarian demands of loyalty. The strength of each must be tested against the evidence of our consistent judgments, subject to errors as they are, about particular cases.

In Chapter Seven I explore the peculiar difficulties associated with establishing obligations of loyalty to large groups, absent an explicit social contract, given that most individuals' actions in support of the larger group appear to have negligible influence upon the group or upon its projects.

I suggest that individuals' actions in a social context are best understood analogously to the way a single person understands a string of her own actions that will collectively issue in a desired result, but none of which separately appears significant. It is our epistemic dependence upon the group (as discussed in Chapter Six), that underpins the conclusion that the rightness of an action should be assessed within its social context, just as an individual views her acts within the context of her plans over time, and therefore that individuals can be bound by obligations of loyalty to large groups.

However, the strength of any obligations of loyalty remains dependent upon two further questions: the value of the larger group; and the importance of acts of loyalty to sustaining it.

The link between objectivity and the moral consensus of the group provides a deeper explanation of the value of a group as the only forum within which the moral enterprise can be pursued, which in turn provides a stronger justification for the group to impose upon its members those obligations needed to perpetuate it. Although loyalty is unfair to those outside the group, and although for that reason amongst others loyalty will compromise some moral ends, the value of the group as moral forum and as a space within which various relational virtues can be exercised, may be sufficient to justify the imposition of (more or less) unconditional obligations, which the individual, to the extent that she recognises the objectivity of the group's assessment and has no reason to suspect that the community is in error on this occasion, is obliged to respect.

The nature of any obligations to the group depends upon the value of the particular good, both as moral forum and as embodiment of relational virtues, of the specific group. In general, I conclude that partialist exceptions to the moral cogency of considerations of fairness and equal concern and respect can be justified in one of the following ways:

- loyal practices may be constitutive of a group's flourishing
- they may be necessary to enable the community to pursue its ends
- they may be necessary to preserve the community itself.

I then turn to the implications for immigration and emigration (entry and exit) to the group, and conclude (contrary to practice prevailing in most democracies) that: entry to the group should be open to all who are willing and able to contribute to the flourishing of the group, exit from the group may be stigmatised in order to ensure fulfilment of obligations to the group.

I conclude with some observations on the nature of the conceptual enquiry that should be pursued by political philosophers. Their object in the context of a particular cultural heritage should be to fuel the inevitable struggle to find the correct balance for that community between the conflicting values of impartiality and of loyalty. In cohesive societies that embody goods of community, highly discriminatory practices may be justified. Societies that emphasise liberal individualism, on the other hand, must be confident that the cost in community values is justified.

The segmentation of the discussion into chapters is inevitably somewhat arbitrary; there are a number of themes that recur. It may be helpful to pick some of them out here:

- the existence of conflicting moral ideals, requiring difficult judgments and compromise: the procedural and perfectionist notions of fairness already discussed, different consequentialist notions of the good (in Chapter Three), the value of flourishing groups (in Chapters Four and Seven). This is in contrast: to political philosophies that denigrate all conceptions of the good relative to an overarching impartial liberalism; to utilitarianism and other moral systems that avoid recognition of value conflict by subordinating all to a single value; and to the blinkered focus of some communitarians upon the value system of their own community, refusing to recognise the existence of alternatives.
  
- the impact of uncertainty (about facts and hence about the evaluation of choices) upon moral structure. (Uncertainty is shown to leave space for partiality, in Chapters Two and Three - though not to the extent that would arise from incommensurability; uncertainty provides part of the imperative of tolerance, in Chapter Five; uncertainty gives

to moral discourse the misleading appearance of subjectivity, in Chapter Six.)

- distinguishing the ontology from the epistemology of moral value: the inadequacy of attempts to justify partiality as a personal prerogative on the basis of the importance of personal evaluative judgment (Chapter Two); and the fuller discussion of moral epistemology in Chapter Six;
- exposing the compatibility of autonomy with the impartialist critique: the wrongheaded notion that we truly have moral autonomy only if we somehow determine what is right and/or good (in Chapter Two); the simple confusion between legal and moral constraints that may be behind this notion (also in Chapter Two); autonomy as an end issues in a prescription of defeasible pragmatic tolerance rather than principled neutralism (Chapter Five)
- exploring the role of empirical research in moral philosophy: the heroic assumptions made by rule utilitarians about the balance of utilitarian advantage from pursuing partialist rules (in Chapter Three); the use of consensus as a test of objectivity of moral judgments (Chapter Six); empirical investigations regarding the degree of partiality needed to sustain good groups and to justify their migration policies (Chapter Seven).

## MORALITY OF GROUPS

### 1 OF FAIRNESS, EQUALITY AND THE DOCTRINE OF IMPARTIALITY

Occasionally even well attested moral theories conflict with each other:

- (I) Kinship or co-membership of a group is not morally relevant -- people should be treated as equals.
  
- (P) People are often right or even obliged to be loyal to or to favour other members of their group, whether family, community or state.

Both these statements generalise a wealth of uncontroversial judgments regarding particular actions, but there is an evident conflict between them.

This chapter seeks to specify more precisely the conflict between partiality and impartiality, noting particularly the canonization of impartiality in political philosophy, and the tension between this doctrine and our endorsement of partiality when practised by individuals (notably in the form of loyalty)<sup>4</sup>.

---

4

The partialist practices that motivate this discussion are not limited to those concerning the individual as agent: for instance almost every contemporary state falls foul of impartiality in its treatment of foreigners. Nevertheless, partialist practices by sovereign states have found fewer defenders. An example of the liberal attitude to such partiality was provided by a leader in the London Financial Times of 2nd June 1994:

"One of the peculiarities of 20th century social security systems is that what might seem a universal moral obligation upon the rich to help the poor is turned into a narrower obligation to help fellow citizens. This treatment of national income as a form of private property is politically inevitable. Its moral justification is rather harder to detect."

Note the assumption that were national income really private property, then such partialist distribution would be justified: implicit is the notion that

What is the ground of impartiality ((I) above): why should people be treated as equals?

The ground for impartiality must be a quality that all share equally. The only plausible candidate is the quality of humanity itself. Impartiality demands that all human properties be considered morally irrelevant (in themselves) except the property of being a human being. Being human is not only self-evidently morally relevant but it is also invariant in intensity between human beings. Hence, only the ascription of moral pre-eminence to this one quality could support a conclusion that we must treat every one as equals.

We may be tempted to define the property of humanity: to specify what it is that differentiates every human being from every other creature without varying between them. Such definition is difficult, but would probably relate to moral agency. It would certainly produce hard borderline cases. However, the concept of humanity is sufficiently entrenched for us not to be perturbed by lack of definition; we can take the property as primitive.

Hence we have a doctrine of equality of human worth, or a principle of impartiality, or of equal concern and respect, (I will use these expressions interchangeably in this chapter to designate the same underlying principle). What follows? The obvious direct implications are for allocation policies at a personal or political level; in particular the doctrine suggests that both in constructing a theory of justice and in making practical decisions that affect the welfare of others, one should strive for fairness.

## **1A      Fairness and Equality**

---

individuals, who do have private property, can use it partially without challenge. This assumption is challenged below.

Fairness can be characterised either positively in terms of an ideal of absolute equality of welfare, or negatively as an injunction not to sacrifice the welfare of one individual to favour another (let alone for some impersonal end). These are the extreme positions, and they often conflict: I suggest that the definitions of fairness adopted by different philosophers, politicians and others can be seen as compromises struck at different positions between these positive and negative views of fairness, both of which have considerable intuitive appeal.

On the first view, the egalitarian view of fairness, we all have a *pro tanto* obligation to redistribute in favour of the deprived - whether deprived of goods, liberties or opportunities. When we adopt this perspective we consider all individuals afresh: all are of equal worth and none has past entitlement or debt; hence all should have the same. Further we take an expansive view of our jurisdiction: we are entitled to determine what the distribution should be and to legislate accordingly.

According to the second, the procedural view of fairness, we assume that the current distribution of resources is the result of free and fair exchange or is otherwise just, and we abjure redistribution on the grounds that any redistribution would involve favouring one individual over another. When we adopt this perspective, we recognise the importance of history in determining entitlement. If individuals have different endowments, it is the result of their choices, their industry, their G-d-given talents, the free choices of others who have bestowed gifts upon them. These are all legitimate grounds of entitlement; differential welfare thus occasioned is fair. Further, we take a more modest view of our jurisdiction<sup>5</sup>: we assume that endowments are fair unless we have grounds for assuming

---

<sup>5</sup> A more interventionist view would be possible here too, to rectify allocations that were the result of illicit transactions, but is impractical because of the informational requirements involved in determining entitlements.

otherwise, and we minimise our interference with the existing distribution accordingly.

The two views may be underpinned by different views of personal identity: the egalitarian view by a conception of the self as evanescent, the procedural view would stress the continuity of the self, and its responsibility for the consequences of earlier action. However, the egalitarian position may also be founded upon a compassionate view of wasted opportunities. (The link between fairness and conceptions of personal responsibility is discussed a little further in the next section.)

This dichotomy in understanding of fairness is obvious in the political arena between capitalists and socialists, and in the philosophical arena between libertarians and liberals. It is also implicit, though less starkly, in the debate within the liberal camp as to whether we should hope or strive for equality of welfare or merely of opportunity. Those adopting equality of welfare as their aim may focus exclusively upon the egalitarian conception of fairness; alternatively they may deny that equality of opportunity is possible given different distributions of natural talents and industry, so that even if in principle they would hold people responsible for their own choices, in practice an egalitarian pursuit of equal outcomes is the best option.

Those who opt for equality of opportunity as their goal are allowing the procedural conception of fairness to moderate the egalitarian aim: redistribution that would put right the outcomes of market processes are deemed an unfair intrusion upon free exchange. The more extreme view, beyond the intermediate position represented by opportunity-egalitarianism, is a purely negative and procedural characterisation of fairness; it claims that redistribution to attain equality of welfare would be unfair not only to the prudent and industrious, but to any whom the fates have favoured. The negative procedural view asks whence the authority for such meddling with natural endowments?



Although in considerable tension, these two notions of fairness both capture elements of moral truth. Hence, I shall consider the theorems of fairness to be twofold:

(f-) "Do not diminish the welfare of another, without their consent, for any end"

(f+) "Raise the welfare of the least advantaged"<sup>6</sup>

As we have seen, the interpretation of f+ (whether disadvantage is defined in terms of overall happiness, or merely access to resources) and the relative weight accorded to f- and f+ will depend upon whether people are thought responsible for their own preferences and choices: only if they are not, can a goal of equal happiness be justified. Answers determine whether fairness generates a liberal-egalitarian, a liberal or a libertarian position (respectively, Dworkin, Rawls or Nozick). Nevertheless, even a libertarian may subscribe to f+ as a *ceteris paribus* rule for the discretionary distribution of personal resources.

Underlying (f-) and (f+), which together embody the injunction to be fair (or to be treated as equals), lies an argument from the equality of human worth:

D All persons are of equal worth, therefore they all deserve to be treated as equals.

---

<sup>6</sup> If this principle is interpreted dynamically, so that the long term interests of the least advantaged are to be maximised, even if that involves increasing or sustaining inequalities, it becomes equivalent to Rawls' difference principle. Like that principle, a more precise statement of it would involve an obligation to raise the welfare of the slightly better off, once the position of the least well off had been maximised. This refinement Rawls calls "the lexical difference principle". (See Rawls 1972 Section 13).

Without such a foundation, the maximand of fairness would seem unmotivated: why worry about fair treatment between beings unless they have equal claims?

This principle (D) employs a foundational sense of desert that is distinguished in common parlance by the qualifier "real" or "true" from deserts deriving from endowments that are not in their turn deserved. Thus although someone may deserve to win a beauty contest because he is very beautiful, there is a sense in which we say that, as he has no greater human worth, he does not really deserve to be beautiful, and thus he does not really deserve to be rewarded for it. (Though this is to be understood as consistent with the view expressed in f- that it would also be unfair to take away the rewards of the beauty with which he happens to be endowed.)

## 1B The Fairness of Punishment

Let us test this against intuitions regarding punishment and reward - where inequality appears to be fully justified. Is punishment ever really-deserved? When society punishes a criminal, is it being unfair to the criminal?

If every person is due the same because each is of equal human worth, how can we punish the wrongdoer? Perhaps punishment is justified on merely consequentialist grounds: it is unfair on those punished, but is nonetheless justified by the greater good. But this is not our intuition: it is not unfair to punish the guilty, whilst it is extremely unfair - unjust - to punish the innocent. (We may on occasion decide that we should punish the innocent, if the consequentialist benefits are very great, but only in tragic circumstances -- we know we are doing wrong by her.) Indeed we would prefer to punish the guilty even if consequentialist considerations (e.g. where the public have an ineradicably mistaken view of who is guilty) indicate harming the innocent.

Thus, an adequate theory of punishment must have a retributionist element -- it must explain why we are never as happy to punish the innocent as the guilty; guilt is a necessary condition for punishment. (However, it may not be a sufficient condition: some consequentialist gain from punishment is also often considered necessary to justify punishment - see below). And an adequate theory of justice or fairness must explain why we are happy to discriminate against the guilty.

Can a retributive theory of punishment be subsumed under a more general principle that does respect criminals equally with the law-abiding? Or are we rather forced to reconsider (D) and recognise that criminals are not considered of equal worth, justifying their treatment as less-than-equals?

Our predilection for punishing the guilty can be squared with the doctrine of equal worth on the basis of the duty to make good torts: damage caused to society by law-breaking requires restitution, and so long as the damage was the result of free choice, the restitution should come from the agent who caused the damage. The nature of the restitution may be determined on consequentialist grounds (e.g. to deter others) but the fairness is underpinned by the procedural notion of fairness, modified to allow deprivations into welfare in reciprocation for like deprivations:

(f-') "Do not diminish the happiness of another, without their consent, for any end, except to make good damage freely done by them to the welfare of others"

Sher (1987) appears to favour a theory of this sort. He suggests that as the guilty have taken from society by breaking the law, justice demands that they repay their debt. However, he interprets this as an egalitarian principle (like f+): criminals are being treated as equals in that they are presumed to be entitled to the same as everyone else, but they have taken more, so they must give some back. Sher acknowledges that this is

an odd sort of argument as other aspects of the distributive balance cannot be taken into account. For instance, if the criminal is poor and has suffered previously, it may be that whatever she gained from the criminal activity in which she indulged, she may still on balance be less well off than average, a relative deprivation that will be exacerbated by punishment.

Sher defends his interpretation by arguing that there can be no trade-off between lack of other goods and the enjoyment of excess liberty, as the two are incommensurable. Thus if the criminal has taken liberties with society, then she must repay in kind, through having her liberty curtailed as a deterrent to others, irrespective of any disadvantages she may suffer in the realm of goods. This is not an intuitive account: it is not her excessive stock of liberties - whatever that might mean - that justifies the criminal's punishment, but the fact that she has freely chosen to harm others.

Thus, a more plausible thesis is obtained if we gloss "taking liberties" as implying that she has damaged society, perhaps by eroding the law abiding culture, rather than, as Sher's locution suggests, implying that the criminal has benefited in some way from excess liberty. (The liberty that the criminal has enjoyed may not seem much compared with the freedom of the wealthy.)

If we attend to the damage that she has done to society, the punishment of the criminal will seem just according to the principle of restitution in  $f^-$ . Nevertheless, our egalitarian sense of fairness  $f^+$ , whereby we strive for equality of outcomes/opportunities for all, remains offended by the further degradation of an impecunious criminal. This is correct.

An alternative explanation for our retributionist preference for punishing the guilty might stem from a view that the guilty are of lesser value than others and thus do not merit equal treatment (akin to the argument that

allows us to use animals in ways that we would not use humans).

Because we value the criminal less, we are willing to harm her to deter others. On this view, punishment of the guilty is fair on a modified view of f+, which shifts from an egalitarian to a moralistic conception of fairness:

(f+)'            "Adjust welfare/opportunities towards a distribution commensurate with real-desert"

This maxim is consistent with Ross's view that allocation in accordance with true virtue is a primitive end:

"If we compare two imaginary states of the universe, alike in the total amounts of virtue and vice and of pleasure and pain present in the two, but in one of which the virtuous were all happy and the vicious miserable, while in the other the virtuous were miserable and the vicious happy, very few people would hesitate to say that the first was a much better state of the universe than the second. It would seem then that, besides virtue and pleasure, we must recognize, as a third independent good, the apportionment of pleasure and pain to the virtuous and the vicious respectively. And it is on the recognition of this as a separate good that the recognition of the duty of justice ... rests."

Ross (1930, Chapter V, p 138)

Ross's argument implies that, at least if the deterrent effect of punishment will raise general welfare correspondingly (so that there is no overall loss in pleasure), the diminution in the pleasure of the wicked is just. I am suggesting that this view embodies a lower valuation placed upon the wicked. (Note that this interpretation of f+ does imply that agents bear some responsibility for their choices, i.e. they had the choice to be vicious or virtuous and bear responsibility for the consequential shift in their entitlements. This is contrary to the interpretation of egalitarian fairness offered above.)

Rawls (1972, section 48) however rejects Ross's argument. He suggests that it is no part of justice to adjust allocations proportionately to moral virtue. Rawls reaches this conclusion via an argument that is embedded in his own theory: moral virtue is defined to be conceptually derivative from justice as fairness; hence justice as fairness cannot itself be defined in terms of moral virtue. Nevertheless, Rawls' intuition is correct even if moral virtue is defined independently - as it can be if the good, and hence striving for the good, is independently defined. The intuition is that it is actions that are good and bad, and that good and bad actions do not undermine the fundamental equality of human worth. This insight is deeply embedded in religious tradition, particularly in the life-long possibility of repentance. Of course the religious tradition is also concerned with the reward of the righteous, as if righteousness was a characteristic of persons rather than of actions, but there is at least ambivalence here too: virtue is its own reward, and there is something repugnant associated with the notion that the virtuous should reap reward for their virtue.

That the equality of human worth is fundamental to our sensibilities is manifest in our attitude to retribution. We find it hard to justify punishment in the absence of consequentialist benefit from punishment. If we were pursuing a Rossian end of distribution commensurate with merit (as in f+'), and we valued the criminal less, we might wish to diminish her happiness even in the absence of benefit to others. However, if our concern is merely that she should rectify damage to society, our reluctance to punish her in the absence of consequentialist gain is explained.

(Is Ross just wrong in his intuition that the world is a better place if rewards are commensurate with virtue? He is certainly not alone in questioning the justice of a world in which "all the workers of iniquity do flourish" (Psalm 92). However, we may interpret the intuition as reflecting the torts-based view that damages should be made good (f-' ) rather than that that the wicked should be punished (f+'.)

We are reluctant to agree with Kant that a man should be punished even if his society is to be dissolved and scattered across the earth (quoted in Honderich, 1969, Chapter 2, from Kant's Philosophy of Law). If no consequentialist ends would be served by punishing him, then there is no need to make him suffer.

We might however explain the few cases where we do feel strongly retributionist about a criminal using the diminished value argument (f+'): in these cases our valuation of the criminal slips into negative territory. We are willing to punish him even if there are no justifying consequences (and to kill him, even if lesser punishment might be reformatory). Nowadays we seem to reserve this sort of judgment to mass murderers of great renown -- Eichmann or Ceaucescu.

We have become more selective in this judgment since Kant's day no doubt as a result of the invention of psychology and sociology; avenues by which evil can be explained away have been multiplied. Thus we are uncertain as to whether the agent chose the wicked path freely and hence as to whether she really-deserves the "evil" tag.

Ironically, the very psychological explanations that mitigate the wickedness of criminals also tend to undermine our valuation of them as moral agents: we substitute a causal explanation for their actions for free will. So if the standard for full human worth is moral agency (which perhaps implies personhood), then the criminal who does not fall short of this standard because she is an immoral or amoral agent, will fall short because she is less than a free agent. Such agents are not of negative value, but they may fall short of full personhood, or full coverage by the injunctions to be fair.

In extreme cases, where we label criminals as psychotic, we explicitly judge them as less than persons because of their lack of self control; and

on these grounds we can justify their incarceration for the greater good. In such cases too, fairness might not be compromised by unequal treatment, for we may judge the agents to be unequal<sup>7</sup>.

Psychology and sociology have also affected our judgment of ourselves; so we are no longer confident of what the standard of free agency should be. Perhaps we are equally psychologically constrained in our benevolence as the criminals are in their malevolence; indeed they may be more deserving than we are in that they may have used what free will they have to restrict their evil-doing. This uncertainty in the attribution of freedom, and hence uncertainty in assessment of real-desert, would reinforce our reluctance to punish (except when there is some consequentialist end served thereby) even were we convinced by Ross's argument.

However, we normally presume that people are free moral agents, with intrinsic human worth. We therefore treat them as equals unless their actions lead us to conclude that they have damaged others so that we can fairly exact restitution (according to f-'). (Of course we do value some people more than others in other contexts, as will be discussed below: but there is a sense in which all have equal worth, and it is this sense that is crucial for the maximands of fairness.)

The reasoning of this section could be applied also to explain the way we deal with awards for valour and other good deeds.

---

7

"Landru (assuming that he really did kill his women) may be pardoned if he did so from financial motives, which it is possible to resist, but not if it was from irresistible sadism." (Marcel Proust, Remembrance of Things Past, The Captive)

A financially motivated murder would certainly merit punishment, but it would not permanently reduce someone's moral status as would a diagnosis of compulsion.



According to f-' we will be willing or obliged to reward merit because of the obligations created by unsolicited kindnesses. (Ross would add - but Rawls would rightly reject - that it is the greater value that we attribute to the freely righteous that obliges us in fairness to increase their allocation.)

To sum up, real deserts do vary between people, because of the good or evil impact that their actions have on others, which can justify restitutive claims, of punishment and of reciprocation. However, that aside, we assume all to be equally deserving, and hence derive the principle of the equality of human worth, and the maxims of fairness.

### **1C Moral Luck**

The maxim of fairness f-' (and also f+' on the Rossian view) enjoins treatment in accordance with real desert, whilst real-deserts reflect free choice. There is in this analysis an underlying assumption of equality of moral opportunity (at least for those whose will is not disturbed by mental illness). In this section I defend this assumption, and the concomitant explanation of how theory of punishment can be integrated into our conception of fairness, against an attack from Bernard Williams.

What is praiseworthy and deserving of reward according to this notion of human worth is the good a person has done or achieved where she could have chosen otherwise. And conversely for blame and punishment. Each should be judged only in relation to the options that were open to her.

Bernard Williams emphasises the importance of this conception of human value at the beginning of his essay on Moral Luck - the purpose of linking value with the correct exercise of free will is to insulate the moral value of a man's life (i.e. its worth as a human life) from luck. In the eyes of morality every one has equal value because all equally have good action open to them. And if moral value "possesses some special,

indeed supreme, kind of dignity or importance" then the thought that this value, unlike others is accessible to all rational agents will allow one "to grasp not only morality's immunity to luck, but one's own partial immunity to luck through morality" (Williams, 1976a, p.21).

But Williams himself does not accept that morality itself, let alone the broad evaluation of life which he advocates, can be rendered immune to luck. He argues that "the aim of making morality immune to luck is bound to be disappointed....the form of the point which is most familiar from discussions of free will is that the dispositions of morality, however far back they are placed in the direction of motive and intention, are as 'conditioned' as anything else ... the bitter truth (I take it to be both) [is] that morality is subject, after all, to constitutive luck."(loc.cit).

Williams is surely right that there are legitimate assessments of humans that are dependent upon constitutive luck, and these assessments may underpin value judgments ("she is lovely"). But by claiming that even the opportunity for virtuous action is unequal - subject to constitutive luck - Williams appears to attack the foundational sense in which we are all equal as moral agents.

However, Williams' argument would only be cogent had the free will debate been resolved in such a way as to undermine completely the justified attribution of praise or blame. *Pace* Williams, if we do believe that man possesses freedom of will in a sense strong enough to be compatible not only with locutions such as "could have done otherwise" (the minimum requirement of the compatibilists), but also with the institutions of praise or blame, then it is possible to proportion claims of restitution to people simply according to whether they have exercised their free will well or ill. And it is this metric of evaluation that coheres with the equality of human worth and the concept of real desert. Of course it is difficult to determine real desert but we can acknowledge uncertainty

about the claims of tort and still be clear about the ultimate equality of human worth.

I claimed earlier that it was mere humanity that was the source of value. Here I appear to be claiming that it is moral agency, and the equal opportunity that we all have to be good or bad, that underlies human worth. This is not an equivalent doctrine: there are those who through handicap or early death never attain moral opportunity. Should we count them less than human?

This is to misconstrue the dispute with Williams. My claim is that were Williams right, were it the case that some were constitutively morally better, then the doctrine of equality of human worth would be impossible to sustain. The dimension of moral virtue trumps the dimension of humanity, such that even if it were granted that we are all equally human, that fact would no longer carry weight in discussing justice claims. If Williams is wrong, however, so that no one is constitutively morally more or less virtuous, then equality of human worth is at least a plausible doctrine, one that can underpin the maxims of fairness. Even the wicked have the possibility of repentance and redemption, and even the virtuous may degenerate: deviation from the norm does not represent intrinsic change. The very young and the incapacitated are different - but they are neither better nor worse on that account, merely innocent.

It is true that we do differentiate intrinsic human value, notwithstanding the equality of human worth, and even when there are no actions really deserving praise. But this demonstrates only an ambivalence towards the doctrine of equality of human worth. For example, we distinguish righteousness from real-desert, and sometimes value righteousness even when it is constitutive.

Suppose Albert Schweitzer was compulsively altruistic from birth, we would perhaps say that he did not really-deserve praise. Indeed,

someone who is righteous may at the same time be thoroughly reprehensible, for it may be that *he* could have been much better. Yet, we would still say Schweitzer was a very good person. We value him for his beneficence, irrespective of how praiseworthy he is for being beneficent.

Williams claims an inconsistency between the attitude of morality and our evaluative practices: how can we evaluate people for virtues and vices they could not help and yet claim that moral virtue is the product of choice? But the tension that he reveals is commonplace: it is no more remarkable than that we value intelligence and strength despite the obvious constitutive luck involved in the possession of these virtues. There is a virtue of righteousness that is related but not equivalent to real-desert. We evaluate people according to their righteousness and according to their other virtues, and separately we evaluate their actions according to the correctness of agents' exercise of free choice (in so far as we can determine this) - and their consequential real-deserts - from a starting point of equality.

Morality in Williams's sense does have a special concern with real-desert, which remains a coherent notion notwithstanding the existence of "moral luck". There is a form of praise (blame) that we reserve for those whose righteousness<sup>8</sup> (criminality) is an expression solely of the exercise

---

<sup>8</sup> Contrast:

Marlow's observation in Joseph Conrad's *Chance* (Part II, Ch.3): "Was the girl born to be a victim; to be always disliked and crushed as if she were too fine for this world? Or too luckless -- since that also is often counted as sin."

with:

"But she was not even grateful to him for it -- it seemed to her that Pierre did not have to make an effort to be good: it appeared to come so naturally to him that there was no merit in his kindness"(Tolstoy, *War and Peace* Book 3 Part 1 Ch.17)]

Conrad's fatalism is consistent with Williams' notion of moral luck;

of free will; and in this sense we may sometimes say that the able do not really-deserve their rewards, that it is not fair (in the f+ or f+' egalitarian senses) that they have a better or worse life than those who are less well endowed who have tried just as hard. This attitude is particularly important when we are considering people's equal entitlements to distributive justice, and also when justifying punishment, for in the case of punishment we are particularly interested in real desert, or culpability.<sup>9</sup>

Behind this is the egalitarian truth embodied in the f-': it is only the exercise of freedom with consequential benefit or cost to others that can justify divergence from equal treatment consistent with principles of fairness. Although, as Williams observes, there are metrics of evaluation that assign different values to different people, there is nevertheless a coherent enough notion of equality of moral opportunity to underpin the doctrine of equality of human worth.

#### **1D The Defeasibility of the Maximand of Fairness**

Notwithstanding their foundation in the principle of equal concern and respect and in D, it is evident that the maxims of fairness are defeasible: there are many circumstances in which we compromise fairness.

---

Tolstoy is emphasising a dimension of evaluation that gives us all equal moral opportunity.

<sup>9</sup> Of course, we do sometimes say that those who are virtuous through constitutive luck do deserve their rewards simply because they possess the virtues for which the rewards are designed. This may reflect the f- procedural sense of fairness: we are defending their right to retain the fruits of their endowments. Our use of "desert" in this less exacting sense even to justify positively favouring those with "constitutive luck" is also evidence of our limited interest in fairness and our willingness to evaluate people as means. See further in Chapter Five. Sher (1987) has an exhaustive analysis of the different forms of desert, few of which rank as real-deserts in the sense used here.

For example, doctors devote more resources to the ill. This may be consistent with impartiality directly if devoting more resources to the ill also happens to maximise fairness: for instance if the ill also happen to be worse off overall than other less ill potential beneficiaries of the doctors' attention. However, if this is not the case, then favouring the ill will only be acceptable to the strict impartialist because such a practice is a corollary of a different moral theorem that is itself impartial in form: i.e. one that does not specify illness or any other characteristic (aside from humanity) as meriting favourable treatment.

We might, for example, be favouring the ill in order to maximise total utility (the ill may not be worst off, but they are most able to benefit from the available resources), in which calculus we value the utility of each equally. In this case, fairness may be sacrificed, but it can at least be claimed that the governing moral theorem does not differentiate between persons. Such theorems and their corollaries can be held to be impartial: although fairness is compromised, it is compromised in the interest of an end that does not necessarily (though it does contingently) favour some humans over others; it does not **command** discrimination. (As we will see in Chapter Five, however, this distinction between direct and incidental discrimination is problematic.)

To see that even the composite notion of fairness ( $f^-$  &  $f^+$ ) is defeasible - consider the allocation of support to the blind.  $F^+$ , taken to its logical conclusion, would dictate that each should be granted sufficient resources to enable them to enjoy a life of equal satisfaction. This might conflict with  $f^-$  if the proposal is compulsory taxation. Suppose, however, the issue is how to allocate funds that have already been raised. The allocation to the blind according to  $f^+$  would be exceedingly high - it may be necessary to grant each blind person a full-time helper and substantial financial resources in order to make available to him anything like the opportunities available to the sighted. On the other hand, utilitarian considerations would perhaps lead us to spend relatively little on the

blind, because similar sums could lead to far greater welfare improvements for the rest of the community.

So we compromise, allocating more to the blind man than his utilitarian due, but less than would be necessary to grant him equal satisfaction with the rest of the community. If the blind man says to us, why are you giving me less than the others (in terms of welfare), we can only respond that it would be inefficient to do otherwise. We may be tempted to say that it would not be fair on the rest of the community, implying that fairness had not been compromised, but he could say, in what way would it be unfair to them: even the most extravagant allocation to me would leave me worse off than they -- who would swap places with me? We might reply, in exasperation, "well, life's unfair", thus shifting responsibility for his relative lack of welfare from us to the Cause of his disability -- but this response confirms that we do recognise the outcome as unfair, just that we rate the unfairness too expensive to offset<sup>10</sup>.

---

10

It may be thought that we could claim not to be being unfair on the basis of Dworkin's insurance hypothesis (Dworkin, 1981). His hypothesis is that our welfare system is designed like an insurance system, with taxation serving as premia: we should be happy to pay the premia, even those heavy enough to provide adequate support for the blind-from-birth, as we would have been happy to pay such premia if we did not know whether we would be born blind. But the critical question is what level of pay-out results. Even from behind a veil of ignorance, it is unlikely that sufficient pay-out would be agreed so as to equalise the opportunities of the blind and the sighted, because such pay-outs would be too expensive, and involve too great a sacrifice of the community's average level of welfare. So, the blind recipient of the insurance pay-out may still claim to have been unfairly treated, as, through no fault of his own, he has ended up with a lower level of welfare than anyone else. We could argue that it is not unfair because he agreed, with everyone else behind the veil of ignorance, what the insurance premia and pay-outs would be. But, of course, he can in justice retort that the veil is fictitious and thus what he agreed there carries no weight. This line of reasoning thus still leaves us with a compromise between fairness and utility maximisation (although it may help us in the difficult business of determining how precisely to strike that compromise).

Now consider the very rich blind man who merely requests exemption from tax. In this case both  $f^-$  and  $f^{+11}$  would indicate that we should refrain from taxation. But in fact we do tax the rich blind - presumably from considerations of welfare maximisation. So it is clear that fairness is being compromised in the interest of utility.

Nevertheless the principles of fairness might be taken to imply:

- that compromises of  $f^+$  and  $f^-$  need justification
- that that justification should be impartial in form.

### **1E      The Doctrine of Impartiality**

Accepting that fairness can be compromised for other ends, and even that our valuation of humans is sometimes inegalitarian, the notion of the equality of moral worth may be embodied into a doctrine of impartiality that circumscribes the ends to which fairness can be compromised:

- (I')            Humanity is the only morally relevant property of human beings (aside from obligations of restitution). No moral law will command differential treatment of persons **according to their significant characteristics, unless it is a corollary of a moral law that does not itself command differential treatment.**

This rule represents a second-line attempt to embody in our practices our conviction of the equality of human worth. Maximising fairness (in its two forms) is one way of embodying this conviction in social practice.

---

11

Under both Nozickean and Rawlsian notions of fairness.



Another, is, even when compromising fairness, to avoid practices that undermine persons' sense of self worth. (Here I am echoing Rawls' insistence upon the importance of avoiding "at almost any cost the social conditions that undermine self-respect" (1972, section 67).)

Discrimination is objectionable when it appears to demonstrate and validate the low esteem in which the victims are held by the majority.

Wherever fairness is compromised, the favoured group will always be picked out under some characteristic (e.g. "those whose welfare would be most increased by receipt of support"). The suggestion here is that this is only objectionable if the discriminating characteristic is one that is significant to the people concerned (perhaps because it is part of their self definition). This clause is shown to be problematic in Chapter Five, but the intention is to outlaw compromises with fairness that are so egregious in their lack of concern for particular individuals as predictably to undermine their self esteem.

In contradistinction (P) becomes:

(P') People are often right or even obliged to be loyal to or to favour other members of their group, whether family, community or state, **even when such discrimination neither serves an end nor fulfils an obligation that can be stated without reference to the group members' identity, or to their particular significant characteristics.**

The impartiality principle, the view that all persons are of equal worth, is foundational of many different political philosophies:

"We may say that individuals have a right to equal concern and respect in the design and administration of the political institutions that govern

them. This is a highly abstract right. Someone might argue, for example, that it is satisfied by political arrangements that provide equal opportunity for office and position on the basis of merit. Someone else might argue to the contrary that it is satisfied only by a system that guarantees absolute equality of income and status, without regard to merit. A third man might argue that equal concern and respect is provided by that system, whatever it is, that improves the average welfare of all citizens counting the welfare of each on the same scale. A fourth might argue, in the name of this fundamental equality, for the priority of liberty, and for the other apparent inequalities of Rawls's two principles." (Dworkin (1977), p.180)

Thus a range of political theories, including both an egalitarian concern with fair outcomes, and full blooded utilitarianism, can be seen as embodying the basic respect for human beings as equals that underpins impartiality. Will Kymlicka has expounded and substantiated this point, and has offered a plausible egalitarian interpretation of a range of modern political theories; he claims that an egalitarian interpretation is a prerequisite for respectability as a political theory:

"It is a matter of debate ... which specific kind of equality is required by the more abstract idea of treating people as equals. Not every political theory ever invented is egalitarian in this broad sense. But if a theory claimed that some people were not entitled to equal consideration from the government, if it claimed that certain kinds of people just do not matter as much as others, then most people in the modern world would reject that theory immediately. Dworkin's suggestion is that the idea that each person matters equally is at the heart of all plausible political theories." (Kymlicka (1990), pp4-5)

Impartiality, as embodied in (I'), is a meta-theorem based upon this apparently universally accepted moral principle that all are of equal worth;

conversely, partial actions, as in (P'), will not accord equal concern and respect to all.

Is it possible to reconcile inegalitarian partial practices (manifestations of loyalty and friendship) with the egalitarianism that appears to be the shibboleth of contemporary political philosophy?

It is true that egalitarianism is stipulated as a necessary attribute only of political theories, designed to provide principles of distributive justice for the state. The primary focus of most partialist practices, on the other hand, is with moral choice for the individual: should he favour his brother, should she pay taxes to her state? So perhaps the conflict can be resolved by distinguishing impartiality as a principle of political philosophy from legitimate partiality at the micro level.

However, the principle of equality of human worth (the equality of "human beings as moral persons" in Rawls' locution (1972, Chapter 77)) that provides the foundation for impartiality in guiding the action of states appears to be equally applicable to the individual as agent; we need argument to prevent that application.

In sum, the problem with partiality is that it fails to accord to one's fellows the equal concern and respect which is their due in view of our common humanity.

Can it nevertheless be justified?

In Chapters Two and Three we explore attempts, respectively, to carve out non-impartiality zones within which individuals' projects may be unfettered, and to show that partial practices can be justified instrumentally.

These attempts fail to do justice to our partialist intuitions, and we are therefore forced to discard the hope of justifying partial practices using impartial principles. Nevertheless, we continue to maintain that they need justification, as the maxims of fairness still bind. In Chapter Four we explore what notion of the good does lie behind partiality. And in Chapter Five we review what this must mean for our commitment to the principle of impartiality and our understanding of fairness.

## **2 FAIRNESS AND IMPARTIALITY: THEIR SCOPE**

This chapter considers a range of strategies for insulating partial practices from any impartialist critique. Some aim to show that there are limits beyond which the moral writ does not run, where partiality is at individuals' discretion (section 2A). Others attempt to found obligations to be partial upon the putative conceptual priority attaching either to individuals' altruistic emotions or to their conceptions of the good life, both of which are likely to give great weight to the welfare of the agents' own families or communities (section 2B).

I contend, however, that most of the arguments brought in support of either permission or obligations to be partial merely support indulgence of partiality. Only recourse to moral subjectivism secures permission to be partial: but we are unhappy with such latitudinarianism, which does little justice to our moral intuitions; rather we are driven to seek objective endorsement of our altruistic emotions and of our personal conceptions of the good. Yet once we seek objective endorsement, we lay ourselves open to impartialist assessment.

We will see that some value attaches to the satisfaction of our partialist inclinations to the extent that such satisfaction brings pleasure, and to the extent that pleasure is of value. However, to pursue such pleasures is an indulgence: once we take the moral step of evaluating our inclinations, we seem to be obliged (in the absence of other moral reasons for being partial), to train ourselves to take pleasure in being impartial.

### **2A Indulging Partiality**

The arguments discussed in this section allege that applying impartial standards at the personal level is illegitimate, either because the resulting morality would be too demanding or because morality is taken to be indifferent to what one does in one's own time, or with one's own property.

Such arguments, if valid, would support a **personal licence to be partial**, though they could not establish obligations to be partial. This is the attitude of neutralist liberals like Rawls and Barry amongst political philosophers, who are concerned that the morality that underpins the institutions of the state should be impartial between competing conceptions of the good. It is likewise the perspective of philosophers like Cottingham whose domain is personal morality and who are anxious to leave individuals discretion to establish their own priorities.

I will contend that their arguments show only that we should be **indulgent** of personal partiality; we recognise that morality is demanding, and do not judge harshly those who fail to live consistently by its teachings, but nonetheless judge that it would be better if they could. It will turn out that a genuine **licence** to be partial (in the sense that a disinterested action would be no more praiseworthy than one favouring a friend or relative) only exists in the degenerate (though not uncommon) case where there is uncertainty as to which is the optimal course from an impartial perspective.

Nevertheless, a stricter demand for impartiality will be shown to apply at the political level in the sense that we can be more demanding (less indulgent) of those involved in a political process than of individuals acting in their own time.

## **i            The demandingness of impartialist morality**

One of the complaints commonly levelled against utilitarianism is that it is too demanding a code. It leaves no time for individuals to pursue their own projects; they must always check first whether there might be something they could do which would add more to the sum of human happiness. One result is that any inclination to do good to kith or kin must

be set aside in favour of an alternative action that is happiness maximising.

The problem is not limited to utilitarianism: any consequentialist doctrine will involve a duty to maximise the good; Kantians and other non-consequentialists also generally accept a virtue or a duty of beneficence which will make demands when other duties are not relevant, and, in the absence of argument to the contrary, this duty of beneficence will be impartial in form<sup>12</sup>. As a result of discomfort with this intrusiveness of morality, a number of philosophers have sought to show that such impartial principles do not apply to all decisions in all contexts. Were such arguments successful, they would also help to exempt individuals from applying the criteria of fairness (as defined in Chapter One) in their practical reasoning in the personal domain.

Cottingham (1983) argues that acceptance of impartiality can be very demanding, forcing one "to give **the same weight** to the interests of large numbers of starving Asians as I give to my own interests." To do so "would be to cut my living standard to the very bone, to reduce my life to one of little more than bare subsistence" (p.91). Cottingham addresses his argument explicitly *ad hominem* at the impartialist: how can you drink beer and give your child a toy when you know that there are people starving in Asia? Now such examples undercut the force of Cottingham's argument, for one is strongly minded to say that the impartialist is right,

---

12

There are important problems associated with a consequentialist general duty of beneficence. These relate to problems of inter-personal comparisons of utility, and the difficulty in excluding unpleasant desires or preferences from the maximisation calculations. However, these are essentially problems of hard cases: in many cases we are able to tell that the course we are inclined to take will produce less welfare than some other course, and such calculations will often tell against the impartialist. Thus, these problems cannot be used as a general escape from the duty of beneficence or the demands of impartialist morality (although, as suggested below, p.59 et seq, the difficulty of the calculations does expand discretion in certain circumstances).

that one ought to give up the beer if one was at all confident that doing so would save a life (even in Asia!<sup>13</sup>). This is an ideal to be aspired to, even if it is difficult to attain.

The demandingness of morality is less of a problem for deontologists than it is for consequentialists, for although most deontologists will recognise a duty of beneficence which will apply all the time, it is open to them to distinguish between the exigency of different duties. Thus Kant himself distinguishes between perfect and imperfect duties. The latter include the duty of beneficence, and allow the agent discretion in the timing and the exact means by which the duties are fulfilled.

Such a distinction between duties allows a break point in our evaluation of our fellows between those who always fulfil their primary duties and occasionally fulfil their imperfect duties, and the heroes and saints who bend their every action to the dictates of duty. Scope is created for the individual to perform morally indifferent actions without attracting censure, so long as primary duties are performed always, and imperfect duties regularly (cf. Baron (1987)). On this model the evaluation of both sinners and saints will nonetheless be with respect to their fulfilment of imperfect duties which are impartial in form. So the partialist case is not advanced.

One might argue that if the individual has discretion over the fulfilment of the imperfect duties, she must be able to bias that fulfilment towards kith and kin. If so, she will be fulfilling his duty of beneficence by favouring her sister over the stranger.

This would be an example of legitimate partiality, but the partiality would be no part of what made the beneficent action good. To the contrary, if she was in a position to help the stranger more at no extra cost, then she

---

<sup>13</sup> Peter Unger (1996), in his essay *Living High and Letting Die*, makes a powerful case for giving equal consideration to the claims of the desperate irrespective of distance.



will be judged less beneficent than she might have been. She may not be criticised for this, but the measure of praise meted out will be less than it would have been had she acted impartially. Impartiality would remain the ideal.

For the consequentialist, it is less easy to find a break-point between the renegade and the saint. If every action is to be based on the maximisation of the good, how is one to know when one has done enough, when one is at liberty to devote time to oneself without censure? But the difficulty should not be overstated. Although it is difficult to form rules to guide the behaviour of a consequentialist who wishes to be rated an ordinary decent chap (doing his duty and no more), it is not in practice very difficult to distinguish such a person (who will maximise occasionally and/or when the demands of the calculus are particularly powerful, blatant or emotionally sympathetic), both from the saint (who will take pains to maximise the good almost all the time), and from the rascal who disregards the interests of the general good invariably<sup>14</sup>.

There is thus no reason to believe that even the most demanding of moralities condemns us all as sinners for our partiality -- the question is only whether morality always dictates an ideal impartial course of action or not. And for both the consequentialist and the deontologist there is in principle always a best thing to do, and one that apparently makes no allowance for partial preferences; there is no theoretical limit to the intrusiveness of morality.

---

<sup>14</sup> Thomson (1997) adopts the simple expedient of defining moral requirements in terms of "the vices": we are only required to avoid being unjust, cruel, or mean, even where it would be better to be just, kind or generous. Given the costs involved, the latter behaviour would be supererogatory.

Nevertheless, considerations of the demandingness of morality do suggest tolerance of a degree of partiality at the personal level. This would support an emasculated version of (P'), but certainly not (P') itself:

(Pa) People are **entitled without censure** to be loyal to or to discriminate in favour of members of a group, whether family, community or state, even when such discrimination neither serves an end nor fulfils an obligation that can be stated without reference to the identity of group members, or their particular characteristics; **but this entitlement applies only in circumstances where the alternative impartial course would be little better.**

Partial action (in the appropriate circumstances) will not justify censure, but impartial action, as supererogatory, **would** always merit praise.

This principle of indulgence will accord much less latitude to the state and its officers than it does to individuals acting on their own account.

Our tolerance of states' common discrimination in favour of their own citizens may be by way of an indulgence of the prejudices of the populace. One can envisage such prejudices being weakened to the point that we would no longer tolerate partiality explicitly embodied in states' laws, though we would still be indulgent of loyalty to kith and kin on the part of individuals.

More straightforwardly, we may well demand higher standards of officers of the state acting in their official capacity than we do of other individuals (though the severity of such standards, whether they tolerate police officers overlooking the crimes of their nephews or merely civil servants securing job interviews for their acquaintances, seems to vary with time and place.)

**ii            Partiality may Rightly be Lawful without being Moral**

That there are partial practices that we treat indulgently does not establish a theoretical limit to the intrusiveness of morality or of the principle of impartiality. A fortiori the existence of partial practices which we do not criminalise does not prove that those practices are morally optimal. Yet, conflation of these attitudes (approbation, indulgence, legal tolerance) has led some to conclude from our unwillingness to outlaw or to condemn classes of partial actions, that they are beyond the reach of morality.

Barry (1995) emphasises repeatedly that the impartiality that is embodied in his theory of "justice as impartiality" is a second order impartiality, by which he means that it is one that operates only at the level at which a society agrees the institutional and legal framework. Unlike first order impartiality, second order impartiality does not require individuals to operate impartially in their day to day lives.

Barry argues that impartial political philosophies, utilitarian and Kantian ethics all respect impartiality at the political level but that all can leave discretion for the individual - to stop short of first order impartiality. Some of the arguments he advances adduce other ethical considerations for allowing individuals to make choices that do not grant equal consideration to each competing interest (e.g. that there are good reasons for allowing or encouraging someone to favour his wife when rescuing people from a fire, either because encouraging such a sentiment will be utility maximising or failure to do so would be inconsistent with the nature of his relationship with his wife). These arguments will be considered below (sections 2B and Chapter Three). They are not really arguments for personal discretion; rather they suggest that there is some other duty or consideration that in the circumstances overrides the obligation of impartiality. (This is a distinction that Barry passes over.)

However, Barry also makes the more general claim that a space should be left for personal discretion:

"...once someone has been assigned something as his legitimate property, he is not normally to be held accountable under the rules of justice (whether legal or moral in form) for the use that he makes of it. It can be consumed by him, given to somebody else, exchanged for something else of value ... or destroyed.

"What William Godwin called - pejoratively - `the magic in the pronoun "my"' can be explained ... by saying that life goes better if what I do with my toothbrush is my business... I should not be under any obligation to ask myself whether somebody else could make out a better case for using my toothbrush than I can, on the basis of impersonally applicable criteria... The point of a system of private property in personal possession is precisely to exclude that.

"... All of us have only a finite amount of time, attention, care and affection to devote to other people... and life would scarcely be worth living if we could not decide for ourselves - once we had met our social obligations - on whom these should be bestowed." (Barry 1995, section 33)

Barry does not wish to argue that there are never circumstances in which one should not make one's property available to others, nor even that choices of personal friends might not sometimes lead to bad consequences. ("This is not to say that an occasion might not arise where the right thing to do was to lend someone my toothbrush."). Yet this is surely an admission that morality does hold sway at the personal level. Once that is accepted, the extent of need in the world leads one to the conclusion that every private use of time in principle needs justification against the alternative uses that one could make of time and resources. It may be morally indifferent which of two television programmes I watch;

but is it morally indifferent if I spend the time watching television or working to alleviate poverty?

Of course, as I emphasised in the previous section, we might not wish to condemn someone for spending more time than strictly necessary in relaxation. The force of Barry's argument trades on the implicit construction of a straw opponent who would condemn or even outlaw such an activity. The references to it being "my business" what I do with my property, and to it being "for us to decide" where we bestow our affections, imply that Barry's opponent would have someone punished for failing to be maximally sensitive to competing demands upon her time or property. Yet no first order impartialist is suggesting that a court should intervene in such cases, or even that failure to share should generally be the subject of social sanction. What they might powerfully argue however is that *ceteris paribus* it would be better to be sensitive to others' claims on one's time and resources.

Barry's argument that there should be personal discretion "once social obligations had been met" is based upon a Scanlonian assessment of the principle. Scanlon argues that

"An act is wrong if its performance under the circumstances would be disallowed by any system of rules for the general regulation of behaviour which no one could reasonably reject as a basis for informed unforced general agreement." (T.M.Scanlon, "Levels of Moral Thinking" in Senor and Fotion Eds, Hare and Critics, 1988, quoted in Barry (1995), section 11)

It is true that Barry wishes to apply the Scanlonian approach only to rules of justice rather than to wider morality. Hence he might claim that he would accept that morality might have something to say about what one does within one's discretionary space. However, this distinction between the assessment of acts under the justice heading and under the "wider

morality" banner begs the question with which I am concerned, viz whether there is a separate realm of moral debate and separate standards applicable to the issues of political philosophy from those which hold sway over individuals in their private capacity. As we shall discuss in Chapters Five and Six, neutral impartialists (including Barry) make this distinction plausible only by suggesting that wider morality is less well attested and hence a matter of individual conscience. He quotes approvingly Mill's definition of justice (in chapter 5 of "Utilitarianism") as "the name for certain classes of moral rules which concern the essentials of human well-being more nearly, and are therefore of more absolute obligation than any other rules for the guidance of life." (quoted in Barry (95), section 12). Hence it is clear that Barry considers the constraints of justice more important than those imposed by "wider morality".

It is also clear from Barry's choice of adversaries that he does not wish to leave first order impartialists with a cogent case in "wider morality". Barry has most respect for Godwin as an impartialist, but his treatment of Godwin nowhere suggests that the rule of impartiality that he advocates might have application outside the domain of justice.

The Scanlonian argument for personal discretion (in the domain of justice) has two stages according to Barry: a priori and empirical. The a priori argument is that...

"In an attempt to secure strict impartiality in all areas of life a huge number of decisions that are now left to private judgement would have to be turned over to public officials; and all decisions left in private hands would be open to scrutiny and censure..." (Barry (1995), section 33)

The conflation of the issue of enforcement with that of the standards of morality is here very clear. All might agree that there should be no policing of private actions of certain kinds, and further that (for a more limited class) acting impartially was supererogatory, whilst nevertheless

maintaining, pace Barry, that universal impartiality was the appropriate norm.

The empirical argument is derived merely from the claim that "commonsense morality" allows for a realm of personal discretion, and that commonsense morality is in this case reliable (there is no reason to think its assessment biased or unreasonable in Scanlon's sense). Both arguments can be challenged:

- Barry notes (op.cit, section 2) that common sense morality respects individuals' right to choose their own friends, yet (notwithstanding the principle of freedom of association) disapproves of golf clubs that bar members of a certain race. Yet one might claim that common sense morality would consider it wrong of somebody to refuse friendship to someone on the grounds of race, whilst nonetheless maintaining that it would be wrong to penalise such behaviour. If this is right, there is no realm of personal discretion within common sense morality
  
- Barry concedes that "up to a point, increasing the stringency of first order partiality works to the advantage of the least advantaged groups in a society" (op cit section 34), yet he refuses to acknowledge that that very fact might have biased society against first order impartiality, and in favour of a rule that protects the more advantaged majority from feeling too bad about the resources and time that they devote to their own projects. Barry's reasoning is merely a repeat of the a priori argument that "the ability, without incurring legal and moral sanctions, to do with or for some people what you would not wish to do with or for other people" is of value to everybody. Yet, might not the homeless under our bridges think this reasoning, justifying

their exclusion from benefit from middle class hospitality, somewhat self-serving? Hence in this case common sense morality appears an unreliable guide to the right.<sup>15</sup>

Cottingham's justification of his claim for personal insulation from the demandingness of morality is no more effective. He uses property rights to amplify his claim that one may neglect impartiality in favour of what he calls (following Broad and Mackie) self-referential altruism:

"Suppose I am not a public official, but an ordinary citizen with some food at my disposal; can it seriously be maintained that it is an improper or unethical piece of favouritism for me to bestow it on my own child?"(1983, p.96)

A little later, Cottingham seeks to forestall an objection:

"To avoid the (tedious) objection that talk of 'his own food' presupposes an uncritical acceptance of existing distributions of property, we may if necessary stipulate that the parent in question gathered the seeds from the wild, sowed them on common land and raised and harvested the food unaided, by the sweat of his brow."(fn.22)

I would like to pursue this tedious objection, for the assumption that the food belongs to me is crucial to the contention that I have every right to give it to my child. The mention of some sort of Lockean acquisition merely serves to highlight the shaky moral foundations of property ownership. Even were it accepted that a parent who gathered seeds from

---

<sup>15</sup> Mendus (1996) notes that Barry equivocates over whether the strictures of justice ought to extend into relations within the home, in particular between man and wife. Again, one might wonder whether the consensus (if such there be) that this is a private matter is one that has arisen without bias.



the wild had the right to give the food to his own child, such an exception to the rule of impartiality would give little comfort to the vast majority of parents, whose ownership claims are more tenuous.

Such a critique of parental partiality is not dependent upon a rejection of property ownership as a commendable political system; only upon a recognition that property ownership may be merely a political system deemed to be better than alternatives. Property ownership confers only legal rights; arguments that it confers moral rights are very hard to sustain<sup>16</sup>. Now if in Cottingham's example we set aside, for moral purposes, conventional property rights, then giving food to my own child in preference to someone more needy will seem questionable.

Cottingham's negative phraseology tends (as with Barry) to make one conflate the legal with the moral. To claim that giving food to one's own child is "improper or unethical" makes it sound as if it should be outlawed (or at least heavily censured). The question is only whether it would not be better to give the food to a hungrier child<sup>17</sup>.

In a later article, Cottingham improves the argument:

---

<sup>16</sup> Barry recognises that property and other rights "are a conclusion, not a premise". But he then leaps to the claim that "they operate so as to establish an area within which people can act as they choose". Even the law restricts what people can do with their property; Barry's claim that there are no moral constraints on property use (for instance a duty of charity) is without support even in commonsense morality.

<sup>17</sup> Cottingham actually relies more heavily upon an alternative line of attack against impartiality, which is not dependent upon either property rights or other claims for personal moral discretion. He suggests that duties to one's kith and kin are universal obligations, justified on the basis that without such obligations "the very object of ethics, human fulfilment, would be defeated". Note the fragility of the empirical base upon which a defence of a marginal partial act would rest ("Best favour your own child lest human fulfilment be impaired!"). This sort of argument is discussed further in Chapters Three and Four.

"There is, in our everyday lives, a subtle but crucial weighting we accord to our own interests. A resource which I legitimately control - my leisure time, for example - may legitimately be devoted to my own recreation rather than that of my neighbour; and to justify this I need point no further than to the fact that my leisure time is my own."

(Cottingham (1991), p804)

One could still argue that ownership of leisure time is not really inalienable -- after all in some countries there is a legal obligation to save life, and in many countries the state demands the time and labour of its citizens through conscription.

Even granting that ownership of leisure time is more firmly based than ownership of property, Cottingham still conflates the legitimacy of an action with its merit. If we are trying to answer the question, "What is the best action open for me to do?", then Cottingham has not shown that helping neighbours, who will benefit more from one's leisure time than oneself, is not commendable.

Of course, if he is interested in the question "What should I be legally obliged to do?", then he is correct. Even the question "What ought I to be morally obliged to do?" (such that omission would justify contumely), should be answered with some respect for the notion that one should be able to spend one's leisure time on oneself without guilt. But even in this case, absence of culpability will depend upon the gap between the benefit to oneself and the potential good to others being relatively small. And this approach could never support a contention that it would not be better to be impartial, i.e. it will not move us from (Pa) to (P').

Confusion between legality and moral desirability recurs in later Cottingham:

"There is nothing wrong as such with preference for a nephew, as may be seen from the fact that (except for the most hard-nosed totalitarian) no blame need automatically attach to the entrepreneur who makes his nephew a director of the family firm that he has worked hard to build up. To see the position of any nephew so preferred as *eo ipso* "tainted" is to take a significant step towards arguing that all family firms should be outlawed, or forcibly prevented from surviving the first generation of those who founded them - a conclusion which on reflection is incompatible with even a minimal conception of property rights let alone fundamental principles of efficiency and incentive."  
(Cottingham (1998))

Oddly, reasons of efficiency might be the basis for a complaint from the other stakeholders in the family firm if the nepotistic owner-manager had passed over a better manager in favour of his nephew. And such a complaint would have a base embedded in the strain of impartiality in first order morality. Yet even such a shareholder may stop short (again for reasons of efficiency and incentive) from wishing to outlaw the practice: what is the best action is one question, what should be legally or socially sanctioned is another.

### **iii**                                          Rights and Duties

Others attempt to establish a sphere of personal moral discretion through the construction of a rights based ethical theory, whereby individuals are constrained not to trespass on others' rights (these being construed negatively, e.g. the right not to be killed rather than the right to life), but are free to do as they wish otherwise. In Kekes' (1981) terminology, the rights-theory would represent the social morality, leaving individuals to determine their personal morality independently, so long as the edicts of social morality are not transgressed.

Alan Gewirth's theory is an example of the difficulties which such an approach encounters (Gewirth (1988)). Because of the sparse morality implied by limiting one's ethics to consideration of the more plausible rights, Gewirth is driven to positing positive rights to life and liberty, which in turn generate obligations on everyone else to try to secure those goods for everyone; thus the moral landscape is enriched. Such enrichment is clearly necessary if the resulting ethical scheme is to bear any resemblance to the thorough critique of human action that is our morality.

But in order to re-establish personal space Gewirth is then forced to introduce secondary rights, and amongst these are included the rights to form various relationships, communities etc., and rights to take upon oneself various consequent obligations. Fulfilling one's obligations to the groups to which one has committed oneself could well involve one in partial actions. By undertaking commitments, one has effectively sealed off areas of decision-making from the claims of people outside the group.

Such rights to undertake commitments would do less to establish a licence to be partial than Gewirth seems to think, for many of the partialist obligations generally thought to hold (particularly from children to parents and from citizens to states) do not stem in any literal sense from voluntary undertakings by those thought to be morally bound. More fundamentally, however, Gewirth has not shown how one is to justify entering into commitments which compromise one's ability to fulfil one's general duty to secure everybody's rights to life, liberty etc.

#### **iv Epistemic limits to the demands of impartiality**

Less grudging license to be partial arises from consideration of epistemic (i.e. practical rather than theoretical) limits to the demandingness of morality: many choices will appear indifferent from the moral point of view, simply for lack of adequate information. The lack of

information may arise either because it is in principle impossible to get the information, or because it would be too expensive (in time or resources) to obtain it -- the cost would outweigh any likely benefit. In such cases, an arbitrary choice must be made, and such an arbitrary choice to favour kith and kin is no less praiseworthy than the contrary course.

For example, suppose someone is trying to decide whether to visit a friend or to spend the day collecting funds for charity. The following considerations render the judgment very uncertain.

- o She may not know which is the better thing to do, because she does not know whether her friend is busy or would welcome a visit (telephoning would not help, for she knows that if she suggests the visit the friend will feign enthusiasm even if he would prefer to be alone), and because she is unsure how much of the money that people put in her charity box would otherwise have been contributed to another collector's box. Both cases of ignorance could be rectified (by discretely consulting her friend's other friends, or by consulting with charity experts, respectively), but the agent is convinced that the cost of knowledge would exceed the

likely benefit<sup>18</sup>, she therefore plumps for one course or the other arbitrarily.

- o She may not know for a more profound reason: that she does not think that there is in principle any way of discovering whether the increase to welfare that would result from her fund-raising would exceed the pleasure they will both gain from her visit to her friend. This may not be because she believes that inter-personal comparisons of pleasure are impossible, but merely because it is not possible to ascertain the benefits of a course if it is not actually pursued.

No moral condemnation or commendation accrues to the practice of partiality in such cases; such uncertainty greatly mitigates morality's demandingness.

In view of these considerations, (Pa) becomes:

- (Pb) People are entitled without censure to be loyal to or to discriminate in favour of members of a group, whether family, community or state, even when such discrimination

---

18

It is a nice question quite how this account is to be drawn up. The agent is ignorant of almost all the relevant pieces of information: she does not know how much it would cost to find out the marginal benefit to friend and to charity, and without doing the research she does not know what is the cost of making a mistake. All must be judged with subjective probability: for instance, she must implicitly judge, if she opts to collect for charity, that the chances of her friend being suicidal are so small as not to warrant even the likely costs of attempting to confirm that he is not. Is it a problem that morality seems to require such calculations? No, because the next stage in the working tells us that the cost of entering into such calculations consciously is too great, and would, in time, prove counterproductive. We have only to be alert (probably without the benefit of rules) to cases where calculations are justified.

neither serves an end nor fulfils an obligation that can be stated without reference to the identity of the group members or to their particular characteristics; but this entitlement applies only in circumstances where the alternative impartial course would be little better or where it is impossible efficiently to discover the best of a range of available actions.

How important is this concession?

Uncertainty is pervasive; yet the form of the concession is such that it does not exempt us from evaluating the different options first from an impartial perspective.

The position would be very different were the difficulty in evaluating the alternatives a matter of principle, as for instance if one claimed that it is impossible in principle to compare the pleasure of one individual with that of another. Such a view would yield the conclusion that whenever there was a choice between favouring a relative or a stranger the ineradicable incommensurability of the outcomes would leave the choice arbitrary, and morality with no purchase.

It is hard to credit a claim that a minor pleasure for one person cannot be compared with a major pleasure for another. But a slightly less dramatic claim of radical incommensurability is put forcefully by Raz (1986, Chapter 13) in discussing a large class of decisions in which the goods or evils are generated in different dimensions. One important example is choice of career: how is one to compare a choice of career as a clarinettist or a lawyer? The issue is not just that so many of the consequences of the choice are unforecastable, but that the sorts of good that would be yielded by success are so different in kind. If Raz is right, all such choices would be rendered immune from considerations of morality in general and impartiality in particular.

Raz argues first that uncertainty as to what to do in such cases can be distinguished from equality of value by a test:

*"The test of incommensurability is failure of transitivity. Two valuable options are incommensurable if (1) neither is better than the other, and (2) there is (or could be) another option which is better than one but is not better than the other."* (p.325)

The obvious response is that there might not be precise equality, but there may nonetheless be rough equality. If there is rough equality, but one cannot judge precisely which is better than the other or by how much, there is no inconsistency in claiming that there would be rough equality between one of the options and a third option, even though that was slightly better than the second. If this is right, there is only a minor victory for discretion: the chooser must evaluate her options and assess their relative value as best she can. Only if she can make a plausible case for rough equality is she justified in determining her choice with non-moral factors. (This argument would not apply to cases in which failure of transitivity involved a third case that was substantially better than one of the options, but still roughly equal to the other. But I doubt whether such an example would be plausible. Suppose the third option is that the student dedicates herself with extra industry to her legal career, to establish an international reputation. This option is substantially better than the option of a mediocre legal career; but transitivity is unlikely to fail in such a case - a first class legal career is also likely to be judged better than a mediocre clarinet career.)

Raz considers and rejects the solution that apparent lack of transitivity merely reflects rough equality. First he suggests that:

*"Two options are roughly equal if and only if it does not matter which one is chosen, if it is right to be indifferent between them...."*



But in the case of the student faced with a choice of careers, to both of which she is equally suited:

"It seems to me that this is the sort of decision that anyone facing it quite rightly cares a lot about. It is a choice that one ought not to be indifferent to, or unconcerned about. To be indifferent to this kind of choice is not to have a proper respect for oneself." (pp 331-2)

Hence, the options are not indifferent but incommensurable. The consequences of this claim from Raz are dramatic, for it leaves large and important decisions marked by "the inability of reason to guide our action". Raz does not claim that one is therefore justified in any decision or any path, for there must be good reason (objectively) to support any option. However, perhaps to a greater degree than he acknowledges, he has radically curbed possible challenges to a chosen path, and thereby built for individuals much greater scope for choosing their projects without regard to the impartialist critique.

However, the claim of radical incommensurability remains implausible. For one thing, changes in one option that leave it within that dimension, can resolve the doubt, and restore comparability. For example, were one of the career options to be sufficiently improved, e.g. by the winning of a major clarinet competition, the student's career uncertainty might well be resolved. But it would be paradoxical if some options in different dimensions were comparable but others, apparently closer in appeal, were incommensurable.

Further, in arguing that two options cannot be roughly equal because we care about the choice, Raz trades illicitly upon an ambiguity in the concept of indifference. Just as one can be disinterested in an outcome whilst being interested in it, so one can be indifferent to a choice whilst caring greatly about it. What one cares about is that one has to make a choice. You cannot decide whether to choose law or the clarinet because

you are greatly attracted to both, and you feel that giving up either is a tragic lost opportunity. Nevertheless, they are of roughly equal attraction; indeed, you might make this momentous choice by tossing a coin without denying the importance of the decision or showing inadequate self-respect. Such an approach would honestly recognise that there is no way of rationally choosing between them: for (and perhaps this follows tautologously) they are of equal merit.

A more powerful example is used by Winch (1965) to reach a similar conclusion to that of Raz - also incorrectly. He considers the case of Billy Budd (in Melville's rendering of the story), an angelic foretopman on trial for striking a satanic officer during a period where the risk of mutiny is reckoned significant. The captain of the ship is faced by a moral dilemma: the arguments for mercy and for following the legal code are both compelling. Winch concludes that in such cases it is the Captain's deliberating and deciding that determines what course of action is right for him in those circumstances, and it is therefore inappropriate to generalise his decision to others facing the same dilemma, or to use other similar decisions to judge him.

As in the Raz case, the competing options are equally balanced, and here too one is uncomfortable concluding that it is a matter of indifference which option to choose - after all a man's life depends upon it. Winch implies that as there is no right answer to be determined by weighing the arguments, it is for the agent to determine what is right. This provides a measure of moral self-determination and discretion.

However, it is more plausible to conclude not that the arguments are incommensurable, but that they are of roughly equal force. Hence, Winch is right to hold back from criticising the captain's decision to execute Budd, not because the decision is "right for him" in some peculiar sense, but simply because both actions are equally justified. Winch's feeling that

he would have acquitted Budd, and would have been right to do so, is then not in contradiction with his justification of the captain. Tertium datur.

This interpretation of such hard cases explains why they are hard. Winch, on the other hand, generalises from this case to impugn universalisability as a principle. But, even were one sympathetic to Winch's characterisation of this case, it would have little relevance to the obligation to universalise one's judgements - or to take account of the dictates of impartial morality - in less vexed cases.

In hard cases, there is no liberation from morality; on the contrary, the tragedy of the choice between evils (or between goods), is exacerbated by its arbitrariness. (See section 6D for further discussion of the implications of hard cases.)

Any licence to be partial based upon epistemic limits will, like the indulgence of partiality, be less accommodating for the state and its officers. Either explicitly or implicitly, law often gives discretion to officials in cases of uncertainty, but we may determine that such discretion should not be exercised to further the partialist ends either of the state or of its officers, for fear of prejudicing the reputation for impartiality that may be critical to the success of the state's institutions. (Had the captain been related to Budd, we would have expected him to hand the judgement to another).

## **2B Partialist Obligations?**

The arguments discussed above establish areas of discretion or licence to be partial, and further areas of indulgence of partiality, such that (as in Theorem (Pb)) there are circumstances in which people are entitled to perform partialist actions without censure. The common consensus however, is that partial practices are not only to be indulged or licensed

but also in many circumstances to be encouraged. Successful positive arguments for partiality must establish that people are often under an obligation to be partial, such that the partial action is the right action -- (as in (P')).

Some of the positive arguments that I will consider attempt to found obligations to be partial upon an alleged conceptual priority attaching either to individuals' altruistic emotions or to their conceptions of the good. For virtue theorists and communitarians, these emotions and conceptions form the foundation of morality. For impartialists, on the other hand, the deliberative mechanisms that determine the structure of rights will have overall priority, yet nonetheless in the domestic sphere (or thereabouts) there is said to be value in a form of **moral self-determination**, whereby individuals establish and commit themselves to their own conception of the good. (Moral self-determination is sometimes misleadingly labelled "autonomy"; but this term is better reserved for freedom to choose what to do; "moral self-determination" is the notion that one can choose what course is right.) I argue that attempts by either camp to establish the value of all goals that individuals set for themselves are confused.

Subsection i discusses the view that spontaneously heeding the **benevolent emotions** forces one to ignore impartiality. I argue that this view fails to recognise that we can and do subject even actions springing from emotion to rational critique. We expect moral agents to be driven spontaneously by benevolent emotions and yet to be sensitive to other calls upon their resources that have less emotional appeal, even where prior ratiocination is inappropriate.

Subsection ii deals with attempts to establish that individuals' conceptions of the good (**their projects**) are of value irrespective of any external evaluation, and hence irrespective of their partiality. Included are

theories that predicate such positive evaluation of projects on their being freely chosen.

I argue that one can recognise the importance of personal projects, and yet insist that these projects can and should be **evaluated** from an impartialist standpoint. This is not to deny that if we have partialist inclinations whose satisfaction yields pleasure, then there is *ipso facto* good in their satisfaction; but we are obliged, if we are able, to adapt our inclinations so that we take pleasure only in impartial projects. (Indeed this ability to reevaluate in the light of objective criteria is one aspect of autonomy properly understood.)

I go on to show that there is no need (*pace* Raz) to posit moral self determination to explain the value of autonomy nor (*pace* Scheffler and Williams) to account for our dissatisfaction with impartial consequentialism (deontological constraints, impartial in form, adequately explain our intuitions).

I conclude that partialist obligations cannot be justified against the impartial principles of justice simply because they form the core of individuals' autonomously chosen projects. To the contrary, those projects are subject to external critique against the standards of fairness and impartiality.

## **i Spontaneity and the Benevolent Emotions**

Blum (1980) argues that there are areas of action in which "impartiality is not applicable" (p64):

"It is evident that in most normal situations of acting from friendship the agent neither acts from the thought, 'My friend needs help and in situations such as this one it is morally permissible to offer that help';

nor does he normally test the proposed action against a principle which specifies when it is permissible to ignore, or to fail to give equal weight to, the interests of those in the range of his beneficence. (Nor, as I have been arguing, is it morally appropriate for him to do either of these things.) [p.62]

There is, Blum argues, a moral space for voluntary actions motivated by emotional urges, not rationally calculated and deliberated. Indeed, some of these actions gain their moral worth from their spontaneity.

Expressions of love and affection would perhaps seem less admirable if they followed consideration of the range of alternative actions.

What is praiseworthy in these cases would be vitiated by cold calculation of what course of action would lead to a fairer allocation, or would maximise welfare. The common criteria of praise or blame are here inconsistent with any rule, which, like impartiality, requires prior ratiocination.

On this basis Blum seeks to allow many actions of partiality. Indeed, he encourages them as he regards actions springing from the altruistic emotions like friendship and love as morally good in themselves.

However, Adrian Piper has questioned whether we do require unthinking spontaneity even in the exercise of the warm-hearted virtues:

"When a friend and a stranger suffer with equal intensity and one empathically imagines the inner states of both with equal vividness, a compassionate person will feel equal sympathy for both, and equally moved to ameliorate the suffering of both. Because the inner state of each bears the same relation to one's own,...compassion evinces a strictly impartial concern for the stranger's as well as the friend's condition. What finally determines one to render aid to one's friend instead of the stranger is not one's heightened compassion for the

friend. What moves one to help the friend are the bonds of mutual trust, loyalty, shared history, responsibility, and respect that uniquely define the relation of friendship." (Piper (1991), p.754)

Piper does not say anything more about the special bonds of friendship (except to indicate that they will only be critical in the event of a tie -- they will not overrule impartiality in her view), but she does convincingly show that we value refinement of the benevolent emotions to the point where their arousal and character is finely discriminated according to the nature of the situation.

Piper's point is corroborated by practice even in cases where lengthy prior deliberation seems inappropriate: we do not limit the requirements of rationality, and therefore do not limit the obligation to take due account of all relevant factors, to cases which allow extended forethought. We demand rationality even when we cannot demand conscious deliberation.

Consider how we would view somebody visiting his bereaved friend when he knew that he could otherwise save an innocent life. For instance, whilst driving he notices that a footbridge over a motorway is close to collapse; he determines to call the police when he gets home. But he finds there a note from his bereaved friend, and is overcome by a surge of Blumian compassion, which inspires him to rush off to see his friend without recollecting the footbridge. The latter subsequently, but avoidably, collapses, killing many. Would we say he could exculpate himself by claiming he was acting under a laudable but non-rational impulse?

Blum implies that this man should be held blameless because he did not consciously decide to ignore the footbridge, and because thought is inappropriate when reacting to a benevolent impulse.

Yet, just as conscious deliberation does not have to precede an action for that action to be ascribed a mental cause and for the agent to be given

responsibility for it, so also the fact that conscious deliberation is not possible does not excuse a crime of omission. Davidson (1963), arguing that reasons are causes, accepts that reasons explain actions by providing "an interpretation, a new description of what he did which fits it into a familiar picture". He posits some causal law connecting the background conditions -- of motives, knowledge, desires plus a trigger event -- with the ensuing action, but he does not require even that the "primary reason" for an action be consciously worked out. Introspection confirms that it would be absurd so to require.

The fact that the man who visits his friend rather than saving a stranger does not consciously think about saving the stranger does not excuse him from having failed to save her (assuming that he had latent knowledge of the stranger's predicament, as in the footbridge example) any more than the fact that he did not consciously evaluate his visit to his friend would remove the praise for that action. In both cases, we can ask him why?: why he visited his friend, and why he did not save the stranger. Intelligible (truthful) answers will provide the basis for assessing his actions, whether or not he was conscious of the reasons in advance of the action/omission.

With regard to the stranger, he may answer that it did not occur to him, it simply did not enter his head. Such a plea may reduce his responsibility, but it will not eliminate it: we can say to him that he should have thought of it. Recall the notions of moral luck and real desert from Chapter One: he may not really deserve to be blamed if he just did not think of it (especially if we judge that he could not have trained himself to remember such things); yet we will nevertheless judge him a less virtuous person for this lapse, notwithstanding that it is sheer bad luck that he suffers this character flaw. He cannot respond by saying that thinking about other calls on his time would have been incompatible with the free exercise of his feelings of friendship, for moral agents are not expected to indulge in



long calculations, but rather to be affected by relevant calls when appropriate.

This also provides at least a partial response to Susan Mendus's interpretation of Williams' famous remark that "the two level approach [that lets a man rescue his wife in preference to a stranger only on condition that circumstances allow discretion<sup>19</sup>] provides the agent with one thought too many: it might have been hoped by some (for instance by his wife) that his motivating thought fully spelt out, would be the thought that it was his wife, not that it was his wife and in situations of this kind it is permissible to save one's wife." ("Persons, Character and Morality" in Williams (1981)). Mendus comments

"The force of Williams' argument ... is to suggest that it is not merely impractical and politically inexpedient to force this extension of the scope of impartiality: it is also, and crucially, a deformation of concepts such as love and friendship, which are what they are precisely because they are not underpinned by completely justificatory explanations. In the example of the man saving his wife, willingness to pose the justificatory question is, in part, an acceptance of this deformed model. ... For the justificatory question, however it is answered in the disastrous case, jeopardises the area of individual discretion and simultaneously threatens a quite general transformation in our understanding of relationships of friendship and of love."  
(Mendus (1996))

This attempt to derive individual discretion from our understanding of friendship is refuted by the recognition that I have been urging, that

---

<sup>19</sup> Clearly Williams' argument would apply a fortiori to a single level impartialist theory. Mendus is using Williams to qualify Barry (1995)'s two level impartialist theory: she suggests that one additional reason for allowing discretion in the lower level is to leave friendship and love undefiled.

sensitivity to situations in which loyalty must be overridden does not require conscious deliberation. The justificatory question does not have to be posed, so long as the agent acts appropriately. Moral discretion is lost - for in many circumstances it is be wrong to be loyal, and agents must act accordingly - but where friendship prevails, it need not be deformed by intrusive calculations.

I will argue in the next Chapter that instrumental justifications of partiality in friendship do deform the nature of friendship in the way that Mendus fears: for within consequentialism, sensitivity to the results of the calculations would be revelatory of deformed and conditional motives for loyalty. By contrast, in the Williams example, where circumstances permit him to save his wife, he does so because he cares about her. The justificatory framework remains in place: friendship and loving relationships are not damaged by recognition that all actions are subject to external critique. Further, even without personal discretion, spontaneity is not compromised. One can spontaneously do the right thing.

In fact Blum does allow that the lack of a prior claim is important in justifying actions springing from benevolent emotions but describes this as a mere background condition necessary before unconstrained partial action is possible.

"The most that can be claimed here is that the fact that the given circumstances are ones in which it is morally permitted to act without taking into account the interests of others whom one might help forms a sort of background condition to the conscientious person's acting from friendship." (Blum (1980) p.62)

But the distinction between such a background condition and "consideration which he actually brings to bear on the situation or proposed action" is vanishingly fine. A very similar mental procedure must be necessary to establish compliance with the background

condition as would be necessary to establish that no better action is currently available to the agent: if the former is incompatible with giving full rein to the benevolent emotions, so must be the latter.

Furthermore, the distinction is hard to justify. If it is relevant whether there might be a prior moral claim by others, why is it not relevant that some other action might create more happiness in the world? After all, Blum admits "the fundamental moral truth in the notion that each person's good is as worthy of pursuit, as is any other's" (p.66).

Blum's attempt to show that there are spheres of action which spring from the benevolent emotions and are thereby exempt from the full requirements of rationality therefore fails<sup>20</sup>.

A more radical claim on behalf of an emotion-based theory of morality would argue that what is right can only be determined by emotion -- that emotion plays a cognitive role, informing us what is the right course of action. This is suggested by Nussbaum (1985)'s treatment, following her interpretation of Aristotle:

Frequently it will be her passional response, rather than detached thinking, that will guide her to the appropriate recognitions. "Here is a case where a friend needs my help": this will often be "seen" first by the feelings that are constituent parts of friendship, rather than by pure

---

20

Similar considerations should address Railton (1984)'s concern that too much careful calculation of moral responsibilities can lead to "alienation" from one's emotions and from one's loved ones. His solution is to posit a form of rule utilitarianism, and to recommend the suppression of deliberation on occasion. If it is accepted that taking responsibility for one's actions does not entail having fully and consciously internally-justified one's actions in advance, then the problem of alienation disappears. To do things which are morally justified is not incompatible with doing them as a result of strong emotions, even if moral justification itself requires careful analysis.

intellect. Intellect will often want to consult these feelings to get information about the true nature of the situation. (p 188)

Note, however, that Nussbaum herself leaves the intellect in the driving seat: not only is it the intellect that is consulting the feelings, the intellect is setting the relevant categories for the emotions to utilise. How would this work in practice? Suppose it turned out that the emotions always aroused us to sympathy and action when a friend was in need and never when a stranger was in need; the intellect would then be obliged to recognise and compensate for the bias revealed in the operation of our emotions. The emotions can only be relied upon to the extent that they have proved reliable in the past when tested by intellect in hindsight.

An uncompromising reliance upon emotion to determine morality could no doubt be constructed; it might be used to provide a justification for partiality. It would however be hard to build a discriminating morality upon such a base: on what basis could the actions of the sadist, issuing as they do from her emotional response to different situations, be criticised? Nussbaum is clearly correct in limiting the emotions to an advisory role, but by the same token they offer no escape from the impartialist critique.

## **ii Projects and Commitments**

According to Kekes (1981), what insulates the individual from the requirement of impartiality is the fact that he has particular projects and personal relationships. Kekes and others have suggested that the fundamental value attaching to this feature of the individual confers a permission or an obligation to be partial in order to pursue his projects.

Kekes quotes Aristotle:

"Everyone who has the power to live according to his own choice should...set up for himself some object for the good life to aim at...by reference to which he will do all that he does, since not to have one's life organised in view of some end is a sign of great folly"  
[Eudaemonian Ethics tr. J. Solomon Oxford 1952 p1214b].

If, as Kekes implies, these personal projects can be chosen independently of considerations of morality, including impartiality, then individuals' duty to be impartial will be strictly limited, for individuals will incorporate their partial valuations of kith and kin into their projects.

But nothing in the quote from Aristotle suggests that the selection of such projects should be independent of morality. Even if we emphasise in the quote the freedom with which the individual is choosing a project, this autonomy by no means implies that any project freely chosen is of value. One may have discretion to choose well or badly, but that does not imply discretion to choose the standards by which one's choice is evaluated. Choosing a project may be a necessary condition to having a fulfilled life, but it is not a sufficient condition: one must also choose wisely.

Two distinct arguments are employed to insulate the selection of and commitment to personal projects from objective critique, including impartialist critique:

- a. an epistemic/metaphysical argument: since individuals' moral perspective embodied in their considered selection of projects is our only source for what is of value, so value must lie in such projects, irrespective of external critique
- b. an argument from autonomy: there is a value in autonomy that is not realised if an independent critique determines what projects individuals should choose.

## **The epistemic/metaphysical argument for immunity.**

According to Williams value inheres in action rooted in integrity:

"...one who displays integrity acts from those dispositions and motives which are most deeply his, and has also the virtues that enable him to do that." (Williams (1976) p.49)

Implicit here is that his own "dispositions and motives" inform the individual what is right, so that he is right not to heed the blandishments of conventional morality. Thus personal choice of projects according to Williams is foundational of value. Project selection itself cannot be subject to the critique of conventional morality; we may be enjoined to be impartial between the projects of others, but this is precisely because we cannot, without hypocrisy, open up their projects to moral scrutiny. Conventional morality, including impartiality, has no locus from which to criticise personal projects.

I take Williams to be making an epistemic claim: the only access we have to objective moral value is the perceptions of moral value that individuals embody in their choices of projects.

This is the line taken by Cottingham:

"Here, in a sense, epistemology and metaphysics coincide: the fact that it is from the "autocentric" perspective ... that we come to see what has worth and significance in our lives is ultimately linked to the fact that we ourselves generate that worth and significance by the intensely personal commitments and preferential networks of mutual interdependence to which we wholeheartedly devote ourselves. This claim is a complex and controversial one; although its acceptance lies (I believe) at the heart of virtue theory, it seems to be denied both by deontologists who define the domain of the moral in abstraction from

the individualities of human desire and affection and by consequentialists who see the value of individual lives as essentially derivative from their contribution to impersonally defined goodness." (Cottingham 1998).

The implicit argument here seems to be that to accept independent evaluation is to deny the worth and significance of personal commitments, whereas it is precisely those commitments that are the epistemic source and the metaphysical locus for what is of value.

It is true that the devotion of many to succouring the sick endorses our observation that such activity is of value, and the lack of any who are committed to turning over blank pieces of paper likewise confirms our judgment that such activity has no value. Yet, commitment is not a sufficient condition for attribution of value to an activity: there are indeed practices that inspire devotion that may without violence to the language be considered worthless (e.g. idolatry).

Further, pace Cottingham, independent evaluation of people's commitments does not have to imply "abstracting from the individualities of human desire" nor limiting consideration to "impersonally defined goodness", or otherwise being insensitive to the particularities of commitments. We can recognise the epistemic priority of individuals' conceptions of the good, and then build upon these individualities a set of moral theorems with which to offer an objective critique, a critique for example of those who choose projects that do not seem sensitive to others' concerns.

Is consensus in moral judgements sufficient to form a perspective for critique? If not, then Cottingham may launch a sceptical defence of personal discretion. In Chapter Six I will argue that this is too pessimistic. For now, note that there is certainly a consensus that projects can be judged by independent criteria: the onus of proof is with the sceptic. For

example, beneficence is an accepted standard against which *ceteris paribus* many projects are judged. The question is whether impartiality should be another such standard.

Scheffler (1982), like Williams and Cottingham, argues that the metric for evaluating projects must be personal. Their worth must not be measured *sub specie aeternitatis*:

"to have a personal point of view is to have a source for the generation and pursuit of personal commitments and concerns that is independent of the impersonal perspective"(p.57).

Scheffler terms such moral self-determination "the agent-centred prerogative" or agent-centred permissions. For Scheffler the epistemic-metaphysical shift involves noticing that we have a personal point of view generated internally and concluding that this is something that ethical principles ought to take into account:

"as Rawls has said 'the correct regulative principle for anything depends on the nature of that thing'... [B]y incorporating a plausible prerogative which allows agents to devote energy and attention to their projects and commitments out of proportion to the weight from the impersonal standpoint of their doing so, hybrid theories [consequentialist theories with agent-centred permissions] recognise and mirror the independence of the personal point of view"(p58).

However, Scheffler's defence of partiality ill describes the personal point of view as it is experienced by most moral agents, including those whose point of view includes recognition of partialist obligations. Most people consider that their personal perspective gives them an insight into what their obligations are from an impersonal perspective (i.e. "really", "objectively", "not merely as a matter of personal opinion".) They justify



their personal projects by appeal to the independent shared values which they feel that their projects are promoting.

In basing his claim for moral self-determination on the fact that we have a personal point of view, Scheffler conflates two different sorts of independence: it is true that we each have an independent view of what is correct; but it is not true that we regard ourselves as being "sources for the generation of value". Only the latter claim would sustain a prerogative to attach greater weight to our projects even when we ourselves recognise them to be sub-optimal from an impersonal (unbiased) standpoint.

Further, to suggest that projects are inherently valuable but not to be evaluated independently is itself an implausible position, being subject to the same form of critique as preference utilitarianism: it would force us to attribute value to the projects of a serial killer or a glutton.

The elaborate institutions of praise and blame and of other forms of moral evaluation witness the fact that, for instance, the projects of the selfless are favoured over those of the egoist, the impartial over the partial. There is scarcely an area of life in which we are not willing to pass judgement on our fellows (though we may well be loathe to wrest responsibility from them through legislation). Utilitarian, egalitarian, or other consequentialist considerations will figure in the metric of evaluation, and partial practices will require justification against these impartialist standards.

The principle of impartiality can legitimately be invoked to criticise particular personal projects and commitments and the actions to which they give rise. Accordingly, it appears that individuals do better to set aside their personal preferences (though not their personal perspective of what is objectively right) when selecting projects -- so that they can select them respecting the equal moral value of all persons.

Sometimes, however we do accept personal preference as an independent source of value not subject to rational critique. For instance if someone in a restaurant says she prefers to eat steak rather than lamb, we are unlikely to demand further explanation other than that she prefers it. The fact that she prefers the steak seems then to have some moral importance: it justifies a particular action. And on occasion it will differentiate between good and bad actions: it may be accounted creditable that someone should provide you with food you enjoy, or someone may be criticised for deliberately giving you lamb when it is known that you detest it.

In these cases, subjective preference gives rise to objective value. Can an argument be built on that foundation that we should accord value to the satisfaction of people's preferences whatever they are? In that case, my preference to see my child happy will justify me in favouring my child over another, especially over an orphan, for in so doing I am satisfying my preference as well as benefiting my child (whereas the orphan has no-one with comparably strong preferences to be satisfied by her happiness).

I shall argue that this argument does have some force, but its force is mitigated by the control we have over our inclinations, and its application is limited to preferences whose satisfaction yields pleasure. I shall argue also that what force it has applies also at the level of the state, justifying the state in recognising the satisfaction that its citizens derive from official discrimination. Hence the argument, even to the limited extent of its cogency, fails to distinguish between personal and political morality.

Nagel (1986, Ch. IX) is not happy with the objectification of the value of the satisfaction of preferences. He suggests that although we should accord value to desires for food, we should be less willing to accord value to the fulfilment of other wishes or projects, for instance an ambition to climb Kilimanjaro, or to play the piano.

The rationale for the distinction rests on the fact that pleasure and pain are objectively good and bad respectively, and what gives immediate pleasure and reduces pain, like food and shelter, should therefore be seen as universal primary requirements, even if the type of food or shelter or relief from pain required varies subjectively from person to person.

Nevertheless, Nagel does not wish to deny the importance of other types of motivation apart from the satisfaction of basic wants. For instance, the ambition to climb Kilimanjaro can provide a reason for action. But such projects (which will include many partialist preferences) provide only "agent relative" reasons for action: they can serve to justify the agent's pursuit of the goal but they cannot justify anyone else in helping him.

It is, however, difficult to justify limiting the attribution of objective (non-agent-relative) value to the satisfaction of basic wants. If a man will gain greater satisfaction from climbing the mountain than from eating well, why should we be justified in helping him to eat but not to climb the mountain?

It is true that on this basis the pleasure taken by the psychopath in his murdering provides us with reason to help him too. A similar argument was used above to discredit the immunity of projects from criticism. However, in this case the only claim is that the psychopath's pleasure weigh in favour of aiding him; the likelihood is that other considerations will overbalance it. Even the murderer himself will have many moral reasons to fail to satisfy his desires, for he is obliged to take into account others' pleasure and pain, as well as his own.

Hence the fact that partial practices yield pleasure does seem to offer a reason for them, even if such a reason will often be outweighed by other considerations. At the political level, if discrimination against minorities or against foreigners yields the populace pleasure, that does provide a reason for discrimination, even if that reason is unlikely to prevail against other considerations.

However, even when immediate pleasure maximisation would indicate that a desire be slaked, an agent must also consider the impact that her action will have on her tendency to have such desires. Restraint will often weaken desire, and satisfaction feed it. If an agent is able to adapt her desires, then she must determine which desires to foster and which to suppress, i.e. to what she should grant objective value by training herself to find it pleasurable.

For example, take again the man favouring his daughter. It is true that favouring his daughter will give him more pleasure than favouring the orphan. He does have a *prima facie* reason to favour his daughter. But, if humans are of equal value, he will also have reason to wish that his pleasure did not lie in favouring his daughter: he would then not have to compromise between maximising recipients' welfare and satisfying his own preferences. Practice favouring others might enable him to foster in himself equal or greater pleasure in succouring needier children. If so, then he may not be justified even now in favouring his daughter.

His pleasure in favouring his daughter generates objective value in exactly the same way as his pleasure in eating beefsteak. In both cases morality can endorse the value of the pleasure *per se*, while at the same time taking a view as to whether the underlying inclination is worthy of cultivation or suppression. If the taste for beefsteak is pernicious (as vegetarians will claim), then we will have less reason to support the beef-eater, and the beef-eater will have reason to fight his inclinations.

Similarly with the partialist. The equal moral worth of all individuals implies that partialist inclinations are to be discouraged, even acknowledging the fact that favouritism generates pleasure, and pleasure is good. The widespread existence of partialist inclinations creates only a limited *pro tanto* reason to endorse partial decisions.

Parallel arguments suggest that states should avoid choosing projects that discriminate against minorities and foreigners, even when under existing norms such policies would yield greater immediate pleasure; rather they should try to educate their people to prefer impartiality, so that non-discrimination, large aid budgets and open borders become vote-winning policies.

Furthermore, objective value attaches to personal preference satisfaction only in so far as it yields pleasure. It will not support the dutiful fulfilment of filial obligations to an unloved parent, notwithstanding a sincere (but perhaps not heartfelt) preference for the parent's well being. The selection of a project will only accord value to a project if the agent will gain some pleasure from its fulfilment. Similarly, there is no reason for a state to heed ideological racialism which is not associated with pleasure taken in discrimination.

Nagel, on the other hand, would allow individuals to attach weight to objectively worthless and pleasureless projects (though he admits that others would have no reason to support them). He uses the example of someone who wishes to build a statue to his god, or who wishes for posthumous fame. Nagel himself unsurprisingly equivocates when considering whether individuals would be justified in favouring such ambitions (supported only by "agent relative" reasons) over the objectively endorsed claims of others (for food for example). Yet he is inclined to think that morality must "leave room for some partiality toward oneself and one's personal concerns" (p.173), which must mean that at the margin such projects would take precedence. This is hard to credit.

Whereas in the case of projects yielding unusual pleasures (like the mountaineer's) Nagel's agent-relative reason for action should be elevated into a full-blown objective reason; in the case of projects which yield no pleasure, reasons which are merely agent-relative are no reasons at all. Certainly, the agents who are supposed to be relying upon

these reasons will not recognise them as merely agent-relative. When someone withholds charity, in order to build a statue to his god, she (the idolater) will not, as Nagel apparently expects her to, accept that building the statue is merely of value-to-her. She will claim that hers is the true god and that it is right, objectively right, that sacrifices should be made to build the statue. Else what sort of argument would she have, to defend her practice?

Nagel suggests (p 168) that things are valuable because we have an interest in them. But this is true only of generators of pleasure like food. Such things should be considered of value in direct proportion to the strength of the pleasure that they yield.

But many of the things which form people's projects are not like that: **we take them to be of interest because we reckon them valuable.**

This distinction is manifest in the ways we expect people to justify their choices. If I ask why you chose lamb rather than beef for lunch, all you need respond is that you prefer lamb to beef; no further explanation is required. On the other hand, if I ask you why you are donating to charity A rather than charity B, I will expect a response that refers to acknowledged (objective) criteria of value: perhaps that A is, in your view, more efficient, or that the recipients are in greater need.

We refuse to help the woman build her statue, or to favour her community, when we think that her project has no value -- and we think further that it should have no value for her. We might indulge her if we think that her project will yield her much pleasure. But otherwise projects are only worthy of support (from their proponent or from society) if they have objective, not merely subjective, value. To determine this, an impartialist critique is appropriate.

## **The Autonomy Argument for Immunity**

It is claimed that the existence of external or objective critique is inconsistent with full autonomy in project selection. The point has already been made earlier in this chapter that an independent critique of personal life does not entail legal or even social restrictions on freedom.

Nevertheless, the thought is that if there is an objective standard for criticising projects and commitment, one is morally bound to choose the project that best meets the standard; undermining one's autonomy.

Of course, if there is an objective standard, and this undermines autonomy, one might simply conclude: so much the worse for autonomy! It would be a brazen sceptic who argued that the value of autonomy is so clear, that we must conclude that there are no objective standards constraining action, at least in areas of personal life.

A more subtle approach is taken by Raz (1986, chapter 14), who recognizes that projects must be chosen because they have value (they do not have value merely because they are chosen), but nonetheless maintains that true autonomy requires choice unconstrained by objective standards. Raz claims that such choices are provided by radical incommensurability between many objectively good options. As discussed above, the prevalence of such choices, and their significance is greatly overstated by Raz. (We concluded that radical choice occurs only when there is rough equality of value.) But his claim that it is only such choices that embody true autonomy is worth investigation.

Raz's account of the value of autonomy is helpful:

"The ruling idea behind the ideal of personal autonomy is that people should make their own lives.

"Autonomy is opposed to a life of coerced choices. It contrasts with a life of no choice, or of drifting through life without ever exercising one's capacity to choose."(loc.cit.)

There is alleged to be a primitive value in making one's own decisions, being a full moral being.

Then Raz introduces an example of "a person who can pursue an occupation of his choice but at the price of committing murder for each option he rejects." Such a person, according to Raz, has no autonomy:

"Autonomy requires a choice of goods. A choice between good and evil is not enough. (Remember that it is personal not moral autonomy we are concerned with. No doubt is cast on the fact that the person in the example is a moral agent and fully responsible for his actions. So are the inmates of concentration camps. But they do not have personal autonomy.)...

"...If he is to be moral he has no choice, just as [a] person struggling for physical survival has no choice..."

The parenthetical distinction is the key one. The Kantian notion of autonomy is surely that of moral autonomy: that is surely the notion with primitive value.

The personal autonomy that Raz promotes consists in the availability of a significant number of valuable options from which to choose. However, the value of such radical choice is mixed: set alongside the sense of making one's own destiny must be the tragedy of rejected opportunities. To the extent that radical choice has value, it seems to be commensurate with other aspects of personal fulfilment. A society that allows individuals to make hard choices without state or social interference, demonstrates its respect for its members, and increases their scope for arbitrary



commitment. Raz claims that making commitments in these cases is creative of value: for making the commitment makes the chosen course (between two valuable options) the right course. This has some plausibility, but it seems a small matter. The choice between being a clarinettist and a lawyer is a difficult one: the ability to commit oneself to one avenue and to stick at it may retroactively justify the choice. But it is hard to understand why Raz places this at the centre of his moral code. It is possible to give or to withhold commitment even in the absence of hard choices: personal autonomy in the making of hard choices is not necessary for that element of personal fulfilment.

On the other hand, moral autonomy is fundamental to the moral value of a human being. The capacity to make moral choices is contingent upon an understanding of the difference between right and wrong. Being granted scope to exercise that capacity is then central to being an active moral agent, to leading a valuable life.

Raz denies this; he argues that there is rarely value in the availability of choice of bad options:

"Autonomy is valuable only if exercised in pursuit of the good. The ideal of autonomy requires only the availability of morally acceptable options..."

But this is a non sequitur. Moral autonomy is of value because without the choice between good and evil we would be unable to choose the good. This by no means implies that for a choice to be valuable it must be between goods.

According to Raz's argument, it would have been better had Eve been prevented from eating the apple, the choice of evil was of no value. But by eating the apple, Eve gained the capacity to choose between good and evil. That choice was sinful, but thereafter she had the capacity for doing

good too. It is that capacity to do what is right, that moral autonomy, that is of importance. And it is the freedom to choose between good and evil that should be respected by government. Raz's version of autonomy, the ability to choose between goods, is at best a subsidiary concern.

It may be that life would be more fulfilling in various ways (perhaps more fun) were there no moral constraints. However, to argue from such an observation to the conclusion that projects are immune from criticism would be akin to arguing from the observation that life would be more fulfilling if we could fly to the conclusion that we can.

A special version of the autonomy argument can be found in Williams (1973, section 5, p113). He argues against the use of a utilitarian critique on the ground that individuals must be allowed to form their own projects and commitments (presumably with whatever ultimate ends they choose) as it is this that will (incidentally) grant them happiness. Even if happiness is a maximand, enforced pursuit of happiness *qua* happiness would be self-defeating.

But it is hard to see why choosing a project according to a utilitarian calculus should be self-defeating. Utilitarians are not required to maximise their own happiness. In choosing a project according to a utilitarian calculus, the incidental happiness which it will grant an individual may come into the calculation, but this is not incoherent. Neither is it incoherent that they are pursuing the happiness of others, even if this involves helping them to pursue projects which are equally aimed at happiness maximisation. There may be a circle here but Williams has not shown that it is a vicious one.

In a utilitarian utopia each will pursue projects aimed at maximising the sum of total happiness. But not all projects will be the same, nor will all agree on what each person's project will be: project choice will depend on judgments regarding individual talent and inclination (because inclination

will dictate likely enthusiasm and own enjoyment) and judgment of consequences. Allowing individuals to exercise their judgment to determine the happiness maximising course is itself likely to prove an optimistic procedure; but granting discretion (autonomy) does not imply granting immunity from utilitarian critique.

Similar considerations weaken the force of the general case that projects must be chosen without reference to independent (impartial) critique. Individuals would no doubt feel cheated of some sense of fulfilment if they could only pursue projects which the state or some other external authority had ordained, but they need not feel so cheated if they themselves are determined to choose for themselves only a project with objective value, even if that value can only be ratified by the tenets of the conventional (impartial) morality of the moral community. (We see here again the benefits of reticence in making laws or in fierce social sanctions, both of which may undermine the sense of moral autonomy, for the attractions of bad choices are artificially reduced; but we see no argument to the effect that objective moral standards are lacking.)

Williams is concerned that pervasive critique undermines personal integrity and responsibility. Yet this need not happen: individuals maintain their integrity by in turn evaluating the morality of their society rather than accepting it blindly. But where they endorse a moral standard (as they should where these standards are correct) then integrity demands that they apply that standard to their own projects.

Scheffler (op.cit) terms his prerogative a "liberation strategy". The strategy liberates individuals "from the demand that their actions and motives always be optimal from the impersonal perspective" (p.62). Yet it is surely a confusion to suggest that we are not free when we pursue the (impersonal) right. We are free to pursue the right or to ignore the right. It is true that we are not free to determine the right; but that freedom is unavailable.

Williams and later Scheffler both use an alternative form of argument to support their contention that we are able to determine the right course for ourselves by reference to our own value system. Scheffler believes that there is a utilitarian standard that ought to be the principal guide to action. Yet he is aware that this standard is often knowingly yet guiltlessly ignored. He concludes that the existence of "the personal point of view" legitimates exceptions from the utilitarian calculus, enabling each to accord extra value to his own projects. Williams uses these exceptions as a general argument against utilitarianism and other forms of consequentialist morality.

There are however ways of explaining the exceptions to utilitarianism without recourse to moral self-determination.

I will focus on two examples of Williams (1973, section 3):

Jim is presented by Pedro, who is well armed, with the option of killing one of a group of Indians who have been randomly selected for execution in order to discourage insurgency in the region. He is told that if he kills the one, the rest of the group will go free; if he does not, the whole group will die.

George is offered a job in a biological weapons research institute. He disapproves of biological warfare, but he cannot obtain another job and life is tough (for him and his family) without one. And if he doesn't take the job, a biological warfare enthusiast, who will have no inhibitions in fulfilling his duties, will get it.

In each case, a utilitarian would find the choice obvious (the details of the cases, which I have omitted, are designed to ensure that this is so), but Williams notes that our intuitions give us strong doubts about the decisions.

"The situations have in common that if the agent does not do a certain disagreeable thing, someone else will, and.....the state of affairs after the other man has acted, if he does, will be worse..... I have already suggested that it is inherent in consequentialism that it offers a strong doctrine of negative responsibility: if I know that if I do X, O1 will eventuate, and if I refrain from doing X, O2 will and that O2 is worse than O1, then I am responsible for O2 if I refrain voluntarily from doing X." [Williams, 1973, section 5, p.108]

Williams thus suggests that what is wrong with consequentialism is that the doctrine of negative responsibility forces one to act in situations in which it would be better not to act. These actions compromise one's personal projects and one's personal integrity, the value of which consequentialism cannot recognise.

The obvious retort to Williams' examples is that they show nothing about integrity and projects; rather they show that there are deontological injunctions which outweigh considerations of the greater good. This is clearly the case for Jim: there is an injunction (the sixth commandment) against killing. It may be that Jim nevertheless ought to kill, but the fact that he hesitates betrays nothing about his personal projects or integrity, only that he gives weight to this injunction irrespective of utilitarian or consequentialist calculations. A similar argument could be constructed for George -- he may reason that the injunction not to kill has as a sub-clause an injunction not to aid the wrongful taking of life by others, giving him good reason to hesitate before taking the offered job.

The fact that Williams does not take this route suggests that he is uncomfortable with these unsupported deontological injunctions. Scheffler (1982) discusses both possible escapes from the unacceptable consequences of consequentialism: the escape via deontological injunctions, and the escape via integrity or moral self-determination (the "agent-centred prerogative" discussed above).

Deontological injunctions are what Nozick terms "side constraints": they prevent one from maximising the good in certain circumstances. Scheffler calls them "agent-centred restrictions", and he considers them to be paradoxical and lacking in moral justification. He finds it paradoxical that an injunction could prevent someone killing one person even when he could thereby prevent a dozen other killings: if killing is bad, surely we should minimise it; if it is not something which we ought to minimise, why is there an injunction against it at all?

If Williams accepts such arguments, then he may take his examples to force him to seek an alternative escape from consequentialism: the virtue of integrity purports to provide such an escape by granting individuals a measure of moral self-determination; this would allow agents like George or Jim to argue that the cost to their own integrity or to their own projects is just too great for them to follow the good-maximising course. (And this moral self-determination would incidentally provide latitude for partiality notwithstanding commitment at the political level to the doctrine of impartiality.)

The following is a stylised example of a deontological injunction:

an agent Alf is prohibited from harming innocent Ian, even where harming Ian will prevent murderous Max from harming, in an exactly similar way, Jack and Jill.

Scheffler goes through many possible explanations of what could be wrong with harming Ian, and in each case he concludes that more harm overall will be done if Alf fails to act than if he acts (this is my example; I am adapting Scheffler's general argument to apply to it). He concludes that the injunction cannot be upheld.

Scheffler or Williams would defend Alf's right not to act on the basis, respectively, of agent-centred prerogative or of integrity. Alf may hold

dear the fact that he has not harmed anyone -- he is not the sort who could harm anyone (even for the sake of others). It is his prerogative to be true to this personal code; perhaps his integrity demands it.

Is there an alternative explanation, involving a defence of the above-specified deontological injunction, for our strong intuition that Alf at least may be justified in not harming Ian?

One way of defending deontological injunctions is to invoke responsibility (and following Williams' lead in focusing on positive rather than negative responsibility). If Alf acts, it is he who is harming Ian; but if he does not act then it is Max who harms Jack and Jill. From Alf's point of view these two states are not equivalent: he is responsible for the harm to Ian but he would not be responsible for the harm to Jack and Jill. Scheffler does mention the argument from responsibility, but he dismisses it shortly, claiming that

"it cannot provide a rationale of the desired sort for the restrictions, for they are part and parcel of the moral view whose motivation is in question."(p104).

But even if "the conceptions of individual responsibility and personal duty appealed to in these suggestions ... stand as much in need of motivation" as the deontological injunctions (p103), the fact that Scheffler is advocating the abandonment of such well-established concepts of ethical practice as responsibility and duty should give us pause. A notion of personal responsibility that discriminates between direct agency and mediated agency supports many moral practices, not least the meting out of punishment.

In any case, the concept of individual responsibility is motivated -- by belief in free choice. Max has a choice. The case is set up in such a way that we assume that Alf is certain that Max will harm Jack and Jill, but

such certainty is not consistent with Max's status as a moral agent, who has a choice over whether he will act wrongly: we can never be certain that he will not relent, and this uncertainty, which underpins Max's responsibility, reduces Alf's. Conversely, Alf would be directly and personally responsible if he harmed Ian. No recourse to Alf's integrity is necessary to explain our sympathy with his predicament.

This sort of argument may not be convincing in every case, however -- for it depends upon at least a minimal level of respect for Max. Attribution of full responsibility to Alf for the harm done by Max treats Max like a machine; but in some cases we may feel that it is disingenuous not to treat the potential harm-inflicter as a machine -- perhaps if he has a track record, or is in turn being threatened by others. The more obviously murderous is Max, the less will Alf's inaction vis-à-vis Jack and Jill be excusable. Alf may then seem as responsible for the harm that Max will do to Jack and Jill as he would be for harming Ian. In such a case, is there any reason other than "integrity" for allowing Alf to hesitate before acting?

There is an additional argument that applies in these cases regardless of Max's character, indeed even if Max actually is a machine; this is the argument that distinguishes Alf's action in harming Ian from his inaction if he allows Max to harm Jack and Jill.

The act/omission distinction is notoriously slippery -- there are many examples in which an omission is more culpable than an action -- and there is also difficulty in definition. Nevertheless, that there is a morally relevant distinction is manifest in our judgement that, even if Max is a machine about to kill Jack and Jill, and neutralising Max could be achieved without cost (Ian would not be harmed), still inertia on Alf's part would not render him guilty of murdering Jack and Jill.

It seems that there are a number of relevant criteria connected to the act/omission distinction, no one of which will individually be sufficient to



incriminate or to exculpate an agent. Criteria that increase the culpability of an agent, and that normally apply to acts rather than omissions, are:

- the agent knew that the act/omission would have harm as a consequence;
- the harm was caused by an intentional voluntary movement;
- the agent altered his behaviour from that normally expected of him (whether through act or omission);
- the harm was the intended result rather than being the unintended side effect of an act/omission undertaken with a different object.

On this basis, Alf can argue that he would be more responsible for the harm he would do to Ian, as he would have passed the first three criteria, whereas in allowing Jack and Jill to be harmed he would be liable only on the first criterion. (Of course, Max, if a person, would be liable according to all four criteria.) Thus we can again defend Alf's hesitation without recourse to "integrity".

The moral significance of the distinction between act and omission may be traced more profoundly to Nozick's arguments against interventions that redistribute between individuals. Alf is contemplating redistributing wellbeing from Ian to Jack and Jill. If Alf says that he has no choice but to sacrifice Ian's wellbeing in order to succour Jack and Jill, Ian will respond that

"He does not get some overbalancing good from his sacrifice, and no one is entitled to force this upon him" (Nozick, 1974, pp32-3).

Scheffler comments on this argument of Nozick's:

"The passage suggests that the rationality of agent-centred restrictions ... as a response to the separateness of persons derives from the fact that the violation of such a restriction harms some victim without compensating him for this harm. But the relevance of this fact is unclear ... why isn't it at least permissible to disrupt one distinct individual life, without compensation, in order to prevent the uncompensated disruption of five equally distinct lives?" (*loc.cit.*)

There is a way to read Nozick's point that avoids this criticism. The lack of compensation is not something which is supposed to distinguish Ian (if he is harmed) from Jack and Jill (if they are harmed); rather the fact that Ian would not benefit from the salvation of Jack and Jill leaves Alf without an argument to put to Ian. Alf has no right to intervene and take from Ian and give to Jack and Jill. No third party has the right to take away one person's life (or physical wellbeing), even in order to award life to others<sup>21</sup>.

Indeed, this is the libertarian explication of the principle of equality of human worth, incorporated in the procedural fairness principle (f-), discussed in Chapter One. Procedural fairness enjoins us not to diminish any individual's happiness: yet that is precisely what the consequentialist would have us do. Hence far from justifying an escape into personal moral self-determination, Williams' cases simply illustrate the tension between utilitarian and fairness objectives, both of which rightly exercise each of us.

The injunction to avoid 'playing G-d' (f-) relies on the distinction between act and omission; otherwise Jack and Jill would be able equally to accuse Alf of taking from them to benefit Ian. The implausibility of such an accusation from Jack and

---

21

Only if Nozick's argument is as I suggest, rather than as Scheffler interprets it, can we make sense of Nozick's point that "his is the only life he has." This emphasises that the wrong suffered is too great for the lack of consent of the victim to be ignored.

Jill should serve to convince that there is a genuine distinction along the lines suggested.

So it is Alf's uncertainty as to whether Max will really harm Jack and Jill, coupled with his reluctance to play G-d and his feeling that action is more culpable than omission, that explains Alf's hesitation in harming Ian. Recourse to agent-centred prerogatives or integrity is not necessary.

Williams may actually accept the universal case for non-intervention in cases where one person will be harmed for the sake of others. In the examples cited in defence of moral self-determination based on integrity (Jim and George), he constructed cases in which action does not result in anyone being harmed who would not be harmed anyway.

The action/omission distinction can still arguably do some work: in the Jim and the Indians case, even were Pedro a machine, Jim would bear less responsibility for the death of the twenty Indians killed by the machine than he would for one that he himself killed. But the lack of a plaintiff who would question Jim's right to intervene makes the reduction in responsibility seem marginal, and Jim's hesitation more like squeamishness. So in these cases, Williams' claim that we need the notion of integrity to explain our sympathy with Jim's predicament seems better founded.

But it is doubtful whether we would actually have any sympathy for Jim in such circumstances. If Pedro is a machine (a machine that will kill twenty, unless it is tampered with in which case it will kill a particular one of the twenty), Jim ought not to hesitate, he ought to do as all the Indians beg of him and shoot the single Indian, notwithstanding the greater culpability he would bear for an action rather than an omission. It is only because Pedro is not a machine that Jim is right to hesitate; he hesitates to consider how Pedro will react to his hesitation, whether

he might not relent after all; perhaps Pedro was never serious, he was just testing the foreigner's appetite for blood<sup>22</sup>.

Thus there are plausible defences of deontological injunctions, based on the responsibility of the mediating agent, the absence of warrant for redistributing life (playing G-d; f-), and the act/omission distinctions. Deontological injunctions will serve to explain our intuitions without resorting to personal prerogatives or to the virtue of integrity.

My position is one which Scheffler acknowledges only in a footnote:

"one can imagine moral conceptions that did include agent-centred restrictions [deontological injunctions] but that did not include any permissions not to produce the best states of affairs except those entailed by the restrictions".

Scheffler denies the existence of agent-centred (deontological) restrictions (Nozick's side constraints) but supports agent-centred prerogatives or permissions. The position I have been developing, and that Scheffler contemplates in his footnote, is one that denies agent-centred permissions, but accepts that there are deontological side-constraints. In Scheffler's utilitarian world, such side-constraints are seen as permissions to deviate from maximising behaviour. The thrust of his argument suggests why he largely ignores my position: he regards agent-centred restrictions as a more extreme form of agent-centredness, one which perversely not only recognises the individual's right to deviate from the promotion of the good but insists that he exercise that right in certain circumstances. He would regard my position as even more perverse: for I (following Nozick) have suggested that where there is

---

22

Similar analysis would also justify George's hesitation: perhaps he should not rely on someone else being wicked enough to do the job more callously than he would.

a permission to deviate from maximising behaviour there is also an obligation to do so.

But this interpretation of agent-centred permissions and restrictions as both part of a "fully-agent-centred" conception of morality, is based on a confusion; for the permissions and the restrictions are agent-centred in fundamentally different ways. The restrictions are agent-centred in that they are non-consequentialist -- they represent restrictions on what any agent may do despite untoward consequences of the omission. The permissions similarly allow an agent to do things despite untoward consequences, but they are also agent-centred in a much more radical way: they allow the agent to define for himself what is right. As we have seen, such autonomy is at odds with what our conception of the "right" and is not adequately motivated by the cases brought by Williams or by Scheffler.

## **2C CONCLUSION**

I have suggested that Scheffler's agent-centred prerogative, like Nagel's agent-relative reason, Williams' recourse to integrity, and Raz's version of autonomy, is at odds with our moral practices -- we act and talk as if we believed ourselves to be pursuing what is objectively right. This is not to deny the importance of personal projects, but rather to insist that the projects can be judged from an impersonal standpoint, with impartiality as a criterion in the critique. Personal projects cannot insulate the individual from his obligation impartially to respect the equal worth of human beings.

Of the defences of partiality considered in this chapter, the most convincing are those that stress that people enjoy being partial. A consequentialist calculus will attach weight to the pleasure arising from partiality to loved ones and to relatives. And if pleasure and joy are reckoned goods, then these concomitants

of partiality must be included in moral decision procedures even by non-consequentialists

Any resulting moral justification for partial practices is not, however, due to any personal 'liberty' to be partial. To the contrary, the defence rests on the presumption that partialist inclinations or desires are not subject to choice -- no freedom is exercised in having these desires.

To the extent that individuals have freedom to restrain partialist desires, or to form relationships that eschew partialist obligations, then second order choices will have to be made. The principle that enjoins us to treat all with equal concern and respect will then dictate that these second order choices minimise partiality. Hence the principle Pb, that indulged partiality in circumstances of uncertainty, or where the sacrifice of fairness or other goods is small, could be modified further to recognise existing prejudices in the moral calculus of right action, so long as the impact of current decisions on future prejudices is also brought into the reckoning.

This contrast between the partiality allowed if one's preferences are taken as given, not to be questioned, and the impartial demands made of those in control of their preferences, desires and emotions, echoes Sandel's (1982) criticism of Rawls' individualism. The latter is said to posit a radically detached self that is prior to its involvements and commitments. Such an individual has no trouble abiding by impartialist norms. But Sandel and other communitarians contend that it makes no sense to detach an individual from her position in society and her commitments - these characteristics are definitional of the individual, without them 'she is no-one'. The implication is that partialist obligations stemming from an individual's place in her family and community are to be respected.

A synthesis of these two positions is possible. It would accept that at any particular time a person is defined by her place in her community and can legitimately take account of that position and the preferences that accompany it in determining how best to act. Yet just as political philosophy asserts that the state should strive to be impartial notwithstanding the prejudices of its citizens,

so personal morality can demand that individuals strive to reform their preferences to reflect the equal worth of all human beings. Permissions and obligations to be partial that stem from the way things are will thus be insecurely based; subject to criticism from a perspective that embraces the way things could be.

This chapter has argued on the assumption that (I') above, the principle of impartiality based on the political conception of equal concern and respect, is sacrosanct; we have seen the limited extent to which partiality is consistent with it. However, this is not the end of the matter, because (P'), stating that we do have partialist obligations (not just permissions based on ignorance or human weakness, or obligations based upon sub-optimal inclinations) is also a well-founded meta-ethical observation. The conflict between impartiality and partiality remains unresolved.

### 3 INDIRECT JUSTIFICATIONS OF PARTIALITY

In discussing the tension between (I') and (P') (see section 1E), I have so far been considering arguments that assume space for partiality must be carved out by limiting the scope of impartiality, on the assumption that many of our partial practices are both justified and inconsistent with the principle of impartiality, (I'). This I believe to be the case, but it is a contentious claim, for consequentialists have claimed that the bulk of our partial practices can be justified within an impartial framework.

Justification would proceed by showing that each duty (or permission) to be partial is subsumed under a universal principle that can be formulated in a way that **does** treat each moral agent as equal -- in just the way required by (I'). In this chapter, I explore the extent to which such justifications are successful.

Several possible instrumental justifications of partiality are explored. The first, and the one most frequently discussed in the literature, is the attempt to subsume partialist practices under the impartialist maxim enjoining, *ceteris paribus*, that welfare should be maximised. Sections 3A and 3B discuss such eudaemonic instrumentalist theories. These sections also discuss some of the empirical and motivational problems that apply to instrumentalist justifications in varying degrees: notably the paradox that belief that partial practices are justified only instrumentally by the welfare they produce will tend not only to offend our sense of what loyalty should be (we prefer loyalty, especially loyalty to individuals, to be untainted by ulterior motives) but also to vitiate their efficacy in delivering welfare.

Section 3C extends discussion to cover aesthetic instrumentalist considerations, which may justify bias towards one's own cultural group.

The uncertain empirical grounds upon which at least some instrumental justifications rest, and the problems inherent in conditionality, render instrumental justifications of partiality at best vexed.



Hence, in Chapter Four, we must consider what theory would support unconditional loyalty. This leads us in Chapter Five to reevaluate our commitment to the doctrine of equal concern and respect, and to the maximand of fairness.

### **3A Eudaemonic instrumentalist justifications**

The moral theorem that follows from recognition that happiness is a good, is expressed by the imperative

(h) "Pursue human happiness, *ceteris paribus*!"

This theorem appears to have very wide acceptance; it also satisfies the egalitarian requirements of (I'). (Some will prefer to pursue pleasure or welfare or preference satisfaction; at this point it is not necessary to distinguish between these maxims, and I use the terms loosely.)

An action that deliberately fails to maximise happiness stands in need of justification. Thus if I am told that there are two charities to which I can contribute £1, one of which will absorb 20 pence in costs and the other 25 pence, such that one is more efficient at relieving hardship than the other, I do not need to give further justification for giving to the more efficient charity, but I **do** need to justify giving to the less efficient one.

Were there no other relevant considerations, this happiness-generalisation would be a form of utilitarianism. But the happiness generalisation, unlike utilitarianism, can allow for exceptions, so long as they have independent justification. Happiness maximisation is in itself neither sufficient nor necessary for the justification of an action, but it is sufficient if there are no other relevant considerations. Thus although the happiness-generalisation is consequentialist in nature, it is consequentialist in what I hope is an inoffensive manner, for it

does not exclude the possibility that there are non-consequentialist (as well as consequentialist) exceptions to it.

Many partial practices seem to represent unjustified exceptions to the happiness generalisation. For instance, someone might give to a less efficient charity because it benefits his own countrymen. One might justify this exception with a maxim such as

(c) "Charity begins at home!"

But this maxim is in conflict with the meta-theorem (I'), as it discriminates between groups.

However, (c) could be reconciled with (I') were it not claimed as a moral theorem with justificatory power, but merely a generalisation over a number of cases, in each of which the partial action was justifiable directly from (h). Perhaps giving to the domestic charity will usually be happiness maximising, even where it is less efficient at relieving hardship, because of the greater happiness that will accrue to the donor when he identifies closely with the recipients. Indeed this greater happiness may explain why people do give more to domestic charities.

Furthermore, additional pleasure may accrue to the recipient when the donor is a kinsman: we value the attention of friends and relatives, whereas we feel demeaned by the charity of strangers.

Along these lines, Brink (1986), citing Sidgwick, claims

"that agents should adopt a differential concern for their own projects and the welfare of others close to them.... Moreover, the possession and pursuit of personal projects and the development of close personal relationships involving mutual concern and commitment are a source of great utility."(p.426).

Brink makes this a point in support of rule-based utilitarianism; it would apply equally to act-based utilitarianism.

However, two considerations detract from the force of this instrumental argument for partiality:

- fairness (f+) will often militate against partial actions, even when welfare maximisation would justify them, for it will often be the case that the stranger whom one could help less with the same effort starts off by being substantially worse off than one's friend
- the implied dependence of the welfare arising from personal relationships upon partiality is questionable. In section 2Bii above, in discussing the pleasure we get from favouring our friends, we suggested that it might nonetheless be better to avoid partiality: the direct pleasure of the partial actions might be outweighed by the long term benefits of suppressing partialist inclinations. As partial inclinations waned, loss of the pleasure that used to accompany satisfaction of partialist inclinations might be made good by pleasure taken in impartiality, whilst better targeting of beneficiaries would increase the sum of their happiness (assuming they too could be weaned off their prejudice against the beneficence of strangers).

How realistic is it that we could wean ourselves off pleasure from partiality? If an impartialist rebel tried to conduct his life without any partiality to friends, relations, compatriots, he could well find himself without friends, disowned by his relatives and spurned by his compatriots. Yet, isolation might simply be the fate of the individual rebel. It is possible that a society might be brought to expect no partiality from its members, without diminishing the welfare arising from marriage, friendship etc. (again remembering that it would still be justifiable to favour the friend more than the stranger when the friend would benefit more than

a stranger in material ways<sup>23</sup>, as might often be the case, given that friends will be more sensitive to each other's needs<sup>24</sup>, and this fact would ensure that the institution of friendship was not weakened inappropriately even in an impartialist world). After all, expectations on these matters have varied greatly, and partiality in the form of patriotism, nepotism, old-boy networks are much less expected, or indeed acceptable, now than they were even two generations ago.

Whether the partial or the impartial stance is more often correct from a consequentialist perspective is thus an empirical question, depending on whether the joy of being partial can be substituted, and on the extent of inefficiency (utility and fairness foregone) that partial practices produce. Does the gain from leaving partial inclinations intact exceed the consequent loss in efficiency?

Relevant to the empirical question is the tractability of agents' partialist preferences. How much impartial practice and of how long a period would be required to remove partialist inclinations? And what would be the emotional cost? It seems likely that partiality is in part instinctive and ineradicable, in which case the question is whether we should attempt to suppress these inclinations somewhat, so as to breed a little more impartiality into the world. Or will we merely engender guilt by attempting to suppress instincts that will anyway persist?

A further difficult empirical concern is that suppressing partiality might result in suppression of beneficence: the pleasure accruing to the agent in being partial

---

<sup>23</sup> i.e. in ways other than as a result of the recipient's preference for receiving favours from kith and kin.

<sup>24</sup> "It is often difficult to know what would benefit others and, even when one does know, one is often in a poor position to produce some of those benefits without great cost to oneself." (Brink (1986), loc cit). This is the omitted section from the passage quoted above; however it does not justify a "differential concern for ... the welfare of others close to them". On the contrary it implies that one will inevitably do more for those close to one even if one maintains strict impartiality.

may be a major motivator of kindness, and bestowing kindness raises welfare for both parties.

Hare (1982) uses the example of a mother's partiality for her newborn child, which impels her to provide for it to a much greater degree than she feels obliged to look after other children:

"Ought a utilitarian to condemn this partiality ?..... If mothers had the propensity to care equally for all the children in the world, it is unlikely that children would be as well provided for even as they are. The dilution of the responsibility would weaken it out of existence". (Section 8.3)

Hare goes on to point out that even 'Evolution (if we may personify her) has had the same idea.'

Yet Hare reaches this conclusion too readily. One could argue that the world would be a better place if mothers came to regard all children as equally deserving of their care as their own. Legal obligation, and, more importantly, the fact that they are likely to be best able to do so might well ensure that mothers looked after their own children, but the children of the rich might find themselves sharing a little more with the children of the poor; orphanages might become depleted.

However, Hare may have in mind, in advocating that this inclination of motherly love be sustained, claims not between her children and other children but between children and other projects. Perhaps the only effective way of ensuring adequate childcare is to accept and even to reinforce instinctive parental love for their own offspring. It may not be psychologically practicable to extend this duty to other children without diluting it.

In this vein, Hooker writes

"Obviously, the world will go better if people care not only about themselves but also about their children and others. Would the world go even better if everyone cared no more about any one person than about anyone else? No, the costs would be too great. Human beings, as currently constituted, are incapable of caring strongly as well as equally about everyone else. So equal concern for all would mean strong concern for none. Better strong concern for some plus weak concern for the rest than only weak concern for all." (Hooker (1995), footnote 10)

These arguments in support of maintaining current levels of partialist commitment, have plausibility, but not the cogency that the authors assume for them. As we have seen, they are based on a swathe of untested empirical hypotheses.

We do get pleasure from partiality to close friends and relatives; this pleasure provides some justification for partial actions; and the case for weaning ourselves away from such partial inclinations might be undermined by a convincing prognosis that we would substitute impartial indifference for partial benevolence.

But whether the evidence is such that we should in general be content with our partiality, entitling us to build partialist generalisations on our acknowledged partial inclinations, or whether we should rather be suppressing partiality to gain the efficiency and fairness advantages that accrue to the impartialist, is uncertain.

Different but equally hard empirical questions arise regarding other partial practices. For example, the claim that chastity (a partialist commitment) is necessary to support the institution of matrimony is plausible, not least because chastity does not appear to be instinctive. But there remains the empirical question of whether the pleasure foregone in chastity is compensated by the pleasure afforded by the institution of marriage. As above, one might defend

marriage by claiming that the alternative to conjugal devotion would be not free love, but loveless licentiousness.

But the outcome of none of these empirical calculi can be gauged with confidence.

### **3B Inadequacy of Conditional Loyalty**

#### **i Exceptional Cases Dilemma**

We have seen that a colourable, if not cogent, case can be made to back a generalisation across certain classes of partial actions that they may tend to be happiness maximising even when compared with a world in which we had attempted to wean ourselves off partiality. Hence there may be no second-order reason for suppressing these partial inclinations.

However this justification of partiality, based as it is upon the pleasure accruing from particular partial actions, is marred by what I will call the "exceptional cases dilemma".

Even could strong empirical support for a partialist generalisation be provided, there would still be exceptional cases. Dealing with such cases presents a dilemma for the instrumental partialist. One horn is accepting that she should only be partial when the particular action is happiness maximising; the other is to adopt a partialist rule and to be partial even in cases when that action, viewed in isolation, is not happiness maximising. I will show that neither horn of the dilemma can be grasped - hence that a satisfactory justification of the partial practices that we applaud is not consistent with consequentialism.

Assume that experimentation proves that favouring one's friends is happiness maximising, because without the added pleasure of bestowing kindness

particularly upon one's friends, fewer kindnesses would be exchanged, diminishing happiness. Hence we would conclude that we should not suppress the bias that leads us to favour our friends. Yet, if this is to be our apologetic justification for partiality, the agent will know that, were it not for the human weakness of being unable to sustain kindness without the pleasure that arises from favouring friends, the happiness-maximising course would be to assess ability to benefit from kindness impartially. Hence the moral agent will feel rightly bound to review in each case whether she would maximise happiness by acting impartially: is the particular pleasure taken in this partial act sufficient to outweigh the greater impact that her kindness would have if directed impartially; after all, she would not be justified in applying the partialist rule where it is not happiness-maximising.

Yet, to appraise partial behaviour by its tendency to maximise happiness is to make transparent to oneself, and perhaps to others, the nature of one's moral motivation. Paradoxically, this revelation is liable to undermine the eudaemonic impact of the partial behaviour that is contemplated: the beneficiary of the partial behaviour will see the partiality as a mere by-product of a general benevolence, albeit one distorted by indulgence of human weakness; he may indeed feel demeaned as the common target of utilitarian charity. For the agent too, regular reminders that it is the bias of her own benevolence that justifies her in her partiality may sap the pleasure she takes in it; she may prefer to think that her actions are guided by altruism, whereas on this calculus they appear self-indulgent. The pleasure of both parties will be impaired.

A more detailed example might help. Suppose people gain happiness from valuing and feeling valued by the other members of the nuclear family, and that empirical research suggests that a certain degree of partiality between family members is requisite to foster this feeling of mutual support and security. Hence family members would have no second order reason for suppressing this bias. Suppose further that someone is informed that his sister is in financial difficulties, but he knows that she will never find out that he knows, indeed that she would be distressed if he knew about her problems. Should the benevolent



brother arrange for a covert transfer of funds to his sister, or should he give his money to some stranger to whom he could provide greater relief than he can to his sister (perhaps because the stranger is threatened with eviction, whilst his sister merely with repossession of her car)?

His sister gains no partialist pleasure from being favoured by him: she knows nothing of the source of her good fortune. So, on the consequentialist calculus, the only reason remaining to him for favouring his sister is that so doing will give him greater pleasure than he would gain from favouring the worthier stranger. Suppose that the extra pleasure that he would gain from succouring his sister would indeed outweigh the shortfall in his sister's benefit compared with that of the third party. Hence he should favour his sister. But a consciously utilitarian calculus might upset this reckoning. For he might suffer a corresponding loss in utility from the knowledge that his action was being steered by his own biased benevolence: it will seem to him a mean thing that he is favouring his sister merely to satisfy his own prejudices if, as we are assuming, he does not credit those prejudices with any independent moral validity (loyalty itself, we are supposing, is not a virtue). This thought might be sufficient to determine that the optimific course is to favour the stranger: that course would appear to the agent to be the selfless one.

Yet not to support his sister in this case would surely be considered by most people to demonstrate a lack of brotherly love. Hence, the partiality justified by a utilitarian calculus would probably be weaker than current practice.

The problem is deeper. If the sister is aware of the nature of her brother's moral theorising, she will see his motivation for favouring her (on occasions when she is aware of it) as impure, being founded on happiness-maximisation. This consciousness will sap her pleasure in being favoured even when he is openly kind to her; in turn further weakening the utilitarian justification for his kindness.

Consider also marriage: assume that evidence is adequate for the propositions that partiality to spouses is necessary to sustain marriage, and that marriage is

an institution that contributes to overall happiness, perhaps because mutual devotion within it serves to underwrite each party's self esteem. This may nonetheless not be enough to justify partial practices. The problem is not merely that devoting an evening to the Samaritans, or to dishing out soup to the homeless, will often increase happiness by more than staying home with one's spouse, even taking account of the incremental benefit to the relationship of staying home; such compromises might be consistent with sufficient acts of partiality to sustain the marriage. The real damage would be done by the mutual awareness that any favours that were granted resulted from a happiness-maximising evaluation in which they counted equally with all others, a calculus in which it is the agents' pleasure in partiality rather than heightened concern for their partner that tended to steer the calculation towards partiality.

In both examples there would still be some comfort gained from partiality: both spouse and sibling might be supported by the thought that their counterpart does takes pleasure in favouring them. My suggestion however is that this comfort will be offset by the knowledge that it is merely pleasure in partiality rather than love<sup>25</sup> or loyalty that is raising their weight in the agent's calculus. In a transparently consequentialist moral world, there is nothing that agents can do to reproduce the eudaemonic benefits of true loyalty.

Stocker (1976) argues (with cogent examples) that someone attempting loyalty within a consequentialist or other contemporary moral system will be driven to a form of schizophrenia in which her reasons are not embodied in her motives. This he sees as a sign of failure to achieve a "good life". I have argued beyond this that this schizophrenia may well undermine the empirical foundation that would justify partial practices within such moral systems. This impairment of the benefits of partiality under a consequentialist framework might tilt the balance in favour of the principled suppression of partialist inclinations: such welfare gains from partiality as might be salvaged from this morass of convoluted motives may

---

<sup>25</sup> Perhaps love is merely pleasure in partiality. But I assume that the beloved would hope that the efficient and defining cause of love, as of loyalty, is the Other.

be more than offset by the cost in inefficient allocation of favours. (Of course, this argument would reinforce Stocker's conclusion that the consequentialist framework is inadequate, but in this chapter the discourse is constrained to that framework.)

To escape this conundrum, to sustain pleasurable partial practices, perhaps the brother ought to help his sister **as if** brotherly love was independently valued. Similarly the spouses might devote themselves to each other without thought for happiness maximisation. But if they thus ignore the utilitarian foundation of partiality, they will also fail to notice the exceptions - the cases where (even taking account of the full pleasure in partiality that both parties enjoy in this unreflective mode) it would be happiness maximising to be impartial, i.e. cases where the need of the stranger is great compared with that of the spouse or sibling.

This constitutes the second horn of the instrumentalist's dilemma when confronted with exceptional cases: to be partial nevertheless - the rule-utilitarianism approach.

The rule-utilitarian theorist may go so far as to conclude that it will be best if people follow the partialist rule without ever knowing that it is founded on happiness-maximisation. Thus the brother should favour his sister without hesitation, because he is aware only of the virtue of loyalty and not of its fundamental source: he will have to **believe** that brotherly love is independently valued.

The consequent opacity of morality may itself be seen as a drawback to this form of utilitarianism. This is compounded into self-deception if one accepts that for the institutions of brotherly love and of matrimony to survive and to be thought of as independent of happiness-maximisation, their value will need independent foundations. These foundations will, according to the theory, be spurious. Yet they must be convincing, otherwise the true source of fraternal and conjugal obligations may be uncovered and the institutions weakened.

Were the evidence of the benignity of partial inclinations powerful, these problems with alienation might be borne. Parfit (1984, section 17) demonstrates (against Williams) the soundness of Sidgwick's view that the regretability of "self-effacing consequentialism" (a consequentialism that concludes that it is best suppressed in favour of some other more incentive-compatible doctrine) is no refutation of its truth. (Parfit judges that self-effacement should have exceptions, that a few should hold on to the truth in case the false theories propagated amongst the masses need modification as circumstances change.)

But the chief problem for the rule-utilitarian in this context is that a much stronger empirical case is needed to justify a partialist rule than is needed to justify heeding partialist inclinations in evaluating particular acts. For on a rule based account, but not on an act-based account, agents will find themselves being partial on some of the occasions in which partiality leads to loss of welfare.

I say "some" rather than "all" because I am assuming that even a partialist rule would be defeasible if the welfare (or other) cost of partiality was very great. (To be loyal does not require renunciation of general welfare as a separate maximand, one that can outweigh the claims of loyalty on occasion.)

Nevertheless, a rule follower is bound to make mistakes from an act utilitarian perspective: for example, in cases where a gross act of disloyalty would cost little happiness in itself but would yield significant benefits (perhaps betraying a dead father's indiscretion to exculpate an innocent stranger), the rule follower will weigh the disloyalty rather than the happiness, and judge that even significant benefits could not excuse such betrayal.

The cost of such disloyalty is inevitably over-valued by the rule follower, who is prevented by the rule from attending to the eudaemonic foundation of the virtue of loyalty. Indeed in the case of partiality this diversion of attention is the very purpose of the rule; for recognition even of the possibility of such exceptions will vitiate the sort of full-blooded loyalty that yields pleasure to both parties. (Other justifications of rule utilitarianism, e.g. those adverting to the demonstration effect of rule-breaking, might be consistent with constant awareness of and

sensitivity to the eudaemonic purpose of the rule; but such rule utilitarianism does not differ from act-utilitarianism from the perspective of the argument here.)

Hence on this horn of the dilemma too, it seems possible that an adequate investigation of the consequences of sustaining partial inclinations and partial practices, in this case by adopting partialist rules, would show that we would be better off without them.

## **ii Hare's Double Think**

The subtlest of the attempts to address some of the problems raised by the exceptional case dilemma is probably Richard Hare's, in "Moral Thinking" (1982). The approach is to find a way in which we can develop a simple reliance upon a partialist rule, backed by an uncomplicated emotional commitment, and yet maintain an ability to notice and to deal with the more egregious exceptions.

Hare's project is to explain how two competing forms of justification (in our case, the partialist and the underlying utilitarian) can co-exist without institutionalised ignorance or false belief of the sort discussed above.

According to Hare, the true answer to any moral difficulty will be determined by what an ideal observer, with perfect access to logic and the facts, would assess to be the outcome that would maximise the satisfaction of preferences, the latter being weighted according to the strength with which they are held. But the structure of his argument would apply equally to other forms of consequentialism.

Hare suggests that we are prone to error in assessing what an ideal observer would make of a particular situation, both because of lack of the time or resources necessary to determine the true outcome, and because, in the heat of the moment or when we have an interest in the outcome, we may be biased in our judgment. Therefore, it behoves us to establish maxims of conduct. These maxims must be inculcated from youth, so that individuals' automatic reaction in

difficult circumstances is to abide by them. Indeed, Hare believes that this is precisely the status of our moral intuitions: they are devices that society has bred into us to compensate for our imperfect ability to calculate the correct decision on the spur of the moment and/or against our own interest.

At least in the case of partiality (and some of the intuitions Hare defends explicitly are partialist) we have seen that we can add to his case for eschewing calculation the thought that calculation can diminish pleasure for both agent and beneficiary.

Along these lines, one might argue that the intuitions of brotherly love are bred into us in order to support the institution of the family, an institution which (we presume) overall promotes happiness but which might be compromised were ad hoc moral decision-making attempted on each occasion. Crucially, Hare would argue, even if thorough examination occasionally prescribes actions contrary to one's brotherly-love-intuitions, and even if it is well-known that this occurs, the intuitions themselves will not be undermined, because they will have been so deeply embedded that they operate at an emotional rather than (or as well as) an intellectual level.

This structure may then be adequate to preserve the pleasure in partiality for agent and beneficiary: both may be aware in theory that the correctness of partial actions can only be established via a utilitarian calculus. However, *ex hypothesi*, the agent's inclinations to be partial will have been reinforced by practice, and only rarely will they have been muddied by extended conscious deliberation ahead of action; hence the agent's pleasure in being partial will be little diminished by her theoretical training. The beneficiary too will delight in the agent's pleasure in giving, seeing it as evidence of a pure regard for his welfare, in practice unmediated by any cold and demeaning utilitarian calculus.

This then is an attempted solution to the exceptional case dilemma: rarely will an agent have the time and the ability to notice such cases, so she will generally rely upon her partialist intuitions, and sustain her pleasure in partiality. Yet some of

the exceptional cases that do arise, for example when she has adequate resources for thorough evaluation and is not subject to strong emotion, will be dealt with dispassionately and appropriately, reducing the adverse consequences of the inculcated intuitions (the rules).

The approach relies upon exceptions being fairly rare. Otherwise, if the exceptions are easily recognised as such, the intuition will inevitably be undermined (as the agent is a conscious utilitarian). Or, if the exceptions are not recognised, so that the partialist intuition is allowed to dictate behaviour, there may be such a loss of welfare that it would have been better not to have inculcated the intuition in the first place.

This need not be a problem for Hare: his theory allows him to criticise our intuitions, not only in particular circumstances but in general, as not being the best maxims to follow. For Hare is not arguing that every intuition is a justified rule of thumb: some are, but others are best eliminated or revised.

However, perhaps to avoid putting his readers in the awkward position of choosing between their own intuitions and Hare's argument, he tends to defend our common moral intuitions, and to try to show how rarely do utilitarian principles conflict with them.

Two stages are involved in Hare's defence of each intuition: (i) that the rule embodying the intuition does, in general (allowing for exceptions), maximise happiness/preference satisfaction; (ii) that recognised exceptions to the rule are sufficiently rare and/or outlandish that they will not undermine the conviction with which the rule is observed.

The trouble is that in the case of partiality, exceptions do not seem to be uncommon. If partialist intuitions are to be over-ridden in these cases, so that the operation of the rule remains beneficial (as (i)), it will by the same token be difficult to avoid the intuition being undermined (against (ii)). Conversely,

allowing people to ignore exceptions, so that they can be unthinkingly and pleurably partialist, the intuitions will not be beneficial.

Take again the example of maternal duty. If time for deliberation is available, this norm may be questioned and it may be determined in exceptional circumstances that it should be over-ridden. If the mother recognises those circumstances to apply, she should attempt to over-rule her inculcated inclination, and set aside her child's interest for a more deserving case. Hare supposes such circumstances to be so rare as not to undermine motherly love; even if a mother is aware of the possibility of such a case (and Hare is clear that he does not want the utilitarian foundation for intuitions hidden), her knowledge of the extreme rarity of such cases will serve to sustain her in her acting as if her commitment to her intuition were totally unqualified - as if motherly love was independently justified.

How rare are the exceptions? Hare ridicules an example of Williams, produced as a challenge for the utilitarian, for its absurdity. It involves the choice in an aircraft disaster between rescuing your son and a distinguished surgeon who, it is supposed, could save many injured passengers' lives. Hare finds it "hard to make Williams' example realistic. How do you know he is so distinguished a surgeon -- perhaps he was only shooting a line when you struck up an acquaintance in the departure lounge ?..... How promising is your son's future?" (op cit 8.4). Here Hare seems to be demanding complete certainty in order to overturn a moral intuition. This would seem appropriate if the moral intuition to be overturned was one which favoured others (in which case we might suspect ourselves of bias) -- but what business have we to ask such questions when other lives (the potential patients of the putative surgeon) are at stake?

Williams' example may be far-fetched, but Hare himself mentions in a footnote a far more likely example (from Godwin), in which Fenelon's chambermaid's daughter has to choose between saving the Archbishop and saving her mother from a fire. Hare concludes, like Godwin, that the Archbishop (assumed to be rather a good man) should have been saved. But fires are, unfortunately, not so



rare, and the circumstances in which utilitarian logic would tell you not to favour your child do not always rely upon such fine calculations as those establishing the relative value of the Archbishop and Fenelon's chambermaid's mother: one may simply be faced with the choice between certainly saving the stranger's child who is near at hand or probably saving one's own who is less accessible. Hare's archangel would surely require one to save the other child, in the absence of knowledge of their relative worth (or on the view that all humans are of equal worth) on the basis of the relative probabilities.

Hare is prepared to bet that the nearer such examples "get to realism and specificity .... the more likely the audience is, on reflection, to accept the utilitarian solution" (loc.cit.). I think this is doubtful for too many cases conflict with our understanding of partialist duties. In almost all cases in which a mother in the affluent world gives material benefit beyond that needed to sustain life to her own child, she would maximise happiness if she gave instead to some other child in greater need. (An argument that for reasons of emotional wellbeing a child needs to receive the attention of its own mother will be much less convincing with respect to the marginal material good than it would be with respect to the lavishing of physical affection. Note also that children are much less likely than adults to feel demeaned by the receipt of charity.)

This argument does not undermine the case for inculcating the intuition of motherly love in a utilitarian world so long as exceptions are respected, on the grounds that such a rule would on balance promote happiness (perhaps for the psychological reasons alluded to above, or in order to promote the habit of beneficence). Nevertheless it must be doubted whether such an intuition could be sustained in the face of these numerous exceptions: how could people so frequently made aware of the utilitarian base of partiality ignore it sufficiently to find pleasure in their own prejudice?

Note that in the case of the motherly love, neither of the two general reasons that Hare uses to justify unreflective reliance upon inculcated partialist intuitions applies. The mother's bias will be reinforced by the partialist rule of thumb, so it

is not serving the purpose of offsetting biased judgement. Second, it will often be the case that there will be adequate time for reflection to establish that partiality should be overruled. Little reflection is necessary for a mother to establish that there is a higher probability of saving the nearby stranger's child than her own from the fire. And there is ample time for her to reflect upon the greater happiness that she would effect by sending a pound anonymously to the shoeless neighbour's child rather than using it to buy her own child a toy. So the cases in which she ought not to be partial will be not only numerous but obvious too.

The only remaining justification for partiality will be the pleasure that she herself will gain from the partiality: this may perhaps be substantial, but may nonetheless be undermined by guilt that the determinant of her action is at bottom a selfish one; she knows that welfare would be enhanced if she could only wean herself off her partial inclinations.

Yet if she decides that she should not favour her child even in such cases, what will be the substance of her moral rule of thumb that motherly love is virtuous? A Harean mother will have to act against her child's interests on utilitarian principles. To preserve her motherly instincts she would have to forget her utilitarian principles not only in rare cases of haste, or bias, but all the time: and the claim that the result of such a policy would yield more happiness than a self-conscious pursuit of the greater good might be difficult to sustain.

Hence, Hare's solution to the problem of impure motivation fails to convince because of the commonplace occurrence of cases in which a Harean utilitarian would have good reason to overrule his partialist inclinations, so that belief that these inclinations are for the good could not be sustained. Only a total renunciation of the maximisation of good could sustain partial inclinations, but such a renunciation might not be happiness maximising.

In the case of motherly love, it is the commonplace occurrence of cases where partiality is not optimal that would undermine the inculcated intuition. In marriage,

mere awareness that partiality could be compromised on utilitarian grounds may sap the pleasure in loyalty: the Harean solution cannot work in cases where the parties are unable to ignore the utilitarian base of partiality at least some of the time.

Suppose a married woman happens upon a man who is desperate for affection, a man who is terminally ill and could not therefore form a lasting bond that would damage her marriage. Should she break her marriage vows on this occasion, make an exception which she knows would give much pleasure to the dying man? If her loyalty is dependent upon the utilitarian calculus her decision is clear cut. (I assume, that in this rare case, her own chaste intuitions are not so strong that her distaste for infidelity outweighs their joint sensual pleasure.)

However, if her spouse is aware that her chastity is dependent upon the non-occurrence of a chance to commit happiness-maximising adultery, their marriage might be undermined whether or not the infidelity actually occurs. And in a Harean world the spouse, being an informed utilitarian, will be so aware. Yet that fact cannot give his wife reason to be chaste, for her chastity in the individual case is insufficient to save her marriage in the face of awareness of its utilitarian base.

Elinor Mason (1997) adopts a position similar to Hare's in her defence of Railton (1984)'s indirect consequentialism against Cocker and Oakley (1995)'s attack. The latter question whether Railton's "counterfactual condition", which would induce someone in a relationship to "seek to lead a different sort of life" if they thought that would be optimific, is consistent with genuine friendship. Mason suggests that Railton is assuming that "human nature is rather inflexible ... the consequentialist has to argue that a disposition to be committed to relationships will have better consequences than a disposition to maximise good on every occasion." In other words, she argues that you cannot hope to sustain friendships if partiality is conditional at every turn upon an assessment of the consequentialist calculation, but for that very reason it may be best to train in oneself a fixed loyal disposition.

Mason, however, accepts that for dispositional consequentialism of this sort, the "empirical claims required to support an indirect consequentialist defence of friendship may on examination turn out to be implausible". This I suggest understates the case. First, a point that is missing from her discussion is that reduced partiality may not be inconsistent with friendship, albeit of a less fulsome kind. Second, as she does recognise, the opportunity cost of adopting a disposition to favour a spouse on every occasion may be severe. Third, neither Railton nor Mason reckon with the alienation (and possibly guilt) associated with the acknowledgement of the philosophical position that they advocate, one in which a large number of actions justified by relationships are seen as wrong in themselves, were it only possible to ignore the relationship on this one occasion.

For these and similar reasons, we cannot be confident that eudaemonic instrumentalist considerations justify many partialist rules or the inculcation of partialist intuitions or dispositions, and even apparently happiness maximising individual partialist actions may lose their eudaemonic effect once the utilitarian base is exposed.

### **3C Aesthetic Instrumentalist Justifications**

#### **i Maximising Happiness and Beauty**

We can widen the range of partial practices that may be consistent with (I') by considering instrumentalist justifications aimed at ends other than happiness maximisation.

We must first discuss whether there are other objectives for which we should strive that are not subsumed under the end of happiness maximisation: is happiness the only good? Then we can turn to see whether pursuit of these goals will instrumentally justify partial practices.

There is nothing incoherent in utilitarianism, still less with a theory that holds happiness as the only good while recognising other duties than the maximisation of happiness (e.g. the duty to refrain from taking life, as discussed towards the end of section 2Bii). Nevertheless, such a position would clearly be inconsistent with our considered judgments and practices -- so that the onus of proof is on those who would claim that other goods are reducible to happiness.

To see that we do not hold happiness to be the unique good consider the result if we were to try to produce a regime wherein people were trained to have that set of preferences that is most easily satisfied -- a happiness-maximising constellation of preferences to the easily satisfied. We would inevitably have to restrict preferences, whether at the individual level (Mill's contented pig) or at the societal level (Aldous Huxley's Brave New World): the more limited the aim the more likely is happiness to be achieved. This is certainly not amongst prevalent visions of utopia; the utilitarian system seems a contorted and narrow system compared with the set of goods that we actually value.

Consider the actions that one might do to benefit oneself. One may devote half an hour to eating ice cream, or half an hour to reading poetry. The same quantity of pleasure may be available from each experience (measured perhaps by how much one would spend to secure them). But second order desires will discriminate: one is less likely to try to encourage within oneself a taste for ice-cream than a thirst for poetry.

Why is it better to encourage a love of poetry? One can make excuses for this intuition: that one will get more enduring pleasure from the poetry, or that poetry reading will eventually give pleasure to others, for reading poetry is liable to make one more altruistic. But counter-examples could be drawn up to suggest that the real basis of the prejudice lies deeper and is indicative of the irreducible value of beauty, and of its appreciation.

That beauty is an independent good is also supported by consideration of the contrary position: were it good merely because it enhanced happiness, then it

would be better to produce pop-art than high art. There is no contradiction in asserting that one work of art gives more pleasure to more people yet is worse than another. Pop-art has utilitarian value, but if it is also very poor art it lacks aesthetic value.

Further, consider the trade-offs that we are prepared to make between the goods of happiness and beauty:

- A It is wrong to do significant harm to another person in the production or in the preservation of beauty. Thus, however great an artist's product turned out to be, murder, or even the desertion of family would not be justified.
- B Conversely, the destruction of art for the preservation of life (though, interestingly, not for its creation) is acceptable.
- C Personal sacrifice for the sake of art (as for eudaemonic ends) is commendable - for instance Michelangelo's suffering in producing the Sistine chapel ceiling is nowhere condemned as excessive. And we are justified in applying ourselves to the production of beauty in preference to applying ourselves to the production of happiness.
- D We value works of art as highly as human life in that we expend greater efforts in their preservation or in their restoration than we do in preventing many human accidents, and in that the mourning for their loss is likely to be as great. Thus, although in prospect we would not have condoned the sacrifices involved in the building of the pyramids, yet in retrospect we are not sure that it would have been better had they not been built. We may not, like Mme Swann (Proust I p.664), "prefer the death of millions to destruction of La Gioconda", but compare expenditure on its security, or on recovering it were it stolen, to expenditure on the security of citizens or the scale of investigations for missing persons.

E Public monies can be expended even in support of new art (despite the uncertainty of getting good value) even when we are certain that greater happiness would be secured with a marginal pound devoted to poverty relief.

Intuitions C, D and E capture our equal valuations of beauty and happiness as goods; they show that the asymmetry between A and B is not to be explained by happiness always being given priority over beauty.

Rather, the asymmetry between A and B arises because of the sacrifice of negative fairness that is involved in the pursuit of beauty at the expense of the happiness of others. (The negative versus positive notions of fairness (f-' and f+) were introduced in Chapter One, and the former was discussed further as a deontological injunction against "playing G-d" in section 2Bii).

(f-') "Do not diminish the happiness of another, without their consent, for any end, except to make good damage done by them to the welfare of others"

This represents a constraint on the operation of any consequentialist maxim; it has no counterpart forbidding destruction of objects of beauty. It explains part of the asymmetry in A and B, for sacrificing beauty for the sake of life and happiness, as in B, will not compromise (f-'), while sacrificing happiness for the sake of art (A) would be. It also explains why there is no contradiction between A and C -- we may be right to sacrifice our own happiness for the sake of art (C) but we would not be right to sacrifice others' happiness for the sake of art (A).

(f-') forbids the sacrifice of another's happiness (but allows the sacrifice of art) for either eudaemonic or aesthetic ends. But matters are more complex. As illustrated in Chapter One, (f-') could certainly be abandoned for a eudaemonic end: if the increase in a third party's happiness is very great, and the sacrifice of the second party's happiness is relatively small, then we would feel justified in harming the second party for the sake of the third.

Could (f-) be compromised for an aesthetic end? Are there circumstances in which we would allow a second party to be harmed in order to produce art, (notwithstanding intuition A)? Again there can be little question that there would if the art is great and the sacrifice small. Indeed such a trade off can be seen in the raising of taxes to subsidise the arts (F).

One reason why we do not approve of an artist damaging someone for the sake of his own art (A) is that the value of his own art is subject to great uncertainty. Conversely, the great value of the Gioconda is now established beyond doubt. Hence for the preservation of what we know to be true art, we are willing to inflict some diminution of happiness on others.

Let us assume therefore that we have established that there is at least one good apart from happiness, and we are labelling this "beauty", construing that term very broadly. As we all have a *prima facie* duty to maximise goodness, each man and woman must be justified in (and indeed obliged to engage in) striving for beauty in their own lives and their own output. As happiness does not appear to have lexical priority, the obligation to strive for beauty will (occasionally or often) entail setting aside the happiness-generalisation.

It might be thought problematic that we now have at least two sorts of good, beauty and happiness, with no means of weighing one against the other. But this problem does not arise just at this level. For the concepts of beauty and of happiness each embrace a multitude of particular goods, none of which is easily comparable with the others. Indeed, this is one of the classic arguments against utilitarianism -- how can one compare the happiness arising from different actions?

Although this is a problem of great practical significance, in itself it does not undermine utilitarianism as a theory, nor even as a guide to action: it is often quite clear that of two outcomes the happiness in one far outweighs the happiness of the other -- even if the happiness in the two cases is enjoyed by different people. Similarly, between happiness and beauty it will sometimes be



clear that the gain in beauty by pursuing one course far outweighs the gain in happiness from an alternate course of action. This is an area in which hard cases do make bad law.<sup>26</sup>

So for the happiness-generalisation at (h), we must substitute a combined happiness/beauty generalisation --

(hb) "*Ceteris paribus*, maximise happiness or beauty or some combination of these!"

Maxims like (hb) indicate the considerations that are relevant to determining action. They do not conflict with (I') because, although they specify a morally relevant quality other than the quality of being human, they do not specify a characteristic significant to a group of human beings, and so do not imply any inequality between the worth of different individuals (though see Chapter Five, where the significance of this defence is questioned).

What is the implication for partiality or partial practices? On the face of it, a whole realm of actions that might have been considered objectionably partialist will be legitimated if the beneficiary can claim artistic or cultural merit (construed broadly).

Much partiality involved in supporting one's own community may be interpreted as subsumed under a duty to sustain cultural activities, this duty being an implication of the second disjunct of (hb). Thus justified, partial practices will be compatible with (I') -- for the injunction to sustain cultural activities need not make any explicit reference to characteristics significant to particular groups of people.

The conditionality of aesthetic justifications is not self-defeating as was conditionality of eudaemonic justifications, for although conditionality could still

---

<sup>26</sup> Raz's contrary view, that difficult comparisons prove radical incommensurability, is discussed above in Chapter Two, section A.iv.

sap the pleasure in partiality of agent and beneficiary, aesthetic instrumentalist justifications do not depend upon that pleasure for their force.

Nonetheless rendering partiality thus conditional upon the contribution of the partial practice to the achievement of some aesthetic end would certainly demand revision of our understanding of partialist obligations, and such revisions may be difficult to sustain. Conventional forms of loyalty to friends or compatriots will be strained by the knowledge that the loyalty is conditional upon the promotion of the aesthetic end -- so that if an agent becomes aware that he is dealing with an exceptional case, in which abiding by the partialist rule would not promote the aesthetic end, then he must be disloyal.

This suggests that instrumentalist justifications do not fully capture our intuitive notion of what we are doing when we favour our friends or compatriots. For instance, suppose the English way of life is considered a good under the aesthetic banner. Not a good that other nations should be encouraged to pursue, but a good worth preserving nevertheless. To preserve the English way of life, a degree of patriotism and partnership amongst the English may be essential. Nonetheless, this does not legitimate all kinds of chauvinism or discrimination -- for the legitimacy of the behaviour is dependent upon the degree to which it fosters the good of 'the English way of life'. Therefore loyalty on this foundation cannot be "blind".

Empirical questions will also loom large in determining the validity of an aesthetic justification of partiality. Each partialist action will have to be scrutinised to see whether the marginal cultural or aesthetic benefit that it promises justifies the sacrifices in efficiency and fairness entailed.

## **ii The Epistemic Argument for Partiality Revisited**

The multiplicity of dimensions of value reinforces the scope for optional and uncensored partiality. In section 2Aiv we noted that the difficulty in ascertaining

the optimal course (from an impartial standpoint) gave scope for a permission to be partial. (The partiality legitimated by uncertainty was however optional; whereas instrumentalist justifications of partiality have the consequence, if successful, of rendering partiality in certain circumstances obligatory.) The scope for partiality based on epistemic limitations is increased if there are many sorts of good, because of the greater difficulty of making firm judgements between the relative value of different sorts of good.

This difficulty can be described as uncertainty as to which is the correct course. The uncertainty limits the application of morality and introduces scope for discretionary partiality.

It may, however, be thought that in such judgments there is often no fact of the matter, so that "uncertainty" ill describes the situation. Rather one might say that the multiplicity of values brings a greater degree of subjectivity into decision-making. In many areas the objective shades into the subjective as consensus gives way to fine judgment. Thus whether of two turquoises one is greener than another may be considered a matter of subjective judgement, while that a third turquoise is bluer than both the first pair may be thought a matter of objective fact. Similarly, whether a man should rather pursue an artistic career or devote himself to housing the homeless may be a matter of subjective judgment, while that he ought to set aside either to save a drowning man may be thought an objective fact<sup>27</sup>.

Little turns on this alternative taxonomy. In the absence of consensus, whether we describe divergence of views as uncertainty of objective fact or rather as an area for subjective judgment seems principally to depend upon the level of conviction people display: if people hold their views on the matter with great

---

<sup>27</sup> That this is consistent with the claim that both courses objectively have roughly equal valuation (*pace* Raz, on whom see above, section 2A.iv), can be seen from the analogy with colour: it is objective fact that the two colours are nearly the same, even if it is a subjective matter which is the bluer. Anti-realism in vague and hard cases is discussed further in section 6Dii.

conviction, notwithstanding the lack of consensus, we speak of subjectivity; if the lack of consensus is matched by a lack of conviction, we speak of uncertainty regarding objective fact.

Uncertainty (or subjectivity) is vastly extended by the need to forecast - if a man knew the results of his artistic or his benevolent endeavours in advance, the judgment of their comparative worth would be much easier for him and for observers.

Allowable discretion is further increased by the desirability of aesthetic diversity in combination with the uncertainty as to which aesthetic excellence is at the margin most required. Because I decide that it is good for me to devote myself to painting does not mean that everyone should devote themselves to painting. But this is not because others might disagree with my value judgements. On the contrary, everyone might agree that painting is a good, and that I ought to paint, but we might also all agree that not everyone should pursue painting as a career. Because my judgments about the best action for me are so intimately connected with judgments about my enthusiasm and abilities that few are able to second-guess, a large area of practical subjective discretion emerges.

This extension of uncertainty and subjectivity in the evaluation of action multiplies the occasions in which we will not be able to decide which is the optimal course of action, and therefore in which the optional exercise of partiality will be unobjectionable. Uncertainty underlies the "burdens of judgment", recognition of which Rawls uses as a criterion of the "reasonable". This view I discuss in Chapter Five. Rawls uses it to insist on tolerance in the political settlement of competing views of the good. However, the arguments here adduced do not suggest that uncertainty is so pervasive as to justify the general scepticism with respect to the good implicit in Rawls' argument. Indeed, uncertainty may arise because although we are certain about some values, these values happen on occasion to conflict. Note that even where we cannot decide upon the best course of action, we will often be quite clear as to which courses of action are sub-optimal.

### **3D Conclusion**

Three sorts of justification have met with a measure of success:

- In chapter two we saw that the exigency of a fully impartial morality, together with the incorrigibility of some partial desires, lead us to tolerate partial practices -- even though we recognise that we would be better to be impartial. This tolerance has narrow limits.
  
- In this chapter we have dealt with instrumental justifications. We found that they meet with two principle difficulties: that there is often insufficient evidence that partial practices are necessary to the goal in question; and that even when a partial practice in general was necessary it would always be subject to scrutiny to see whether each act of partiality was necessary to the end, a scrutiny that may tend to undermine the eudaemonic returns from partiality. Nevertheless, in some circumstances, partiality may be justified instrumentally.
  
- Both in chapter two, and in this chapter, we have seen that the extent of epistemic uncertainty as to what is the correct course of action (or perhaps metaphysical doubt as to whether there is one), creates considerable latitude in decision making, and there is in general nothing reprehensible in allowing partial considerations to steer action in these cases. But no obligations to be partial are created.

However, we have also seen that justifications consistent with (I') miss the emotive core of partiality -- unconditional loyalty. Stocker (1976) puts it like this:

"...it is essential to the very concept of love that one care for the beloved, that one be prepared to act for the sake of the beloved. More strongly,

one must care for the beloved and act for that person's sake as a final goal; the beloved, or the beloved's welfare or interest must be a final goal of one's concern and action." (p.533)

Loyalty is only really tested in circumstances such that whilst there are reasons for not being loyal, there is no ulterior motive for being loyal. Yet the principle of impartiality requires there to be some excuse for loyalty. Whatever instrumental justification we could adduce for loyalty will only be valid in circumstances in which loyalty furthers another end, and therefore in which loyalty is conditional and thus impure.

The next stage in the argument is to examine what does underpin our partial intuitions. We will then (in Chapter Five) be able to reevaluate our commitment to impartiality and fairness in the light of a better understanding of our attachment to partiality.

#### **4 HISTORICAL TIES AS SOURCES OF PARTIAL OBLIGATIONS**

Our consideration of instrumental justifications of partiality focused on happiness and beauty as the ends justifying partial practices. Apart from empirical issues, the chief problem with this approach was that the resulting loyalty was conditional.

Unconditional loyalty will result when the circumstance that identifies the particular individual or group to be favoured is historic, and hence not dependent upon future contingencies. The obligation may nevertheless be defeasible -- there will be circumstances in which claims of loyalty must be overridden, but if the loyalty is unconditional there will be no circumstances in which loyalty ceases to make a claim.

Given our conclusions in the last chapter, we may claim further that the reverse holds: unconditional partiality may always be backward looking. It will therefore tend to be in conflict both with maximisation of the good and with the egalitarian understanding of fairness (labelled  $f^+$  in Chapter One): maximisation of the good requires actions to be determined by forecasts of future benefits; egalitarian fairness demands that all individuals be considered equally worthy.

What are the circumstances we take to justify such non-optimific obligations to specified individuals? Two are: contracts giving rise to obligations to counterparties; and beneficence giving rise to debts of reciprocity (including those grounded in fair play, and in gratitude)<sup>28</sup>.

---

<sup>28</sup> Reciprocity, to be discussed below, is akin to a parochial version of the procedural notion of fairness  $f^-$ , a version in which Robin's favourable treatment of Simon is justified not by Simon's general beneficence but merely by kindness or work done to Robin's benefit.

Although obligations based upon contract and reciprocity accord greater respect to the individuals picked out than to others, the infringement of impartiality is one that we recognise even in political theory (the source of the doctrine of impartiality). Hence they are a promising area for exploration when attempting to reconcile the political and the personal. It is true that amongst partial practices, only marriage and business relationships are really sustained by contract, and the support that reciprocity gives to obligations of children and citizens is nebulous in concept and vague in extent. Nevertheless, the derogation from impartiality sanctioned in these practices is important and may be generalisable: historical contingencies creating relationships between moral agents are recognised in these cases as of moral importance.

In the next two sections, we will examine contract and reciprocity in turn to see to what extent they justify partiality. In section 4C, we take a crucial additional step in order to justify the bulk of partial practices between group members: we recognise a further good, that of groups in themselves. This recognition underpins the historical obligations of individuals to each other, for groups are constituted of such connexions, and are thus fostered by fulfilling such obligations. (For these purposes groups include relationships.) Rooting loyalty in the recognition of groups as goods also renders it conditional upon the groups actually being good (we do not recognise all groups as good) and upon individual actions actually contributing to the welfare of the group. However, this conditionality is less destructive than the conditionality discussed in chapter three.

#### **4A Contract**

As discussed above, it is not outlandish to assume that marriage is itself a happiness maximising institution, even granting the restrictions involved. The marriage contract helps the parties to sustain marital fidelity without continually questioning whether a particular indiscretion could be justified on the consequentialist calculus, a questioning that could undermine the relationship.



If the obligation to abide by contracts is itself grounded in consequentialist considerations, then it too might be subject to exceptions in the same way, and little extra strength will be added to marital bonds. However, perhaps contractual obligations can be defended on the utilitarian grounds that the institution of contract is beneficial to mankind. Such an empirical generalisation gives rise to a dilemma similar to those discussed above. If each promise itself is subject to the utilitarian calculus, then the institution of contract will be undermined because it will doubtless often be the case that the benefit accruing from breaking my word on this occasion will exceed the consequential damage to the institution of contract. It would be the fact that contract is thus contingent on the utilitarian calculus that would undermine it - rather than the impact of individual contract-breaches, which would occur only where the benefit outweighed the cost.

The other horn of the dilemma would be to accept that contracts should be kept regardless of individual circumstances. Perhaps in the case of contract the empirical case that generalised keeping of promises improves welfare is better attested than in the case of dispositions or inclinations to be partial to friends and relatives. Indeed, economics is now rife with theories of social capital that point to the indispensability of such trust to capitalist economies. This sort of justification is parallel to that attributed by Wolff (1996, Chapter 4) to Mill as a utilitarian theory of rights: in this case, the utilitarian calculus would point to the value of respecting contractual rights.

There is, however, a problem with such justifications from the perspective of the individual, at least one who is aware of the theory underpinning the contractual right. Although the utilitarian calculus can justify the state in instituting laws for the enforcement of rights (and, Wolff expounds the theory in that context), it does not follow that the individual is obliged to keep promises. The individual is aware that the institution of contract, the social capital associated with it, will not be undermined by her misdemeanour. So it seems that she is forced back upon a separate utilitarian calculus.

There is I believe a solution to this sort of free rider problem, which I will discuss in Chapter Seven. However, for the time being we may posit the keeping of contracts as an independent moral obligation, not instrumentally justified.

Although a non-instrumentalist obligation to uphold contracts would be in conflict with the doctrine of impartiality, as it requires one to favour the other party to the contract, such an obligation does seem to be widely recognised<sup>29</sup>. Ross (1930, p.38), suggests that it is a primitive duty, not requiring further justification. We all know<sup>30</sup>, as securely as we know that we ought, *ceteris paribus*, to maximise happiness, that we ought to keep our promises (and *a fortiori*: our formal contracts). Ross insists that we also know that happiness-maximisation cannot be the foundation of contractual obligation, for we are prepared to sacrifice at least a small quantity of happiness to keep a promise.

The belief that there is some independent justification for contractual duty is deeply embedded in our moral system, and should therefore only be given up reluctantly. Breaking a promise, particularly a promise enshrined in a formal contract (assuming contracts embody promises), seems a self-defeating action, almost sacrilegious. (The relevant biblical injunction, Numbers 30:3, uses the same Hebrew term as that used for profaning the holy.) The rationale may be that language, being the bearer of culture, shares whatever value is attributed to the common culture, and its abuse through lying, hypocrisy or contract-breaking, represents a betrayal of the culture.

---

<sup>29</sup> Indeed a non-consequentialist obligation to honour contracts is so widely recognised that the suggestion that it is in conflict with a duty of impartiality seems far-fetched. However, to claim that impartiality allows for discrimination in order to fulfil contracts begs the question: why is having entered into a contract with someone a relevant consideration in justice whilst being born of the same parents, or having shared a romance, is not?

<sup>30</sup> That we are sure about the particular practical judgments that underpin these moral theorems is to be taken as evidence that they are true. (The relationship between moral epistemology and truth is discussed further in Chapter Six.)

Furthermore, the infringement of impartiality involved in honouring contracts seems intuitively unobjectionable. Although, as we saw in Chapter One, impartiality is a firmly entrenched doctrine in political philosophy, and few would justify its infringement by government, yet states' duties to honour their treaty obligations are not questioned. Perhaps the fact that the existence of treaty and contractual obligations is widely recognised is itself reason for not worrying that the results will be partialist: we might expect even those discriminated against when a contract is honoured also to recognise the obligation, hence their self-esteem will not be threatened.

Recognition of an obligation to honour a contract would not, in the absence of other justification, allow one to enter a contract in the first place if that contract were not optimistic. (How could one bind oneself to do other than maximise the good?) However if we assume that marriage does tend to increase happiness, then entering the contract is legitimate. Once inside the contract, the spouses have, in the contract, an additional reason for favouring each other even in the ex-hypothesi-exceptional circumstance in which they know such favouritism to be non-happiness-maximising.

The contract helps by allowing them to separate the original happiness-dependent motivation that permitted them to enter into the contract from their day to day behaviour -- which will be motivated rather by respect for the contract embodying their mutual love.

Consider again the example of the married woman confronted by the chance of a once-off pleasurable illicit encounter. If her obligation to be chaste is founded merely on a utilitarian base, she will only be right to withhold herself if the lasting damage to her marriage will outweigh her immediate pleasure. As we discussed in the previous chapter, the mere awareness that her chastity is thus dependent on the non-occurrence of such a contingency could undermine her marriage whether or not she is disloyal.

The mere possibility of such transgressions breaks the trust that sustains marriage. Marriage involves a renunciation of happiness maximisation for the relationship; to be credible this renunciation must be embodied in a bond that trumps consequentialist calculation: this is the marriage contract.

Thus only the existence of the marriage vows themselves is sufficient to safeguard her chastity and her husband's trust. She would not only be being unfaithful -- failing to be partial to her husband -- she would also be breaking her vows. The contract in effect allows her to separate through a temporal break the two levels of moral thinking of Hare's analysis - she committed herself on her wedding day to be chaste and is now no longer obliged to reflect as to whether chastity is the best course. The knowledge that entering the contract was only justified because it was expected to maximise welfare will not now affect her observance of it.

It is true that the principle that contracts should be honoured will rarely underwrite partialist obligations other than those of spouses and of business partners. Nevertheless, the point of principle is that morality does recognise historical ties between people: the fact that we have had a certain relationship justifies me in future in treating you differently from others. This is fundamental to our conception of loyalty.

#### **4B Debts of Reciprocity**

Partialist obligations to parents, friends, and society are sometimes grounded in kindnesses received, perhaps on the basis of a moral theorem such as:

- (r) "Reciprocate receipt of benefits, even if unsolicited, wherever the required sacrifice of other goods is not too great!"

I here take debts of reciprocity, grounded in the benefits received, to include both obligations to repay kindness - debts of gratitude, and obligations to play

one's part in a common scheme - grounded in the notion of fair play. (There may in addition be a duty to feel gratitude, but this is not relevant to our discussion.) Simmons (1979, Chapters V and VII) differentiates these grounds of obligation, but neither ground is well defined, and the circumstances in which obligations are triggered are best considered together.

Theorem (r) is in conflict with the doctrine of impartiality, as it will lead one to favour individuals or groups merely because they have benefited you. It thus opens the way to a cycle of partiality. It might allow someone to favour her parents over others in recognition of benefits received, justifying the parents in reciprocating, and so on.

Ross claims that repayment of a debt is morally significant. He points out that it would generally be accepted that it is better to give a benefit of 1000 to one to whom you are indebted than a benefit of 1000 to another, if such is the choice. (Ross, 1930, p35). Ross was speaking of duties arising either from promising (in which case they are a subset of contractual obligations), or in repayment of debts, or a restitution of wrongs.

Such obligations seem much more powerful than those grounded in acts of unsolicited kindness, such as those received in childhood from one's parents and/or society.

Can we say that it is better to give a benefit of 1000 (or even 999) to someone to whom a debt of gratitude for an unsolicited kindness is owed, than to give a benefit of 1000 to another?

What, if any, obligation does come from receipt of such benefits? Take the case of an unsolicited gift from a stranger. If one is in a position to refuse it, acceptance may create an obligation. Thus wilful acceptance by the recipient is one criterion that will determine obligation.

But the child or the citizen is more on a par with someone who receives an unsolicited gift which cannot be refused. Perhaps someone arranges for a favourite piece of music to be played for you in a public place. Until you are informed that she arranged it, you believe your hearing it to have been a happy chance. Once you are made aware that someone has put herself out for you, however, it may be accepted that you are in her debt (perhaps only "morally"). The recipient's enjoyment of real benefit seems sufficient to create some obligation.

Simmons (1979a) in "The Principle of Fair Play" examines whether receipt of benefit can be adjudged a ground for an obligation to participate in a community's institutions.

The case he is examining is slightly different, for he is assuming that the benefit is conferred with the understanding on the part of the donor that a specific obligation is generated. The resulting obligation upon the recipient seems both stronger and weaker than that of the recipient of an unsolicited kindness: stronger because the donor could claim that the benefit would not have been conferred had she not thought that payment would be forthcoming; weaker because the act of conferring a benefit with an expectation of payment is not in itself laudable --it was not a gift, there was no kindness. Indeed, non-payment would seem to serve her right; she was trying to make a forced sale. Thus the kindness of the donor increases the obligation, as does any expectation of recompense, while guile diminishes it.

We now have five intuitively plausible criteria determining the existence and extent of obligations of reciprocity: two dependent upon the recipient (wilful acceptance, real benefit); three on the donor (kindness, expectation of

recompense, absence of guile)<sup>31</sup>. The degree to which any given act meets these criteria will determine the extent of any consequential debt.

Simmons emphasises the weakness of the position of the donor-expecting-receipt, quoting Nozick's objections to people having obligations forced upon them "One cannot, whatever one's purposes, just act so as to give people benefits and then demand (or seize) payment. Nor can a group of persons do this" (Nozick, 1974, p95).

But one might claim, following the above analysis, that the problem with this sort of case is not that there is no *prima facie* obligation on the recipient of an unsolicited good or service, assuming that it is a benefit and enjoyed as such, but rather that the guile of the donor in taking for granted your acceptance, offsets your obligation to her.

Nozick and Simmons emphasise the absence of wilful acceptance of the recipient as if it were a *sine qua non* for an obligation to arise. But what if we cannot tell whether the benefit would have been accepted if choice had been

---

<sup>31</sup> Simmons (1979) has separate lists of requirements for obligations of gratitude and for obligations of fair play (chapters V and VII). For example, he insists that some action of consent is necessary for an obligation of fair play, whilst the intentions of the donor are critical for an obligation of gratitude. This bifurcation excludes cases where a number of the other important criteria of obligation are fulfilled (e.g. significant benefits are received, absence of guile). Prime amongst such cases is receipt of benefits from the state, for which there is rarely active consent on the part of the beneficiary, and regarding which the intentions of the collective donor are hard to ascertain. Yet such cases do well on the other criteria, and thus intuitively seem to establish some form of obligation. It is hard, but unnecessary, to say whether such obligations are grounded in fair play or in gratitude - perhaps a bit of both. Hence, by considering the possible grounds of obligation separately, chapter by chapter, Simmons prejudices the argument against the existence of obligations to the state. As we will discuss in Chapter Seven below, the real problem with obligations to the state, whether founded in reciprocity or otherwise, lies in the difficulty in establishing for an individual that her acts will be of sufficient benefit to the state to override her other obligations (both partial and impartial).

possible? It is plausible to suppose that obligation is only categorically escaped if a benefit is wilfully refused.

Simmons accepts Nozick's view that unsolicited goods do not generate obligations, and extends the argument to a case of public or open goods, for instance the provision of a public address system, or an aesthetically-pleasing environment, emphasising that in this case, as much as in the case of unsolicited mail, there is no willing acceptance of benefits, the sort of willing acceptance which generates an obligation.

But the provider of public goods need not be doing anything wrong (and is thus not guilty of guile), for she may have been requested to provide the service, and promised payment, by other members of the community, (whereas e.g. to clean someone's car windscreen at the traffic lights without permission generates a weaker, perhaps negligible, obligation because it constitutes an attempted forced sale).

Furthermore, it is not always the case that consent could or should have been obtained. Even if a member of the community explicitly refused consent, this does not mean that the public good should not have been provided, if there was a strong consensus in favour of its provision, and it was objectively in the public interest. In many cases, universal consent is not merely not required, it is also impossible -- particularly when the public good, and payment for it, lasts more than one generation.

Simmons accepts that receipt of benefits can sometimes generate obligations in lieu of consent -- he mentions the case of someone who takes water from a public well. He also acknowledges that in cases of public goods the borderline between wilful acceptance and mere receipt of benefits may be fuzzy. But he firmly puts the onus on the donor to prove that the recipient *really wished to accept the good*. "If open benefits can be accepted [it must be by] taking the benefit willingly and knowingly". In particular, we escape obligation if we



"regard the benefits as having been thrust upon us against our will, or think that the benefits are not worth the price we must pay for them".

Simmons considers a case in which a member of a community refuses to do his bit towards the upkeep of the environment, arguing that, although he enjoys the improvement to the environment achieved by other members of the community, *his* time is too valuable -- he would prefer to have an ugly environment than a beautiful one purchased with *his* labour. Simmons admits that this man would probably be condemned by "ordinary feeling" but suggests that in this case, following Nozick, ordinary feeling may be wrong. He alleges that "ordinary feeling" is inconsistent in criticising this anti-social character while accepting that in the case of an individuated, unsolicited service, there is no obligation to pay any debt.

But the two cases can be distinguished as suggested above: there is no obligation in the latter case because of the element of deceit in the behaviour of the donor, and because in such cases obtaining consent is practicable; Simmons' unwilling gardener on the other hand, cannot complain that his consent was not obtained, for the scheme to improve the neighbourhood was entered into blamelessly. Thus there is no guile. Yet there is an expectation of repayment by the community in creating the garden, not least because the recipient acknowledges that he enjoys the garden. Perhaps in this case Simmons' "ordinary feeling" is right.

Simmons seeks to cast doubt on the unwilling gardener's responsibilities by considering the case of a non-resident who works in the neighbourhood and enjoys the benefits of the improvement to the environment generated by the scheme. Simmons says that an attempt by the council to get such a commuter to work would be laughable. The appeal of his example may be in the fact that the salesman, not living in the location, might find physical contribution more onerous than the residents. But financial contribution by non-residents is frequently -- and legitimately -- required, in the form of business taxes and sales taxes. This may form a counter-example to the old saw "no taxation without

representation" (in other words he did not give his consent to the scheme), but representation for visitors is impractical, its lack is thus neither culpable nor an excuse for non-payment.

That there is some obligation in such cases seems hard to deny: real benefits are legitimately (i.e. without guile) conferred, prior consent is impossible to obtain, and there is an expectation of recompense through taxation.

This case is very similar to that of medical, educational and defence services provided by the state for children, who are then expected to pay taxes when they grow up. Similarly, the unwilling gardener is comparable to a beneficiary of current government expenditure who voted for a tax-cutting party.

If it be argued -- as Nozick argues -- that this is just not good enough, that the state has no right to confer such benefits and expect repayment, consider the position from the state's viewpoint. Suppose there is an orphan child. The authorities have not the wherewithal to finance his upkeep and education except through a loan secured on his future earnings. (The authorities might argue it would be unfair to impose this burden on others through taxation.) Clearly, the orphan is not in a position to consent or refuse this arrangement, so the authorities go ahead anyway. When the orphan grows up, he is surely under a moral obligation to repay the loan. And in the case of current tax payers, the unwilling minority cannot deny they get some benefit from the state, which cannot be withheld from them, so is it unreasonable to oblige them to pay for it?

Similarly, the Talmud (Baba Metziah) considers a case in which a man had instructed labourers to plough his field, but they have mistakenly ploughed a neighbour's. The neighbour is required to compensate him. He was in receipt of a benefit, and, although he should ideally have been consulted as to whether he wished to receive the benefit, the fact that consent was not obtained in no way implies deceit on the part of the donor. In English law, also, in certain circumstances a provider of an unsolicited service can claim payment if a genuine benefit has been conferred through a misunderstanding of some kind.

Children's obligations to their parents may similarly be grounded in services received. There is usually no presumption that the services performed by parents to children are conditional upon repayment by the children. In a sense, therefore, the parents' position is weaker than that of the state - they cannot claim that they would not have undertaken the upbringing of their children had they not expected support later (although such a claim may be appropriate in some societies). On the other hand, the fact that the kindness was unconditional (and perhaps also that there is an identifiable donor whose sacrifices are measurable) would appear to strengthen the children's obligations.

There is however a problem in determining the extent of such obligations of reciprocity<sup>32</sup>. What can one say to the citizen who little values the services conferred by the state, claims that her debt is properly discharged with a year's rather than a lifetime's taxation, and that she would never have agreed to such a bargain, still less to more onerous duties to the state such as military service (the price expected for defence services received)?

Should the valuation of the services rendered be based on the counterfactual analysis of what the recipient would willingly have paid for the service, or on the cost of the services received?

Either way, reciprocity provides a problematic base for our obligations to family and state. Obligations grounded in reciprocity would be roughly proportional to the benefit received (on whatever valuation base is chosen), but obligations of daughter to parent or of citizen to state, do not in general appear to be so. A problem peculiar to relations of citizen to the state is that the state rather than the citizen generally stipulates the coin in which the obligation of reciprocity is to be

---

<sup>32</sup> Much more precise are the obligations founded upon tort: payment to put right damages inflicted or wrongs done. These are likewise obligations arising from historical incident, which will force diversion of resources from uses that would optimise welfare or fairness. As noted above, they represent the parochial version of f-', which sanctioned punishment of individuals in reciprocation for damage done to society.

repaid. In Chapter Seven we will suggest that this does not fit well with our general understanding of reciprocation.

Further, it seems plausible that the obligation a child has to his parent should be similar in ground to his obligations to his siblings and to the parents' obligations to the child. Similarly, obligations to the state ought to be linked to obligations to be partial to compatriots. In neither case would a ground in reciprocity forge such links.

Reciprocity like contract does appear to be of some importance in our moral life, and is at odds with a principle of impartiality that would ignore historical ties. And, as with contract, this exception to the principle of impartiality applies also at the political level. To the extent that partiality is justified by obligations of reciprocity, the inconsistency between personal and political morality identified in Chapter One is resolved.

However, as we have seen there are difficulties in basing partial obligations upon reciprocity. More is needed to capture the full range of our prima facie partial obligations.

#### **4C Groups as Goods in Themselves: Communitarian Obligations**

The derogation from impartiality sanctioned in reciprocity and contract is important and may be generalisable: historical contingencies creating relationships between moral beings are recognised as of moral importance.

(The reluctance of ethical theorists to canonise such obligations may relate to uncertainty regarding criteria of personal identity over time. Obligations by individuals to other individuals depend upon two such claims. Yet no firm criteria of identity over time have been established. It may be thought unsatisfactory for ethical obligations to be based upon distinctions without criteria. Similar problems arise in the case of retributive theories of punishment. However, the dependence of our ethical framework upon the concept of human identity over

time is not so much a problem for the ethical framework as one for the philosophy of mind.)

Both in the case of contract and of reciprocity the moral significance of the historical relationship was dictated by an apparently *ad hoc* non-consequentialist principle.

Perhaps there is an underlying source of value that gives significance to historical ties, and justifies both these practices and wider partiality including loyalty. In a nutshell the argument is this:

- Loyalty is only appropriate in the context of an existing historical relationship, of friendship, kinship, community. These relationships have intrinsic value, just as happiness and beauty have intrinsic value. Contract and reciprocity are **minimal requirements** that aid the forging of relationships based upon good faith and trust. Loyalty involves **strengthening** relationships that have value and that are already in existence.
  
- Groups and relationships are constituted of historically forged connexions between individuals who mutually recognise the significance of these connexions. (Is this an adequate definition of a group; does it capture what is morally significant about groups without restricting the class? The moral underpinning of groups is developed further in Chapter Seven.) The intrinsically valuable flourishing of good groups and relationships underpins historically based obligations of loyalty to these groups and relationships, as loyalty serves to sustain the groups/relationships.

There may thus be an ethically primitive obligation towards good groups of which one finds oneself a member. This might support: obligations arising from one's birth into a group; the responsibility of ex-patriates to provide financial support to their homelands; the responsibility of members of ethnic groups to

maintain their cultural traditions. Such obligations are at odds with the notion of equal worth (non-members whom you do not succour might have none to care for them); so they need justification.

Rooting loyalty in the intrinsic value of good groups also renders it conditional upon the groups actually being good (we do not recognise all groups as good), or at least not being evil.

In the rest of this section, I explore:

- i whether good groups and relationships do indeed have intrinsic value
- ii whether they give rise to obligations of loyalty
- iii whether loyalty to groups or relationships rather than to particular individuals is genuine loyalty
- iv the conditionality of the intrinsic value of groups/relationships upon their goodness
- v criteria for obligations of loyalty towards groups and relationships.

## **i Intrinsic Value of Groups and Relationships**

We demonstrated in Chapter Three that happiness is not the only good, and suggested beauty as a second good, but this list was not intended to be exhaustive. Ross argues for the existence of just three goods, all, according to him, belonging to states of mind: knowledge, related to cognition; virtuous disposition or action, motivated by the desire to do our duty (or to increase another good, or happiness); and pleasure.

As we have seen, Ross's method in proving that these goods are intrinsically good is simply to consider, for example, "whether, of two states of the universe having equal amounts of pleasure, we should really think no better of one in which the actions and dispositions of all the persons in it were thoroughly virtuous than of one in which they were highly vicious." (Ross, 1930, Chapter V, p.134) Fully independent valuation accrues if we go a step further and

acknowledge that we would prefer the virtuous world even if it were somewhat, perhaps just an iota, less happy. And it follows from that that we should be willing, in some circumstances, to sacrifice some happiness (at least our own) to virtue.

Similar arguments serve to prove the independent value of pleasure and of knowledge. Ross does not allow beauty as an independent good, identifying it with the power to produce aesthetic enjoyment. But his method could nevertheless be used to show that this power is valued independently of other pleasure, in that we would prefer a beautiful world to an ugly world even if the sum of pleasure in each world were identical.

Are such hypothetical arguments valid? Even if they are not, it would not be difficult to instance real life consensus decisions that manifest the independent value which we attribute to these goods (one example used earlier was the allocation of public funds to subsidise cultural activities).

It is not important for our purposes to produce a list of goods that is exhaustive. Nor does it matter whether they necessarily inhere, as Ross claims, in qualities of mind; if knowledge is good then anything productive of knowledge will be instrumentally good -- therefore things which are not mind will also legitimately be characterised as good.

Ross's simple methodology does however enable us to test whether groups are viewed as goods in themselves. This is certainly a category of goodness outside Ross's list of virtue, knowledge and pleasure. But all of his goods pertain to individuals; he did not consider valuations pertaining to groups. Hence his failure to include these goods need not be taken as indicative that he would not have acknowledged them as such.

We do value relationships to the extent that we sacrifice other goods to them -- accepting the pain with which they tend to be associated. This is very clear at the level of very small groups, of two or three, of loves and friendships: these

relationships are a prerequisite of a recognisably human life; and certain human excellences or virtues uniquely apply to relationships with others, the virtues of sensitivity and of love. These virtues or excellences, like the relationships in which they are manifest, may be considered primitives as goods, for the sake of which we are willing to sacrifice other goods, particularly happiness. Better to have loved and lost than not to have loved at all, even though the latter state may leave one with a higher level of cumulated happiness; it also leaves one less fulfilled or developed as a human being.

This collection of intuitions may command sufficient assent to justify inclusion of relationships as independent goods. Their independence is not compromised by the suggestion that relationships are valuable because without them we would not be fully human or fulfilled: the relationships are not instrumental to achieving fulfilment; rather just as partial actions are constitutive of relationships, so relationships are constitutive of human fulfilment.

## ii **Obligations of Loyalty to Groups and Relationships**

A eudaemonic instrumentalist might say that one ought to favour one's friend because without such favouritism friendship will fade, and happiness will be diminished. The last clause is an unproven empirical claim. But if friendship is good in itself, then we will be justified in partiality if we merely show our partiality to be constitutive of friendship.

Thus the empirical problems that bedevilled instrumentalist justifications of partiality (in Chapter Three) are removed by appeal to the intrinsic value of good groups. Furthermore no causal link need be established between partial actions and the flourishing of the group: partial actions constitute the flourishing of the group and are *ipso facto* justified (*pro tanto*).

Thus if good groups have intrinsic value, it is plausible to suppose that they found partialist obligations.



### **iii Whether Loyalty to Groups or Relationships rather than to Particular Individuals is Genuine Loyalty**

In Chapter Three, we were concerned that instrumental justifications of partiality failed to embody the true virtue loyalty, in that genuine loyalty had to be directed at specific individuals rather than at a further good (whether happiness or some aesthetic end). The present proposal is that the partial action is justified only in so far as it constitutes the flourishing of a good relationship. Is loyalty that aims to sustain the friendship rather than the friend pure enough?

Recall that Stocker (1976) insists that "it is essential to the very concept of love that one care for the beloved... the beloved, or the beloved's welfare ... must be a final goal." He goes on (p.537) to point out that a friend would feel betrayed if he really believed that he was being visited in hospital merely out of a sense of duty. Such a friend would feel less betrayed if it was their friendship that inspired his friend rather than abstract duty - but there might still be a sense of falling short of pure loyalty.

To test this thought, we need a case in which loyalty to a friend would inspire action, but loyalty to the friendship would not. Our example from Chapter Three, of the brother thinking to succour his sister financially without her knowledge might serve. In her ignorance, his subvention will not sustain their relationship. (Let us assume that his own fraternal feelings towards her will be unaffected: neither strengthened by vicarious interest nor weakened through rancour.) But here we can again appeal to the point that his loyalty to her is not instrumental to but rather constitutive of the relationship between them. In which case, he ought to help her simply as a brother, because sibling-relationships are good. Hence we find no practical difference between loyalty to the friendship and loyalty to the friend.

### **iv Conditionality of the Intrinsic Value of Groups/Relationships upon their Goodness**

Our recognition of relationships as goods is not without qualification. Not all relationships are thought good: some relationships are destructive of both parties.

As a result we predicate our endorsement of friendship upon an evaluation of the friends and friendships concerned: are the prospective friends good or bad people, will they enlighten or corrupt us?

(Our justification of partiality here begins to depart radically from the vision that informed our endorsement of impartiality -- that all people are of equal worth.)

Some have thought it odd to support friendship conditionally, to expect people to vet their emotional attachments. Kekes discusses the position of outright impartialists, who must of course also demand that attachments are vetted, though in their case the criteria may be impersonal:

"We are instructed `to prefer no human being to another, because that being is my father, my wife or my son, but because, for reasons which equally appeal to all individuals, that being is entitled to preference' (William Godwin, Political Justice). The passage of almost two hundred years has not mitigated the absurdity of the same sentiment that leads to advocating `a principle of mutual love requiring that people should not show personal affection and love to others on the basis of arbitrary physical characteristics alone, but rather on the basis of traits of personality and character related to acting on moral principles' (DAJ Richards, A Theory of Reasons for Action, Oxford, 1971)

"Imagine questioning the morality of Cordelia's love for Lear, Romeo's for Juliet, Pierre's for Natasha, the love of a mother for her mentally defective child on the grounds that their preference is arbitrary, because it does not apply to all people" (Kekes, 1981)

But Kekes' anti-impartialist polemic is not well taken. Even the extreme impartialist would not question the fact that love is a powerful and spontaneous emotion; rather they question what follows from the fact that someone loves another. Godwin claims that nothing follows, no extra obligations or permissions accrue to the lover or the beloved. Each particular act of partiality by Pierre or Romeo would need justification independently of the fact of love. This may be a hard line, but it is not an impossible one, nor is it inconsistent with the fact of love.

We may agree with Kekes that love and friendship are goods; that the existence of the emotions that sustain these relationships is morally relevant. Nevertheless, we may claim, with Kekes' antagonist, that partialist obligations will not flow from a love-relationship unconditionally. In general it is plausible that such obligations should only attach to relationships if the agents believe their love or friendship to be ennobling. However, it is also plausible that most love and friendship is ennobling, and thus that obligations to partiality, which sustains, and may even constitute these relationships (insofar as the relationships exceed mere emotion) will be commonplace. Nevertheless, each obligation will presuppose a fundamental evaluation of the relationship.

In the case of friendship, this argument is controvertible. However, few would deny that any goodness that inheres in larger groups is conditional on the characteristics of the group and thus on the members.

More clearly than in the case of friendship, we will be unwilling to view every group as instantiating a good to be fostered; if nationhood is a good, it is nevertheless only so for good nations; and whether a nation is good will depend upon what it does, and what it produces. This does not mean that the group is only good as a means to a further end; but its intrinsic value is dependent on its members' pursuing other goods<sup>33</sup>. Anarchy may be better than an evil state.

---

<sup>33</sup> Groups' instrumental contribution to other goods includes, for instance, their role in fostering the development of virtues by their members. The claim that good groups have intrinsic value entails that

(There may be honour among thieves, but the band of thieves to whom this virtue is applicable is likely in the popular imagination to be assimilated more to Robin Hood's than to Al Capone's; abhorrent crime appears to be made worse when carried out by a group.)

Is conditionality upon the goodness of a community consistent with analysis of what it is about community that renders it valuable?

Aristotle emphasises the incompleteness, the lack of self-sufficiency of a solitary man, who is compared to a solitary piece in a game of draughts, or a detached human limb (Politics Iii 1253a), arguing for a full-blooded articulated community. His imagery, justified in terms of the striving for human excellence, implies an obligation on every man to attach himself to a society in order to enhance his life.

But if the purpose of community lies in the enhancement of the lives of its members, it is clear that attachment to a society that is so corrupt that it will corrupt its members is worse than attachment to no society at all.<sup>34</sup>

Dependence upon groups in a biological sense or a psychological sense, which might not be conditional upon the morality of the group, will be insufficient to generate strong moral obligations. It may be the case that most people cannot cope with a solitary life, and therefore they are well advised to participate in society to avoid madness, but this does not confer upon them a significant obligation to support the group at the expense of other goods. (Such reasoning might at best generate a permission to be partial to the group, to participate in

---

there is some additional value in having such a group over and above its instrumental value.

34

cf. Maimonides, Book of Knowledge, Laws of Ideas, Chapter 6, "if all the countries of which one has personal knowledge, or concerning which he hears reports, follow a course that is not right - as is the case in our times - he shall live by himself in seclusion, as it is said "let him sit down and keep silence"" (Lamentations 3:28).

its self-supporting activities, akin to the eudaemonic justifications discussed in Chapters Two and Three.)

The view that there is value in community, but only in particular forms of community, is implicit in Marxism, and Communitarianism and other Perfectionist political philosophies. But the partialist implications of these philosophies are not always clearly stated: where one's own community is moving towards the desired end but other communities are not, partiality towards one's own group would seem to be justified by the good that is comprised by the group's characteristics.

Does the conditional nature of the support for the group undermine the purity of that support, generating the problems associated with instrumentalism? As in the case of personal friendships, the answer is that loyalty is not instrumental but constitutive of the good sought: the flourishing of the relationship or of the group.

This is also the answer to what might seem a problematic question: which group should I support? If the grounds for obligation to a group are a) that man is unfulfilled without a group and b) that a particular group will be advancing particular goods, then it could seem that switching to another group, whose ends perhaps appear better, is perfectly legitimate, even obligatory. A similar question might be asked with regard to family obligations: if the ground of familial obligation is the value of "the family", in what way does a person transgress if she abandons one family for another? Is she to be upbraided for broken objectives or disloyalty or commended if she reckons that she can better enhance the new family than the old?

This question - to which group shall I be loyal? - has a specious feel - it is not a question that anyone imbued with the virtue of loyalty would ever ask. This reveals the artificiality of attempting to divide the obligation from the good, as if one were instrumental to the other. It is in the nature of loyalty that it relates only to groups of which one is a member; its members' being loyal in ways that

support the peculiar good of that community is what constitutes the group's flourishing. Of course, it is possible to change groups, but perhaps not as an act of loyalty to the new group: the question of group membership is discussed further in Chapter Seven (section 7D).

There is a sense in which the loyalty to the good group, and to the good friend, although not instrumental, is nevertheless conditional, and this sense may seem to diminish the loyalty. For in both cases the loyalty will only last as long as the group or the friendship maintains those characteristics which render them good: if the group or the friendship becomes destructive, then their needs will cease to justify action.

Such loyalty falls short of that which may be idealised, especially in the case of friendship -- the friend may change his character, but we will less readily concede that the friend has changed sufficiently to abrogate duties of loyalty. However, recall that our loyalty is conditional upon the character of the friendship rather than the friend. There are circumstances in which friendship can be renounced, and most conspicuous amongst those will be where it becomes a damaging sour relationship. At this point in the discussion, that it is the friendship rather than the friend that we have justified as the focus of partiality aligns well with intuition.

#### **v Criteria for Obligations of Loyalty towards Groups/Relationships**

The good of large groups appears to be conditional upon the nature of the groups -- the quality of the people, the goods being produced, the moral culture. The nature of such goods will be particular to each group or relationship.

I do not wish to offer a general set of criteria of goodness in groups or relationships (though see the discussion in Section 7B); however, it may be helpful to review the criteria that Dworkin has suggested for determining the

existence of "obligations of association", which would include obligations of loyalty.

Dworkin (1986, Chapter Six, see especially section on "Obligations of Community") develops a defence of political obligation based upon the obligations of association, which extend across the variety of groups that we have been considering (friends, families, unions, political groupings, nations). Dworkin does not explore the foundation for these obligations of association, merely claiming that if such obligations are recognised, and if it can be shown that they do not depend upon personal emotional ties subsisting between members, then similar obligations can be claimed for members of political communities. This, as he acknowledges, leaves him without defence against those who "deny all associative obligations", and with no argument with which to persuade an official that he has special obligations towards his own community, and that he should not necessarily subordinate the interests of his own community "to improve justice overall". Recognition of the intrinsic value of good groups would address this deficit, bringing out that there is something of value that would be lacking in a world that did not recognise associative obligations.

Dworkin does however identify some features that he would require of a community's attitudes to their mutual responsibilities for these to count as genuinely "fraternal" and thereby justified (in so far as any associative obligations are justified). In our crude terminology these would be the features of a group that render it "good" and hence also "good in itself" underpinning obligations of loyalty. For Dworkin, they transform a bare community into "a true community". These features are:

- a the group's members must recognise "*special*" obligations "holding distinctly within the group, rather than as general duties its members owe equally to persons outside it."

- b responsibilities are "*personal*": they "run directly from each member to each other member, not just to the group as a whole"
- c responsibilities flow "from a more general responsibility each has of *concern* for the well being of others in the group"
- d "*equal* concern for all members"

and (Dworkin lists this separately)

- e "special responsibilities to my brother in virtue of our brotherhood are sensitive to the degree to which he accepts such responsibilities towards me; my responsibilities to those who claim that we are friends or lovers or neighbours or colleagues or countrymen are equally contingent on *reciprocity*".

The first of these, (a) special obligations, should perhaps be considered as that which is being justified by ascribing goodness to a community, rather than as a prerequisite for such an ascription. *Personal* responsibilities (b), and the requirement of *equal* concern (d) are I think special features of Dworkin's liberal conception of political life, rather than features that we should require of any true association. The requirement that responsibilities are personal is breached when a brother prioritises the good of the family over that of an individual member of it; proscribing this would promote the individual over the group - which may be a feature of the sort of political association that Dworkin favours, but is hard to justify as a necessary feature of, say, a string quartet or a rugby team.

Equal concern for all members is at the centre of Dworkin's conception of a just political association, but it is far fetched to claim this as a necessary feature of



any true community - Dworkin is surely here guilty of building conditions into his general characterisation of community in order to prove his favoured conclusion for the political community. See how Dworkin strains to fit familial or military associations into his mould:

"Fraternal associations .. may be structured, even hierarchical, in the way a family is, but the structure and hierarchy must reflect the group's assumption that its roles and rules are equally in the interest of all, that no one's life is more important than anyone else's. Armies may be fraternal organizations if that condition is met."

It is hard to see how hierarchy can reflect equality.

This characterisation manifests the liberal individualist's concern that obligations should be in the interests of each individual, rather than of some other good - like the flourishing of the group itself (under which head might fall the defence of the realm). This assumption is also behind the stress upon *reciprocity* (e): it is true that lack of reciprocity may indicate to an individual that she is not being respected by the group, but she may nonetheless be obligated to support the group because of the value of the group itself.

The most appealing of Dworkin's criteria is that of pervasive and persisting mutual *concern* (c), as a feature of good communities that we would wish to see flourish. But although this feature, like the others Dworkin identifies, may contribute to our valuation of a group *qua* group, a case has not been made to consider it a necessary feature. (By extension, the support that Dworkin claims from this foundation for his claim that "integrity" is a necessary feature of political community, is shaky.)

Nevertheless, Dworkin's insight that political associations are forms of communities, and that loyalty to political institutions and their ordinances can be justified in the same way as can loyalty to families and other true communities, is fundamental. It is the tension between this perception and the requirements of

justice (of "equal concern and respect") that leads him to characterise the good of community so narrowly. Dworkin concedes that, even within this narrow characterisation, loyalty to community can conflict with the requirements of justice "if the consequences for strangers to the group are grave". However, he does not appear to recognise that special obligations to one's community are *per se* in conflict with equal concern for members of other groups; and he therefore fails to notice that loyalty can only be justified if the flourishing of the groups commanding loyalty is an independent good, competing with justice and fairness, and occasionally overriding them.

## 5 QUESTIONING IMPARTIALITY AND EQUAL CONCERN & RESPECT, AND COMPROMISING IMPARTIALITY

As we noted in Chapter One (1D and 1E), a great many practices that are universally accepted as right involve some compromise of fairness in that they often result in distributions that are less equal in any sense than the prior distribution, and/or they forcibly diminish the welfare of some for the benefit of others. (A paradigm example was raising tax from the wealthy blind to fund the distribution of medical aid to those with minor but readily alleviated ailments.) We argued that such compromises with the two maxims of fairness<sup>35</sup> may nonetheless respect the principle of equal concern and respect, so long as the moral practice is a corollary of another that does not itself command differential treatment. The meta-principle embodying this idea was:

- (I') Humanity is the only morally relevant property of human beings. No moral law will command differential treatment of persons according to their significant characteristics, unless it is a corollary of a moral law that does not itself command differential treatment.

In this chapter I wish to test this principle against the partialist practices discussed in the last two chapters, those justified instrumentally and those justified as embodiments of our recognition of the value of relationships and of community.

In section 5A I review a range of discriminatory practices that are more or less objectionable. It turns out that the principle of non-discrimination (I') fails to mark

---

<sup>35</sup> See Chapter One, where we concluded that two maxims, though in tension, are needed to exhaust our intuitions:

- (f-') "Do not diminish the happiness of another, without their consent, for any end, except to make good damage freely done by them to the welfare of others."
- (f+) "Raise the welfare of the least advantaged."

a line between objectionable and unobjectionable discrimination. There is objectionable discrimination that accords with I'. The reason seems to be that we fear that we may undermine the self esteem of at least some groups even when discrimination is motivated by some further end.

Any attempt to tighten the principle, to outlaw all discrimination that can undermine self worth, will outlaw many practices that contribute greatly to other ends. Yet we have no reason to elevate the goal of demonstrating an equal valuation of all humans to the point where it is always and without limit allowed to compromise attainment of other goals. Rather it seems that what matters is **the extent** to which fairness is compromised, and to what extent practices undermine self esteem. These are important concerns, but not ones that trump other considerations. (There are no trumps.) Thus although the doctrine of equality of human worth underpins consideration of fairness and concern for self-esteem, it does not, as many suppose, provide an indefeasible constraint upon moral practices.

In section 5B, the source of the political philosophers' commitment to impartiality is debated. There are legitimate motives for emphasising impartiality: promoting autonomy, social harmony, discovery of moral truth, as well as fairness and self-esteem. These are implicit and explicit in modern political philosophy. But none justifies according impartiality the doctrinal status that it seems to have achieved. One argument that could in principle grant impartiality that status is the argument from scepticism. But moral scepticism fails obviously to differentiate between the right and the good in the way required to support a doctrine of impartiality. However, to demonstrate that the good has sufficient objective status to compromise the right requires an answer to the sceptical challenge. This is the topic of Chapter Six.

## **5A The Doctrines of Impartiality and of Equal Concern & Respect Compromised**

Eudaemonic compromises of fairness justified by I' can claim to be heeding the equality of human worth, not through limiting discrimination, but by granting to everyone equal concern and respect in the decision procedure which determines the allocation of goods, as the happiness or the preferences of each are accorded equal weight. They may not receive equal allowances of goods or opportunities (on any definition), but at least they are treated as equals. Rawls identifies this as the key notion of equality that should underpin a theory of justice (Rawls 1972, Section 77). Such a doctrine can be justified in part by the preservation of self-esteem that is its intended consequence (Section 67).

Yet policies consistent with non-discrimination will prejudice the interests of some individuals, and these may see themselves as victims, notwithstanding the fact that they have been "treated as equals". Take, for example, the blind taxpayer discussed in Chapter One. We concluded that the state would have to admit to being unfair in taxing the wealthy blind, but suggested that the state, in redistributing from the blind to those whom the state could help more efficiently, could at least claim impartiality: the blind man's potential for benefit is weighed equally with that of alternative recipients, and to that extent the decision mechanism is impartial.

Would the blind man think it ingenuous for the state to claim impartiality when pursuing an end (distribution proportional to ability to benefit), if its pursuit will predictably and clearly favour one group (the sighted) over another (the blind)? Is it a defence to claim that this favouring is merely incidental to an end that makes no reference either to the favoured or the ill-starred group?

This is a case in which I' does give us a defence against accusations of discrimination, but it is not a very cogent defence. It rather has the flavour of an insincere rationalisation for a discriminatory practice that we are not willing to relinquish.

There are other cases that also accord with 'I' but which so clearly breach our intuitive notion of the acceptable limits of discrimination that we resist them, notwithstanding the putative instrumentalist justification that we could use.

Most simply, objectionable incidental discrimination could arise from account being taken of the prejudices of the majority against minority groups. It may be possible to exclude such practices on other grounds: perhaps because in the long run welfare will be maximised if people are trained away from their prejudices - as discussed at the end of Chapter Two. But it is implausible that such calculations are the ground for our distaste for such discrimination.

Even without paying heed to prejudice, impartial welfare maximisation may still beget discrimination. Consider a police practice of stopping and searching people acting suspiciously. Suppose the police have weaker criteria for "suspicious behaviour" for black youths than for other groups: they are more readily searched. Such practices have been justified on the grounds that a higher proportion of blacks commit crimes. Were this true (and of course the statistics may be biased by the practice), it could be attributable to other factors driving criminality, like poverty, that happen to be correlated with skin colour. But as skin colour is easier to detect, there would be instrumental justification for discrimination. The discrimination would itself have a welfare cost borne by those wrongly arrested and might make the system less fair (f+), but this might be offset by the greater efficiency and accuracy of law enforcement if fewer searches using this policy yielded the same number of convictions.

What is it that is here objectionable beyond unfairness to some individuals - who are made worse off in the interests of the greater good?

The doctrine of impartiality ('I') is a defence against practices that undermine self-worth: one's self worth will be undermined by policies that discriminate against a group with which one identifies. A practice that picks out a group for discrimination upon the basis of a characteristic that was viewed as self- and group-defining, would appear to imply a lower valuation of that group by society.

Members of that group would be the subject of intentional rather than incidental discrimination.

However, the examples show that the scope of this defence is too narrow. It suggests that actions that are unfair, in that they allocate resources in ways that do not maximise fairness, may be considered impartial if they are justified by a non-discriminatory rule, one that does not pick out a particular group to be favoured or discriminated against. But our examples suggest that even discrimination that is justified instrumentally can undermine self respect of those discriminated against. In such cases, concern for self esteem leads us to eschew policies even if the damage to particular groups is clearly incidental.

But perhaps these groups are being over-sensitive. If the true goal of the police is to minimise crime, why should one racial group feel undermined simply because it happens to bear characteristics that are correlated with criminality? After all, change the example to gender, and few would complain: the greater criminality of men is well attested, and it would be surprising if the police did not tend to use different criteria for "suspicious behaviour" for men than for women.

Consider also the class of all those stopped for "acting suspiciously". A high proportion of those questioned will be without fault. They have been unfairly treated, discriminated against using some set of criteria. In our example this group will be larger if a colour blind stop and search policy is implemented.

Nonetheless the latter policy is intuitively the better one because it is non-discriminatory.

To appreciate the distinction between these cases, between significant and insignificant characteristics, takes us again back to the ground of our concern about discrimination: maintaining self esteem. Discrimination against a racial group, even where incidental, may undermine their self-esteem in two ways:

- the intentions behind a practice are not transparent: hence both the group itself and society more generally may view the expressed objective of the policy with scepticism; and indeed there may be no fact of the matter as to what is the true purpose of a policy
- particularly where a group has been subject to discrimination, or has low self-esteem for other reasons, failure to be sensitive to the impact of a practice upon that group may be seen as itself a symptom of the low value in which the group is held.

Hence, although the distinction between intentional and incidental discrimination may be an important one, it is clearly only an imperfect guide aiding us in determining the extent to which self esteem is compromised by a practice.

If 'I' is not strong enough to exclude objectionable practices of discrimination, perhaps it should be strengthened. **Perhaps any practice that predictably discriminates directly or indirectly against a collection of individuals that identifies themselves as a group, should be ruled out on grounds that it will weaken self-esteem; it will display a lower valuation of that group.**

What is the criterion for grouphood in this context? One might attempt to discover whether the characteristic upon which discrimination will be based is part of people's self definition and (perhaps a separate question) find out whether those with the characteristic treat each other as members of a group. If not, then discrimination on the basis of that characteristic would not contravene the doctrine of equal concern and respect.

This distinction might allow allocation of medical resources to those most able to benefit rather than to those most in need: that you have an illness for which no cost-effective treatment is available is not essential to your self-definition, nor do you identify with others in a similar predicament. Hence this partiality or discrimination seems less offensive than one based on race or colour. Similarly



as people "acting suspiciously" do not consider themselves a group, the police would be able to investigate them - so long as skin colour did not enter the list of criteria justifying suspicion.

But this new criterion is much too strict: there are many practices that discriminate against groups of people, and may weaken their self esteem to some degree, which we nonetheless pursue for the sake of some other end.

We do want the police to be able to search men more readily than women. Sufferers of some diseases come to define themselves as such; and form mutual support groups that underpin this self-definition. Yet we may find ourselves taking resources from such groups to give to others easier to treat. We do tax the blind. We fail to provide marginal relief to MS sufferers on grounds of expense.

A practice that may be justified initially will tend to undermine itself according to this criterion: persons discriminated against tend to band themselves together, and to begin to define themselves by their disadvantage. It is a commonplace of social history that discrimination forges a sense of group identity.

A housing policy that reduces security of tenure may be unfair to tenants under both fairness maxims. It may diminish the happiness of tenants who currently enjoy security of tenure, and these may well be amongst the least advantaged members of the community. This policy may nonetheless be justified instrumentally for maximisation of welfare, for it may increase property availability and labour mobility. However, being a tenant rather than a landowner may well in such circumstances become a badge of group identity, even if it does not start out as such. The tenants could define themselves as "victims of economic consequentialism", and then claim that the policy was discriminatory, and undermined their self esteem by demonstrating how little society cared for their sense of security. The victims of the market economic system may well see themselves as such. And indeed supporters of the market system are

compromising the wellbeing of its victims, and hence implicitly they are placing a low value on those individuals.

A welfare-maximising policy of compulsory car insurance may be unfair (f-) to the more careful members of a score of different groups who have to pay higher premia than would be determined by an omniscient actuary (who would set an individual's premium according to her own behaviour, not that of her actuarial group). These people will feel that they are discriminated against: as they are cautious drivers, they pose little risk to the insurance company, yet they have to pay more merely because they live in inner cities, or they are young, or are men, or have only just qualified, as the case may be. Many of these categories will be ones by which people define themselves, and are certainly ones that people would come to identify with if the discrimination became widespread. They may feel devalued by the practice of the insurance companies.

Practices fostering aesthetic ends are yet more likely to favour groups defining themselves as such by the very criteria used to discriminate. For example, arts funding may be determined by the pool of talent available in a country: a country with a singing tradition may build an opera house; who benefits will depend upon the accident of musical taste. Opera-lovers may very well identify themselves as such; and, once the resources involved became significant, non-opera lovers likewise. If our concern is the self esteem of different parts of the population, the fact that the policy is motivated by concern for art rather than for art-lovers may cut little ice with the *hoi poloi*.

More generally, many practices that reward those who can contribute most to a maximand involve discrimination on the basis of identity-defining characteristics. Individuals who can promote happiness or beauty, are likely to be able to do so largely because of their skill or virtuosity. Yet even Rawls' maximin principle will justify greater rewards for the skilled worker in order to coax forth her effort for the general good: the software house will not thrive without rewarding its programmers for their skills. These policies discriminate against the unskilled. The unskilled may not see themselves as a group so-called, but may well define

themselves as unskilled in the relevant sense individually: they see their worth as undervalued by a society that discriminates in favour of the skilled. The unfairness of favouring the naturally gifted is deep: it is fundamentally in tension with an egalitarian evaluation of humans; it is very likely to undermine self-esteem.

Furthermore we saw in Chapter One that even righteousness - virtue in the modern sense - is partly a question of how a person is (her constitutive luck) rather than of what she has chosen to do. In which case, there is no reason for treating it separately from other human excellences, from virtue in the ancient sense. Human value consists in a variety of characteristics - health, intelligence, correct motivation, physical beauty - some attainable by all in some degree, some by just a few. In all these cases we sometimes reward those who instantiate these virtues whether or not such rewards are really-deserved<sup>36</sup>, perhaps in order:

- to provide an incentive for their production and display;
- to give the virtuous the wherewithal to be virtuous;
- to entrench and manifest our positive evaluation of these virtues.

It is thus evident that in practice we are willing to discriminate even on the basis of characteristics that people regard as essential to their self-definition and even where the self esteem of those not favoured will inevitably be compromised.

---

<sup>36</sup> I.e. whether or not they have made a contribution through the correct exercise of their free-will, a contribution deserving reciprocation. The concept of real-desert was explained in Chapter One (1B and 1C). All differential treatment that is not really-deserved is unfair procedurally (f-').

**CATEGORISATION OF UNFAIR PRACTICES (PRACTICES THAT REDISTRIBUTE WELFARE WITHOUT REGARD TO FAIRNESS)**

	<b>OBJECTIONABLE PRACTICES</b>	<b>JUSTIFIED PRACTICES</b>
<b>Incidentally discriminatory practices that pick out groups defining themselves as groups.</b>	Welfare maximising stop-and-search policy that disproportionately stops a particular racial minority	Welfare maximising policies that discriminate against tenants; that compel car insurance.
<b>Incidentally discriminatory practices that do not.</b>	Welfare maximising "punishments" of the innocent to deter crime; random re-allocation of resources (e.g. a compulsory lottery).	Welfare maximising health policy that allocates health care to those who can most benefit from it rather than to those suffering most; rewards for the skilled to induce effort.
<b>Directly Discriminatory practices that pick out groups defining themselves as groups</b>	Mutual support amongst bandits; discrimination against racial groups that is not motivated instrumentally.	Discrimination against bandit groups. Mutual support within good groups and relationships.
<b>Directly discriminatory practices that do not.</b>	Discrimination against red-heads.	Awards to excellence that are inaccessible to those lacking natural talents.

Many of the examples above are of practices that have been contentious. For example, many have argued that the market system is fundamentally immoral, perhaps because it undermines the self esteem of those who fail to thrive in it. (Such an attack would be consistent with recognition that the market system

does produce more overall, and even that it produces more for the worst off than any alternative.)

Such positions are arguable, but they are not all overwhelmingly cogent. Given the range of the practices that incidentally degrade some members of society, it is clear that impartiality is defeasible. These practices compromise fairness and self-esteem to a greater or lesser degree, and they achieve other ends more or less. Compromise is necessary, with sensitivity to the circumstances of each case.

The conclusion of this discussion is that there is no doctrine of impartiality, there is only concern to support self esteem and to avoid practices that undermine it.

Once this is recognised, there is no reason to rule out even the intentional direct discrimination that was justified in the discussion in Chapter Four by the objective of sustaining relationships and communities. For positive discrimination of this sort need not be particularly damaging to anyone's self esteem: and of course it can be sustaining of the self esteem of the members favoured. And ideally everyone would be a member of a flourishing and mutually supportive group.

The doctrine of the equality of human worth does survive, but only as a conceptual support to two defeasible objectives: to maximise fairness, and to minimise damage to individuals' self esteem. Self esteem will generally be damaged by unfair practices, which may roughly be ranked in terms of their impact upon self esteem as follows:

- discrimination that identifies a self-defining group for ill-treatment for no end other than to eradicate or harm it; this is deeply offensive, but may be justified if the group itself is the bearer of a culture that is wicked (e.g. a group that defines itself by its racial bigotry)

- discrimination that identifies a self defining group for ill-treatment for another end; (e.g. the racial policing case discussed above); this is very offensive and could be justified only by overwhelming general benefit
- discrimination in favour of a group without ulterior motive (e.g. positive discrimination); this incidentally discriminates against others, but may not be thought offensive depending upon circumstances
- discrimination in favour of a group for some end; (e.g. the favouring of the skilled in a market system); this incidentally discriminates against others and is probably acceptable if the gains are significant.

From the perspective of self-esteem, it also matters who is doing the discriminating. Perhaps it may matter more when the state discriminates; but this is not an absolute rule. It also matters less if a group discriminates against you in favour of its own members (flourishing behaviour) than if it discriminates against you in favour of a third party. For example I will mind less that you favour your daughter over mine than I would if you favoured Esmeralda's daughter over mine.

This applies to a state too: it may discriminate in favour of its own without causing offence to other nations, so long as it is equally prejudiced against all other nations. This is the basis of international trade law: tariffs are accepted, but discrimination against particular nations in the form of particular surcharges or quotas or indeed trade sanctions are seen as an instrument of aggression.

Difficulties arise where some individuals or groups are treated as part of the outside world yet see themselves as members of the nation; their self-esteem may certainly be vulnerable in these circumstances. Barry (1995, p165 note c) for example considers the "advantages of establishment enjoyed by the Church

of England or by the Lutheran Church in Sweden". He comments that these advantages

"are scarcely on a scale to lead anyone to feel seriously discriminated against. In contrast, denying the vote to Roman Catholics or requiring subscription to the Church of England as a condition of entry to Oxford or Cambridge did constitute a serious source of grievance. Strict adherence to justice as impartiality would, no doubt, be incompatible with the existence of an established church at all. But departures from it are venial so long as nobody is put at a significant disadvantage .... on the basis of his religious beliefs."

Barry's treatment of this issue illustrates the problem with treating fairness in a doctrinal matter: breaching impartiality is seen as illegitimate, although minor infringements will be tolerated. The sense of tension that results in the compromise represented by the slight advantages enjoyed by the Church of England is entirely absent from the analysis. On the understanding of fairness and of the nature of our support for groups that I have been developing, on the other hand, the tension is easy to understand:

- the Church of England represents the religious expression of English culture; as English culture and the flourishing of the English community that bears that culture is considered a good, institutions that support that culture are legitimate
- however, there is a wider more inclusive multi-cultural understanding of England that includes members of all religions within its ambit; even those who do not recognise this broader community as of value must nonetheless acknowledge that discrimination (even merely symbolic kinds of discrimination that Barry would tolerate) in favour of the established Church may undermine the self-esteem of non-CofE people who consider themselves English.

Pace Barry, this issue cannot be settled merely by reference to the materiality of the injury involved in the discrimination: the value of the group being favoured (is an England that is defined in part by religion a community worth sustaining?) and the threat to self esteem and to wider communities (how hurtful is the discrimination to how many people?) must be assessed in the particular circumstances. Of course the members of the wider groups may themselves favour a discrimination that is discriminatory to yet other individuals. It is paradoxical but true that in order to avoid objectionable discrimination (discrimination that unwarrantedly diminishes self-esteem) it is necessary to be sensitive to others' favouritism towards their own groups.

These are all empirical questions. This is the point. What is important is to be guided not by a doctrine of impartiality, but rather by particular judgments balancing the impact on self-esteem (perhaps the impact that might reasonably be expected), against the benefits secured by discrimination.

Issues of membership, and the legitimacy of different larger groups' self conceptions, are discussed further in Chapter Seven.

That impartiality is seen as a mere embodiment of a defeasible commitment to consider fairness and self-esteem amongst other guides to action, implies that (P') is correct, and (I') is mistaken:

(P') People are often right or even obliged to be loyal to or to favour other members of their group, whether family, community or state, even when such discrimination neither serves an end nor fulfils an obligation that can be stated without reference to the group members' identity, or to their particular significant characteristics.

Impartiality, interpreted as an injunction to sustain self esteem, turns out **not** to be a meta-ethical principle, never to be compromised; rather it is a theorem of no greater status than, for instance, the injunction to maximise happiness - they can both equally be compromised where other maximands, or deontological



injunctions, are more cogent (there are few absolutes in a world of conflicting ends and injunctions).

Fairness nevertheless does remain a maximand, and the impact upon self-esteem remains of moral importance; both are founded upon the doctrine of equality of human worth. But, it is apparent that we are not overly taxed by these considerations. We are willing to harm or benefit people, even if it will undermine their self worth or that of their peers, if other ends justify this policy.

Although people are of equal worth in themselves, they may be of variable worth instrumentally, as means to various objectives rather than ends in themselves (more or less valuable doctors, policemen, partners in sustaining relationships and groups). We will support members of our community because their way of life seems to us good and valuable in itself; they may embody virtues that we consider particularly noble; and because supporting members of a group constitutes a contribution to the flourishing of the group. Do members of good communities **really-deserve** support because they are lucky enough to embody the correct morality? No! But although it may not be fair, it may nevertheless be right, justified and perhaps obligatory to give them support.

There are, in sum, many occasions in which we discriminate between individuals on the basis of valuations that are deeply undermining of their self esteem. Such discrimination is by no means limited to partial practices; indeed exceptions to the principle of equal treatment (D in Section 1A) are legion. Hence, partial practices are in good company amongst actions justified by differential valuations of people.

Nevertheless we do recognise that such practices are unfair. And it is this very recognition of unfairness and of the relevance of questions of self esteem that is our recognition of the equality of humans as moral agents. People really-deserve to be treated equally, and with manifest equal respect, but achievement of ends other than fairness, and other moral restraints, often oblige us to treat people unfairly and partially.

The best theory to explain the way in which we justify some unfair practices and find others objectionable builds not upon an infeasible doctrine of impartiality but an appreciation of the tensions caused by competing maximands of fairness, utility, beauty, self-esteem, relationships and community, et cetera.

## **5B The Basis for the Doctrine of Equal Concern in Political Philosophy**

If the principle of the equal worth of persons implies no more than that fairness and sensitivity to self-esteem should be amongst the many considerations guiding action, why does it figure so prominently in political philosophy?

The Kymlycka thesis quoted in Chapter One, that "if a theory claimed that some people were not entitled to equal consideration from the government, if it claimed that certain kinds of people just do not matter as much as others, then most people in the modern world would reject that theory immediately" (Kymlicka (1990), pp4-5), which echoes Dworkin and Rawls, may now seem less plausible.

**Although we may agree that regarding the maximand of fairness, or regarding individuals as moral agents, ends rather than means<sup>37</sup>, each matters equally, yet we must also recognise that some people do matter more than others in that they can contribute more to the achievement of certain moral ends.**

The liberal tradition may claim that the differing instrumental value of human beings should not be recognised by political institutions, and hence that **political** philosophy should attend only to the equal valuation of men and women as moral

---

<sup>37</sup> Valuation of people as means will tend to lead to differential valuation: some are better than others at contributing to various objectives. Conversely, as ends in themselves, it is possible to conceive of people as of equal worth (assuming, as suggested in Chapter One, that it is actions rather than people who, in general, are wicked or good).

agents and hence as ends, as if that fact must dominate all other moral truths and constrain the shape of our political institutions. Yet it is not credible to claim as Kymlycka does, that all plausible political philosophies must share this narrow focus: in particular such a stance would too easily dismiss utilitarianism, which puts a value on individuals as means to enhanced welfare as well as as ends, implicitly a differentially favourable valuation of the talented. (It turns out to be difficult to sustain Kymlycka's claim that utilitarianism can be construed as an egalitarian philosophy, in any practical sense<sup>38</sup>.)

The question begged by Kymlycka, and before him by Dworkin and Rawls is Why?: why should other considerations apart from fairness be ignored when framing the political regime?

To make this question more precise, consider one mechanism for embodying the equality of human worth, and the practical consequences of according that mechanism priority. The liberal tradition insists that equality requires the maximisation of the freedom that each enjoys to pursue the good as she sees it. Rawls codifies this as the priority of the right over the good (Rawls 1972, Section 6). Discrimination against individuals' conceptions of the good (apart from those that conflict with the impartialist doctrine itself) is unfair.

Defining liberty without assuming a value system is not easy: different concepts of the good require different forms of liberty and different resources for their pursuit. Rawls resorts to a list of primary liberties or goods (Rawls 1996 p.181)

---

<sup>38</sup> Kymlicka (1989) is at pains to establish that at least one version of utilitarianism can be understood as spelling out the idea that "from the moral point of view people's interests matter equally." (p.26) This deontological version "defines the right in terms of treating people as equals, which lead to the utilitarian counting procedure, which happens to maximise the good."(p.28) My point is that notwithstanding this theoretical grounding, the inevitable practical consequence of the utilitarian counting system is to accord greater value to those who, if appropriately rewarded, will be able to contribute more to the satisfaction of desires. It is therefore disingenuous to pretend that utilitarianism is egalitarian.

to define that without which someone's liberty is infringed; but Rawls acknowledges that the liberal principles of Justice as Fairness, like any political conception, "inevitably encourage some ways of life and discourage others, or even exclude them altogether."(p.195)

How should a liberal answer the charge that a liberal constitution is unfair to those who hold incompatible conceptions of the good? Why should concern for others' liberty prevent our view of the good from informing our judgement in creating a political system? Whence this dominant role for fairness and equal concern and respect as paramount desiderata behind the veil of ignorance such that they should have priority over our view of the good? Rawls argues that the principle of tolerance that is at the heart of Justice as Fairness can be cogently advocated within most of the competing comprehensive moral theories. But we are situated within our moral framework, and the extent to which we are tolerant must be dictated by our "comprehensive" conception of the right and the good. Rawls must convince us in our own conception that the good should be sacrificed to the right.

Three contrasting liberal impartialist responses are possible, depending upon whether the liberal is value-based or procedural or sceptical with respect to conceptions of the good. Rawls himself provides arguments for all three positions, though I suggest that he relies most heavily on an implicit scepticism.

### **i Value-based Defences of Liberal Impartiality**

Value-based liberals adopt liberal principles for the sake of the values that liberalism embodies, most importantly the values of autonomy, of a society embodying mutual respect and social harmony, and of an open search for truth.

Rawls hopes to establish his political theory ("Justice as Fairness") as independent of "any particular comprehensive religious, philosophical or moral doctrine; ...". Nevertheless Rawls is committed to certain values, and it

sometimes appears that these values underpin the priority attached to the particular liberal constitution that he endorses.

And of course there are fundamental value-based arguments that will support a Rawlsian conception of political justice against many other considerations:

- Rawls' conception embodies "a fair system of cooperation between free and equal persons"; which is a mechanism for maximising fairness ("the good of justice and the social bases of ... mutual self-respect" (1996, p.203)).
- Rawls' conception provides "a workable conception of political justice", in which we avoid arbitrating between contending philosophical views, "so that social cooperation on the basis of mutual respect can be maintained". (Rawls, 1985) The priority of right over good and the rest of his theory of justice may thus be a formula for maintaining social harmony - a primitive value.
- "[E]stablishing and successfully conducting reasonably just .... democratic institutions over a long period of time... is a great social good". (1996, p.204) This is (I believe) the good of community discussed in Chapter Four.
- Given that there are large areas of uncertainty with respect to the best thing to do, tolerance is dictated for lack of conviction that one course is better than another. This argument is parallel to the argument from epistemic uncertainty that justified tolerance of partiality in the Sections 2Aiv and 3Cii; we are now suggesting that it should be used at the political level as a support for a measure of impartiality, non-discrimination against others with different moralities. This argument justifies some degree of tolerance even without endorsing the radical and extensive

uncertainty regarding the good that Rawls calls the "burdens of judgment" (1996, p54), of which more below.

- A more hopeful argument for tolerance (based upon a less radical uncertainty) is found in Mill: "it is important to give the freest scope possible to uncustomary things, in order that it may in time appear which of these are fit to be converted into customs." (Mill, 1859, p.125). Tolerance also fosters diversity allowing instantiation of a plurality of values.
- Raz justifies liberal tolerance in order to underpin moral autonomy defined as the availability of choice between goods: however as we discussed in Section 2Bii, the autonomy yielded in such circumstances seems an odd value to promote; but it is in any case not one that requires a very extensive freedom.
- Where we are confident of the rightness of a course, autonomy is choice as to whether to follow the right rather than the wrong course. This is also a value, and a profound one. It accords individuals the right and the means to take moral responsibility for their own lives.

These are all important desiderata of political philosophy: they must be considered and weighed against competing values when determining action. Should they be given preeminence over other ends?

As we have noted, Justice as Fairness and other liberal doctrines emphasise these particular values by giving priority to the Right. Technically this is accomplished for Rawls by excluding particular conceptions of the good from the original position, so that the means and the wherewithal to pursue as wide as possible a range of conceptions becomes by default the aim of each of the negotiators. (The original position is a procedural device. I include it here rather than in the section below discussing procedural liberals because the original

position is a theoretical construct rather than a practical procedure, and hence must be interpreted as articulating a substantive moral view.)

Habermas (1995) suggests that there is a more direct route to the priority of the right over the good. He characterises the distinction as one between "norms" which "inform decisions as to what one ought to do" and "values" which "inform decisions as to what is most valuable". Amongst other distinctions he claims that

"The obligatory force of norms has the absolute meaning of an unconditional and universal duty: what one ought to do is what is equally good for all (that is for all addressees).... different norms must not contradict each other .... Different values, by contrast, compete for priority ..., they constitute shifting configurations fraught with tension."(p.115)

Why should norms be absolute and unconditional whilst values compete? I have been suggesting that no moral desiderata are indefeasible, whilst all are (or can in some circumstances be) unconditional in the sense that they remain considerations in the practical calculus even where they do not determine action. Habermas suggests that such "a leveling of the deontological dimension" whilst it might be appropriate for a utilitarian or Aristotelian ethic, ill-accords with a "theory of rights" which "proceeds from the concept of autonomy". (p.114<sup>39</sup>)

Hence it is apparent that it is the concept of autonomy that is doing the work, and that that value is allowed to dominate others - unjustifiably.

Most of these justifications for tolerance apply particularly at the political level, and go some way to explaining the reticence that we expect from political institutions with respect to personal morality; yet there are countervailing values specific to the political level that militate against tolerance - the values of

---

<sup>39</sup> Habermas says that Rawls himself levels the norms and values in his articulation of the original position by limiting the contracting parties to a first person perspective. Habermas claims that Rawls rectifies this by "acknowledging as primary goods only those which are expedient for the life plans and the development of the moral faculties of citizens as free and equal persons." (loc.cit.)

coherence and community for example. In any case, even at the political level, these justifications are all defeasible - maximands are identified that are advanced by tolerance and by the allocation of resources to support individuality, but priority is not established.

All the justifications are at best proofs that some one end will be furthered by tolerance. Such support for tolerance and fairness will by no means entail a ban on partiality, where partial practices will better promote the good. Yet few moral or political philosophers explicitly undertake to show that the benefits of partiality are systematically trumped by the benefits of impartiality; Rawls for example effectively avoids this debate by means of the veil of ignorance.

Barry (1995) does explicitly subordinate other ends to that of achieving a single good - agreement - but without providing any argument for giving this objective such overriding importance:

"...I shall offer a general argument designed to give a reason for voluntarily constraining the pursuit of the good within the limits set by justice as impartiality... my argument presupposes the existence of a certain desire: the desire to live in a society whose members all freely accept its rules of justice and its major institutions." (p.164)

Barry builds his argument for neutrality between different conceptions of the good upon his claim (following Scanlon) that the desire for agreement is "quite strong in most people". Noting that disagreement between conceptions of the good is widespread (an empirical point we will discuss and qualify in Chapter Six), he concludes that only a neutral state can expect to secure agreement.

Yet it is perfectly consistent for someone to share Barry's desire to live in a society whose members all freely accept its rules, and yet to think that that objective must be compromised if it proves incompatible with efforts to design a society that promotes the good. Tolerance would help to secure the first end



but may have to be circumscribed to secure the second. (And of course a person with such convictions will not "freely accept" the rules and institutions of a neutral state!) Barry supplements the appeal to the desire for agreement with an appeal to scepticism (see below); but fails to notice that agreement will only accrue to a neutralist state on the implausible assumption that everyone reasonably shares his scepticism about the good.

The communitarian view is precisely that freedom should not be allowed to work in ways that undermine the communal aspect of social life; i.e. liberalism can and should be compromised in pursuit of communal harmony. Clearly, partialist and patriotic practices in support of any other ends will fit within the same model. The cost of ascribing priority to "justice as fairness" is too great in terms of the forgone opportunities to pursue other ends.

Even impartiality and maintenance of self-respect, which we have seen to be one of the motivators of impartiality, can conflict, and lead to hard choices; all the more so for other values. To illustrate, consider an example that Barry (1995) uses to make the case of the impartialist against proponents of an "ethic of care":

"Let us imagine that [someone concerned to be caring] is on the board of governors of the school that the daughter attends and is proposing to exploit the influence that this gives her in order to gain for her daughter some unjustified advantage - for example, special musical tuition for which some other child is much better qualified. She would only have to ask herself how she would feel if she were the mother of the other child to see how morally repugnant her plan to abuse her position is." (p.254)

Barry's description of the case is loaded. Suppose the parent governor's commitment to impartiality (after reading Barry) becomes so strong that she even forbears from securing an interview for her child for the music scholarship. Generally such preferential treatment is seen as a perk of the job of governor, in which case the daughter's own self-esteem might be undermined if the mother

did not so use her influence. (The other mother's sense of maternal obligations may also be offended by such coldly impartial behaviour.) Of course, one might argue that it would be better if the more rigorous impartial ethic became more widespread: but this would come at a cost in opportunities to display affection and care. There is a tension between ends and values that the impartialist political system ignores.

## **ii Procedural Defences of Liberal Impartiality**

The second response of the liberal to the complaint that the liberal political system is prejudicial to the pursuit of certain visions of the good life, particularly visions involving partiality to the bearers of a valued culture, is procedural. It is claimed that a fair procedure for determining the political structure yields a liberal constitution.

This view too is found within Rawls in the idea of an overlapping consensus, if this is taken not as a sign of the achievement of the social value of harmony, but rather as an epistemic constraint on a political system: the procedure for determining the correct political system is to find one to which as many as possible of rival moral conceptions can subscribe. Rawls wants his political conception to be "formulated ... in terms of certain fundamental intuitive ideas viewed as latent in the public political culture of a democratic society." (Rawls 1988). The political conception should be "the focus of an overlapping consensus" ... and "should not presuppose any particular fully (or partially) comprehensive doctrine."

For Rawls the identification of the common core across different conceptions of the good raises its achievement to a point of lexical priority within each different conception of the good, such that the political institutions that delivered the common core would *ipso facto* demand our support. However, although delivery of "the consensus set of goods and liberties" may be thought a good within each conception, Rawls fails to argue that it will be an overriding end in each of the

conceptions of the good across which the consensus prevails, or indeed from some overarching perspective. For each conception, that bundle of goods and liberties has only the value ascribed to it within its comprehensive moral system. Rawls' political conception is only as valuable for any group in the community as its place within their comprehensive philosophy allows.

Habermas more explicitly builds an epistemic procedure into the foundations of his comprehensive ethical theory:

"Discourse ethics ... views the moral point of view as embodied in an intersubjective practice of argumentation which enjoins those involved to an idealizing enlargement of their interpretive perspectives." (p.117)

The question is what would result from such a procedure? It is not obvious that were it conducted without Habermas' prejudice in favour of determining absolute norms, that it would have the liberal consequences that Habermas expects.

I will discuss in Chapter Six an epistemic role for the consensus in discovering ethical truth. My claim will be that the core of moral consensus attaches to particular judgments about particular situations, and that in such cases consensus is universal. Rawls can merely appeal to a pragmatic attachment to tolerance amongst a number of competing moral ideologies within the Western world. This has epistemic force, but the force is only commensurate with the value accorded to tolerance within each ideology. Of course, were all other moral views undermined by such cross-cultural consideration, then the remaining core might have priority by default: this sceptical conclusion is considered below.

A procedural approach similar to that of Habermas is taken by O'Neill (1996), another modern Kantian: she claims that reasons for action *must* be held *capable* of being followed or adopted by others."<sup>40</sup> When this principle is

---

<sup>40</sup> The O'Neill procedure has a strong echo of the call within Barry (1996) for political systems to maximise agreement. But in the case of O'Neill agreement is seen as an epistemic value rather than as a moral

pressed into effective service it turns into a requirement to use arguments about which there is a consensus - which is in O'Neill to be understood as a formal constraint on a system, one that is supposed to yield a liberal conclusion. A narrow interpretation of "capable of being followed", one that demands that reasons are compatible with all competing moral visions, would yield the liberal conclusion that those visions of the good that are incompatible with a liberal regime, e.g. those that are intolerant, are *ipso facto* unreasonable and beyond the pale. But the moral epistemic argument that is needed to support O'Neill's view will not support a narrow interpretation, for epistemology must allow for the possibility of error. As we will discuss in Chapter Six, there is no shortage of cogent error theories to explain how it is that competing moral visions are mistaken. It is not obviously unreasonable to hold a vision of the good that is not capable of being followed by others, if you have reason to believe that those others are in the grip of some error.

### **iii Sceptical Defences of Liberal Impartiality**

The most plausible interpretation of Rawls is that he prioritises the right because he views the rival conceptions of the good with implicit scepticism. Scepticism is implicit in his account of "the reasonable": and reasonableness is a criterion applied repeatedly in the development of Justice as Fairness. Only reasonable views are allowed to influence the development of the Fair system of Justice.

There are, according to Rawls, two aspects to the reasonable. The first "is the willingness to propose fair terms of cooperation and to abide by them provided others do. The second ... is the willingness to recognise the burdens of judgements and to accept their consequences ..." (1996, p.54). It is in the explication of the "burdens of judgment" that Rawls builds his scepticism. Rawls has already concluded that "reasonable pluralism" concerning conceptions of the

---

value, hence I reckon her amongst the procedural liberals.

good will not be diminished in the way that disagreements in science are expected to diminish in the long run. This conclusion is a sceptical one (as we will discuss in Chapter Six, section G). It provides a response to those disenfranchised by the political settlement because their conception of the good is not consistent with liberal tolerance of other views.

The sceptical conclusion about conflicting conceptions of the good is supported by the fact that the evidence needed to reach moral judgments is (a) conflicting and complex, and (b) difficult to weight, and that theorising about the evidence (c) uses vague concepts, and (d) is susceptible to cultural bias. A fifth and sixth reason refers to the pluralism of values, which make overall assessments difficult, and render a compromise possibly excluding some values, necessary. (Rawls, 1996, pp.56-7).

These are all of course sound points, and should cause any policy maker or moralist to hesitate before pontificating. But they are not valid arguments at a meta-ethical level. They are all quite compatible with a cogent case being made for the abrogation of a particular liberal right at a particular time, for the sake of some particular aspect of someone's vision of the good. Through his reference to the "burdens of judgment", Rawls seeks to silence such an advocate by disparaging her as unreasonable.

Rawls' attitude is not explicitly sceptical: he wishes his merely "political" liberalism to be compatible with cognitivists amongst those holding comprehensive theories of the good. However, I suggest that in this objective he fails: his view is political and not comprehensive only because it is implicitly sceptical, and for that reason it could not be endorsed by any comprehensive conception of the good. Indeed I suggest that this is the explanation that he seeks when he comments

"It is a great puzzle to me why political liberalism was not worked out much earlier: it seems such a natural way to present the idea of liberalism, given the fact of reasonable pluralism in political life." (1996, p.374 fn1)

The "fact of reasonable pluralism" is for Rawls (as for Barry) a starting point. He claims that this is compatible with objectivism, and indeed that it may be a matter of regret that some comprehensive theories fail to thrive in a liberal society (pp197-8). In this he aligns himself with Isaiah Berlin:

"For Berlin the realm of values is objective, but values clash and the full range of value is too extensive to fit into any one social world; not only are they incompatible with one another, imposing conflicting requirements on institutions; but there exists no family of workable institutions that can allow sufficient space for them all."

However, Rawls claims nonetheless that a "just liberal society may have far more space than other social worlds". (1996, p197 fn.32).

The question is why is "space" the appropriate criterion for choosing a political system, given a conflict in values. A pluralist about values may nonetheless say that regrettably (for example) the value of autonomy must be sacrificed to the value of community (at least in this community at this time).

Rawls regrets the narrowing of the space "of reasonable pluralism". Pluralism starts as a fact, but appears to become a value in itself, driving the whole political-philosophical enterprise: producing a society that will foster such pluralism by preventing any conclusions from being embodied at the institutional level. This is Rawls' neutrality, and it is unjustified.

Rawls or Barry might argue however that fairness itself demands not only allocation of goods and liberties in accordance with real-deserts, in so far as this can be achieved without undue compromise of other ends or injunctions, but also equal respect to each man's vision of the good. They may claim that there is no conception of the good that can trump fairness, for all conceptions are equalised simply in view of individuals' fundamental equality as moral agents.

Such an argument may be cogent on a non-cognitivist analysis of conceptions of the good - that is if one denies that there are right and wrong conceptions of the

good; but non-cognitivism leaves the political philosopher with no tools with which to argue. For cognitivists there is no reason to accept that individuals' equality implies equal status to the moral conceptions which they espouse.

A cognitivist may nonetheless take a sceptical stance particular towards conceptions of the good: Barry (1995) for example claims "that no conception of the good can justifiably be held with a degree of certainty that warrants its imposition on those who reject it" (p169). But what is the argument for this conclusion that would not equally undermine claims to the objectivity of the Right?

Underlying Rawls' position is a "political conception" of "citizens as free and equal persons". Persons are free and equal in virtue of their possession of "two moral powers connected with the elements in the idea of social cooperation ... namely, a capacity for a sense of justice and a capacity for a conception of the good. A sense of justice is the capacity to understand, to apply, and to act from the public conception of justice which characterizes the fair terms of social cooperation. The capacity for a conception of the good is the capacity to form, to revise, and rationally to pursue a conception of one's rational advantage or good ... this good must not be understood narrowly but rather as a conception of what is valuable in human life." (Rawls, 1985)

Contrast the definite article in "the public conception of justice" with the indefinite one in "a conception of what is valuable in human life." The implication is that any conception of what is valuable in human life will do, but only *the* conception of justice is adequate.

How is it possible on the one hand to maintain this indifference to the content of individual conceptions of the good, and yet to claim it is precisely these conceptions of the good that provide life with its point - and indeed that it is precisely in order to allow the pursuit of these conceptions that justice as fairness is promoted?

"It is essential to stress that citizens in their personal affairs, or in the internal life of associations to which they belong, may regard their final ends and attachments in a way very different from the way the political conception involves. Citizens may have and normally do have at any given times affections, devotions and loyalties that they believe they would not and indeed could and should not stand apart from and objectively evaluate from the point of view of their purely rational good."  
(Rawls 1985,p241)

Rawls is attributing to citizens the belief that they would not wish to question their loyalties even were they irrational: this is an extraordinary piece of (surely unwitting) condescension. Harking back to a theme of Chapter Two, we must emphasise that each citizen views her own conception of the good as cogent not just for herself but without qualification; and few would allow that their commitment would not withstand "objective evaluation". The correct conception of the good - the conception that we all strive to attain, would justify compromises with other ends, including fairness.

The problem here is thus much deeper than that the political indifference between goods urged by Rawls is contestable. Prioritization of the right, indifference between different conceptions of the good, is also inconsistent with the conception of the purpose of human existence that underpins the Rawlsian specification of fairness. Rawls summarises his view of the matter in a phrase: "justice draws the limit, the good shows the point." (Rawls 1988 p252). If the good shows the point, then surely the nature of the good should at least influence the shape of political institutions.

#### iv Conclusion

Of the three liberal responses, the only credible arguments for impartiality against the partialists are ones that argue from first level values, i.e. value-based liberalism. However, the equality of human worth is not the only moral truth, nor



are fairness, autonomy and agreement the only moral values within our conception (which is of course, as far as we can establish it, the correct conception). Happiness and beauty and flourishing good groups and productive relationships are also of value. Achievement of these other values may be compromised too severely if we design political institutions that ignore them.

This is not to argue against the tolerance that Rawls advocates, but it is to claim that it is merely a means to the achievement of other goods (including fairness) rather than a constraint that must apply to any system.

Conversely if our conception of the good is correct and should be promoted, what says that we cannot discriminate in order to promote it?

## **6 THE SCEPTICAL FOUNDATIONS OF IMPARTIALISM**

### **UNDERMINED**

***ISAIAH 5:20) "who call good bad and bad good, they make the light into darkness"***

Many contemporary moral philosophers take their cue from recognition of the plurality of visions of the good. To challenge impartialism is to challenge the validity of the arguments deriving impartialism from the fact of pluralism.

Neutralist liberals, wishing to establish political theories that achieve social harmony and underpin personal autonomy yet are robust across different conceptions of the good, have reacted to pluralism by eschewing what Rawls calls comprehensive theories, aiming rather to create the minimum just framework that allows to each the freedom to develop their life according to their own theory of the good. Barry (1995) convincingly argues that it is not merely a recognition of pluralism but consequent "scepticism [that] supplies the premise that is needed to get from the desire for agreement on reasonable terms to the conclusion that no conception of the good should be built into the constitution of the principles of justice." (p.172) Only by taking a sceptical view of other goods are neutralist liberals able to ascribe to personal autonomy and social harmony the preeminent position needed to justify impartiality.

Communitarians and virtue theorists have paradoxically found value in the very diversity of visions of the good: that each live according to the value system of the community with whom they identify is a form of Aristotelian flourishing. Yet implicit in such openmindedness is a similar scepticism regarding the possibility of establishing an objective metric that might discriminate against some communities' or individuals' projects, or that might arbitrate between conceptions that are evidently inconsistent with each other.

Neither of these positions is sound. The liberal neutralists beg the question why the core liberal values of autonomy and social harmony should be immune from scepticism. The communitarians fail to explain how value can be ascribed to local practices even by the practitioners given present awareness of diversity, an awareness (that can scarcely be wished away) that the practices and moral theories of different individuals and groups are in conflict with each other.

My claim is that a more robust response to apparent pluralism is appropriate. Core<sup>41</sup> judgments regarding both the ends that motivate us and the restrictions that constrain us are equally objective.

I seek to show how it is possible to avoid moral scepticism even whilst recognising the plurality of competing visions of the good. A methodology is sketched for evaluating the objectivity of particular moral and descriptive judgments. From this base, a defence of the particular partialist judgments discussed in chapters four and five can be mounted against sceptical attack.

This discussion serves not only to address the challenge presented by the implicit scepticism of much current political philosophy, but also as a foundation for arguments launched in Chapter Seven, arguments that build upon an objectivist moral base rooted epistemically in consensus to demonstrate how individuals can have obligations even to large groups to whose flourishing they appear to contribute little, and to explain the nature of the value that inheres in good groups.

The argument of this chapter is distributed across eight subsections, summarised as follows:

A. Meaning-theoretic Considerations relating to the Attribution of Objectivity to a Domain of Discourse. Theory of meaning since Wittgenstein and Quine establishes that there are no *a priori* reasons for denying objectivity to

---

<sup>41</sup> A qualification to be explained below.

moral judgments. This conclusion is consistent with Dummett's meaning theoretic treatment of other realism/anti-realism disputes (the view that objectivity in a domain of discourse is equivalent to bivalence-without-reinterpretation). A criterion is nonetheless needed for determining *a posteriori* the objectivity of discourse.

B. A Public Language Test of Objectivity. A continuing degree of consensus in the use of predicates "in normal circumstances", and agreement when judgments must be adjusted, is the criterion for ascription of bivalence and objectivity to "core" judgments; (it is this consensus that is unavailable for a private language). Consensus regarding the extension of a predicate is an indication that it has a determinate projectible sense. (Meaning = use.)

C. Objectivity in Moral Discourse #1: Focus on the Particular. The continuing consensus requisite to sustain the claim of objectivity in ethical discourse pertains to singular judgments (observation statements); arguments for moral subjectivism or moral relativism fail to notice the extent of consensus about such judgments, focusing irrelevantly on the diversity of more general claims.

D. Objectivity in Moral Discourse #2: a Research Programme for Testing Error Theories and for Establishing which Moral and Descriptive Judgments Warrant attribution of Objectivity. There are of course many disagreements even about singular moral judgments. However, these can be explained: as the product of conflict between equally valuable goods, as the consequence of the vagueness of terms, and by plausible error theories (including lack of expertise). It is plausible that an empirical research programme, sensitive to these explanations, would justify objective status for core uses of moral terms.

E. Objectivity in Moral Discourse #3: how Charity in Interpretation should Limit attribution of Immoral Judgments to Other Cultures. A Davidsonian

principle of charity should limit transcultural moral conflict about particular judgments.

F. Moral Objectivity is not Reducible to Descriptive Content. The cogency of moral judgments (their action guiding character) is also objective: moral terms and judgments are not obviously reducible to descriptive terms; and the onus of empirical proof is therefore upon the non-cognitivists and evolutionists to explain why this aspect of the phenomenology of moral judgment should not be taken at face value.

G. The Possibility of Convergence of Moral Theories. Upon this cognitivist base it is conceivable (*pace* Williams) that even moral theories will eventually converge - providing confirmation that moral theorising is "world-guided".

H. Implications for Partialism. That we can make objective moral judgements is critical to our ability to offer a critique of the practices of our own group as well as those with a different cultural heritage. Such a critique allows us to determine the extent to which partiality to support particular groups can be justified. The possibility of objectivity undermines the general argument to impartiality founded upon pluralism: the fact of moral pluralism at the level of theory is consistent with judgments that particular groups and relationships have objective value sufficient to justify loyalty.

## **6A Meaning-theoretic Considerations relating to the Attribution of Objectivity to a Domain of Discourse.**

The notion that statements of fact have some privileged metaphysical position relative to statements of value has been dealt fatal blows by developments in the theory of meaning over the last half century.

At the beginning of that period, when a correspondence theory of truth implicitly or explicitly held sway<sup>42</sup>, and, further, analyticity was seen as a guarantee of truth, the contrast between the descriptive and the analytic on the one hand and the evaluative on the other appeared dramatic. Descriptions of the world were anchored by analytic meaning statements that determined their correspondence to the observed world. Scientific theory was in turn founded upon these descriptive statements<sup>43</sup>. Mathematical truth was anchored through deducibility from a set of self-evident tautologies<sup>44</sup>. All these statements had decision procedures determining their truth. What comparable anchor was available for statements of value? Hence alternative explanations of moral discourse, shorn of claims to objectivity, were born.

The first blow to this view of the fast connexion between descriptive, analytic and logical truths and the facts of maths and science was struck by Gödel (1931), who finally undermined the project to found mathematics on the self-evident axioms of logic.

---

<sup>42</sup> This view of the world is neatly exemplified in Wittgenstein's *Tractatus Logico-Philosophicus*:  
"2.1 We picture facts to ourselves.  
"2.12 A picture is a model of reality.  
"2.1514 The pictorial relationship consists of the correlations of the picture's elements with things.  
"2.1515 These correlations are, as it were, the feelers of the picture's elements, with which the picture touches reality.  
"3.01 The totality of true thoughts is a picture of the world."  
(Wittgenstein 1921)

<sup>43</sup> Though the intractability of the problem of induction qualified this foundation, the success of scientific theory in predicting observation convinced (however circularly) that this was a problem that would be solved.

<sup>44</sup> "6.22 The logic of the world, which is shown in tautologies by the propositions of logic, is shown in equations by mathematics."  
(Wittgenstein 1921)

But the critical blow was delivered by Quine(1951), who attacked both the observational and the analytic anchors to the objectivity of descriptive statements:

"Modern empiricism has been conditioned in large part by two dogmas. One is a belief in some fundamental cleavage between truths which are *analytic*, or grounded in meanings independently of matters of fact, and truths which are *synthetic*, or grounded in fact. The other dogma is reductionism: the belief that each meaningful statement is equivalent to some logical construct upon terms which refer to immediate experience. Both ... are ill-founded."

Quine's attack on the privileged position of both analytic statements and of observation reports is built upon one key insight, that although "the truth of statements does obviously depend both upon language and upon extralinguistic fact" yet "it is nonsense and the root of much nonsense to speak of a linguistic component and a factual component in the truth of any individual statement. Taken collectively, science has its double dependence upon language and experience; but this duality is not significantly traceable into the statements of science taken one by one."

So it cannot be said that the objectivity of any statement or class of statements is peculiarly anchored in the world, in the sense that once its sense is understood its truth value is deducible from the state of the world: there is no one-to-one mapping from perceptions of the world onto true statements about the world. And the analytic core is equally subject to pragmatic revision<sup>45</sup>.

---

<sup>45</sup> The questioning of the analytic, particularly of the inner sanctum of logic, seems to have taken far longer to penetrate modern philosophy. I take it that one of the objectives of Smiley (1995) is to illustrate how choices between logical systems can properly be influenced by pragmatic considerations - i.e. in the light of experience.

This insight, which Quine employs to deflate the claims both of the analytic and of sense-data reports, can equally be brought to bear against any claim for epistemic privilege for one domain of discourse over others: with no metaphysical procedure to establish the correspondence of the statements within the privileged domain to the world or to the a priori, the onus of proof is shifted onto the empiricist. Quine argues that "One effect of abandoning [the dogmas] is .. a blurring of the supposed boundary between speculative metaphysics and natural science." The epistemic boundary between statements of fact and statements of value is equally blurred.

A parallel view of language is provided in Wittgenstein's later philosophy. As Sabina Lovibond (1983) puts it, in her exploration of the possibility of developing a Wittgensteinian moral realism, Wittgenstein's is "a conception free from invidious comparisons between different regions of discourse". (Lovibond, 1983, Section 6).

Like Quine, Wittgenstein denies the separation between questions of meaning and questions of fact that alone would allow the specification of rigid decision procedures to underpin claims to objectivity:

"We do not learn the practice of making empirical judgments by learning rules: we are taught *judgments* and their connexion with other judgments. A *totality* of judgments is made plausible to us. ...

"But wouldn't one have to say then, that there is no sharp boundary between propositions of logic and empirical propositions? The lack of sharpness *is* that of the boundary between *rule* and empirical proposition." (Wittgenstein 1969 #319)

Hence, there is no language-theoretic reason for taking moral judgments, judgments that on their surface behave as empirical propositions, as less objective than descriptive judgments: in neither case will the truth value of the judgments be determinable by analytic rules relating the judgments to "experience" or "the world".



Yet we do not accord objective status to all judgments. There are cases where disagreement is considered genuinely to refer to a matter of fact, and others where it is thought insubstantial - a matter of opinion.

Lovibond distinguishes two types of cases in which objectivity fails: in the first, exemplified by the normal use of the predicate "fun",

"complete agreement is not only absent, but the pursuit of it would actually run counter to the spirit of the game.... the 'pull towards objectivity' hardly acts upon me at all." (Lovibond 1983 Section 17).

There are other cases where the purposes of the language game would be better served by objectivity, but where "the reach of intellectual authority is relatively short". Lovibond takes moral discourse to suffer somewhat in this way. I will argue that moral judgments as a class are not short of authority. However, in moral discourse as in descriptive (and perhaps more so in aesthetics<sup>46</sup>), there are many cases where we are obliged to accept that differences are not susceptible to resolution, and hence we are driven to conclude that there is no fact of the matter.

If we cannot use presence or absence of decision procedures - rules of meaning - to determine whether a statement or a domain of discourse is to be accorded objective status, what can we use? Where else can we turn to discriminate between objective and subjective domains of discourse?

To discriminate, we want a definition of objectivity. Dummett's definition of realism will serve well:

---

<sup>46</sup> Even in aesthetics, however, there are facts. Although it is said that "beauty is in the eye of the beholder", there are objects - e.g. a Francis Bacon painting - which to classify as beautiful would be to invite incredulity.

"... realism cannot be characterised in purely metaphysical terms: it essentially involves the semantic notion of denotation, as well as the semantic notions of truth and falsity. Integral to any given version of realism are both the principle of bivalence for statements of the disputed class, and the interpretation of those statements at face value, that is to say, as genuinely having the semantic form that they appear on their surface to have." [Dummett 1991 p325]

The realist/objectivist need not (and, if Quine's attack on the correspondence picture is taken seriously, cannot) claim "linkage with reality" as the foundation for her views.

What she will claim is that a statement that is judged "objectively true" is binding upon the rational in a way that "subjectively true" (probably an oxymoron) statements, or statements that are interpreted away from their face value are not. Hence, if "one ought to pay taxes" is true, then that fact will weigh in a rational agent's deliberations in a way that it would not if: i) the statement was thought to be mere opinion, such that bivalence did not hold; or ii) if it were interpreted not at face value, perhaps as reducible to some other facts (e.g. about speaker's preferences, or local custom). Moral statements, if taken as objectively true, have the power to motivate, or better, to provide justificatory reasons for action, in a way that a subjective claims could not<sup>47</sup>.

On Dummett's view, the focus of the debate is not on metaphysics - what is in the world - but on how we can best explain our understanding of the terms involved. The question then is to justify the authority that we attribute to moral statements to the extent that we treat them at face value and interpret them bivalently. To this we will return in the next section.

Blackburn (1981), however, is unhappy with this language theoretic approach to the metaphysics of morality. He does credit moral judgments with bivalence and

---

<sup>47</sup> cf Dancy (1993, Chapters 1 and 2), following McDowell (1978)

does interpret them at face value, but claims that the attribution of objectivity, whether a realist or merely a "quasi realist" understanding is appropriate, remains as a further question:

"The sentence 'a is good' is indeed true, in English, if and only if a is good.... That is, if and only if we are committed to the goodness of a will we allow that the English sentence is true. That is its rule of use. But saying this tells us nothing about the kind of commitment it is: it is quite irrelevant to the metaphysics." (p.180)

Yet, by glossing the truth of *a* as bespeaking our *commitment* to *a*'s goodness, Blackburn is presupposing his own metaphysical view - that values are projected on to the world. Following Dummett, we can argue that if bivalence holds, and if we are prepared to speak about moral facts (a face-value interpretation which Blackburn endorses), then a fullblooded realism is appropriate.

In a footnote (6) Blackburn says that he does not understand the view that would derive metaphysics from theory of meaning. The answer is surely that there is nowhere else to turn. Dummett again:

"How, then, can such disputes be resolved? My contention is that all these metaphysical issues turn on questions about the correct meaning-theory for our language. We must not try to resolve the metaphysical questions first, and then construct a meaning theory in the light of our answers. We should investigate how our language actually functions, and how we can construct a workable systematic description of how it functions; the answers to those questions will then determine the answers to the metaphysical ones. For the metaphysical questions are formulated in terms of the appropriate *picture* of the reality to which our statements relate:.... The realist argues that an independently existing material universe is the only hypothesis that explains the regularities in our experience. The idealist retorts by asking, with Berkeley, what content the belief in an autonomous realm of matter can have. [Note the parallel

bafflement of the non-cognitivists or quasi-realists confronting a claim that values are of "the world"]. It is however useless to carry on a debate in favour of one or other of these competing pictures as if they were rival hypotheses to be supported by evidence. What we need to do is to formulate theses which are no longer in pictorial language but which embody the intended applications of these pictures. If we do that those theses will be found to be theses belonging to the theory of meaning, theses about the correct meaning-theory for statements of one or another kind." (op.cit. p338-9)

Dummett thus deflates the realist:anti-realist debate, removing the metaphysical rhetoric. A similar deflation of the metaphysical question of objectivity in ethics (moral cognitivism) is also appropriate, for the same reason: there is nothing cogent that can be said by either side in the language of metaphysics that could convince the other to accept their picture of what is in the world, and what merely projected upon it.

What should guide us in determining whether we have got an adequate meaning-theory for an area of discourse, if we are to eschew reliance upon the resulting metaphysical picture? Dummett offers three pointers. These will be useful in exploring the case for moral realism.

Notwithstanding his sympathy with anti-realism, Dummett first concedes that:

"...classical logic may be said to be in possession...any meaning-theory that can be shown to be workable and that validates classical logic is to be preferred."

This would certainly advantage cognitivism over some forms of non-cognitivism in ethics. (Whilst as we will see even a cognitivist meaning theory for ethics will have to allow that some judgements, for example where the lack of consensus indicates to us that we are at the uncertain penumbra of the application of the

moral terms, are indeed lacking determinate truth value, non-cognitivism in its simplest form denies determinate truth value to all evaluative statements.)

The second requirement, that a workable meaning theory "should accord with our practice, to the greatest extent that is possible", extends the advantage against those non-cognitivist theories that maintain bivalence but analyze away the appearance of objectivity in ethical claims (e.g. by claiming that evaluative claims are true only relative to a culture, belying their authority in our practice).

Common usage lies with the cognitivists: few (least of all political philosophers) can resist using moral discourse as if moral claims had determinate truth value. Taking non-cognitivism seriously would require a revolution in our mode of thought and life.

The tension between non-cognitivism and common usage extends also to self-justification. What form of argument can one use, even in developing one's own ethical theory, if moral beliefs are based on personal preference? This point is put by MacIntyre (1981, "After Virtue"), where he maintains that although moral discourse today is without foundation, yet the terms used are remnants from a golden age -- an age in which moral discourse was anchored in a common conception of the virtues necessary for the pursuit of the good life.

The problem with modern discourse, according to MacIntyre, is the lack of authority : for the definitive difference between a moral principle and a mere whim may be taken to be the authority that such a principle has over us -- the fact that we are bound by it. The authority that moral principles possess for me must derive from the reasons for my 'choice' of principles. If, as is the case with Kirkegaard (whom MacIntyre regards as the father of emotivism), moral choice is not founded upon moral fact, but is rather radical choice, without reason, the principles selected must be 'devoid of authority' (1981, p.40). I might indeed adopt such a principle from whim or caprice, or for some arbitrary purpose -- I just happen to like acting in this way, or just happen to like others to act in that way.

It is clear from our use of moral prescriptions as binding on self and others that the language of morality assumes that the moral claims do possess the cogency of authority, a cogency conferred by objectivity. Following Dummett, MacIntyre might therefore have taken the dissonance between the emotivist theory that prevailed in theoretical discourse and the authority evident in the use of moral language, as evidence that the emotivist theory was false.

Confessed emotivists, like Ayer, do not concede that their moral claims are indistinguishable from whims. For it seems to them that they can distinguish between those of their prescriptive beliefs which are motivated by whim, and those which represent moral principle. But it is not easy to see how such a distinction is to be sustained in the absence of objectivity.

Distinctions between the altruistic and the selfish, which are intuitively appealing, are also unhelpful, as they presuppose a particular narrow view of morality.

Don't we know when a universal precept is a moral one?; we simply ascribe to it a certain authority. But how is one to guard against self-deception? How to be sure for instance, that the high moral principle of 'family integrity' lies behind one's claim as to the acceptability of nepotism, rather than a desire to justify one's own conduct in accepting a job in the family business? How can conscientious objectors sustain a claim to be better than cowards?<sup>48</sup> Only the claim of objectivity, with the concomitant legitimacy of a search for the truth, allows discrimination.

An assumption of objectivity is at the heart of our ethical discourse<sup>49</sup>.

---

<sup>48</sup> A parallel concern may lie behind Russell's comment: "I cannot see how to refute the arguments for the subjectivity of ethical values, but I find myself incapable of believing that all that is wrong with wanton cruelty is that I don't like it", Russell (1960) quoted in Wiggins (1987).

<sup>49</sup> Brink (1989) offers many further examples of the implicit and explicit realism of moral discourse.

Yet, this begs the question against the moral sceptic. The appearance of objectivity may be mere sham; cynics may claim that if we examine our linguistic and nonlinguistic practice more closely, we will find that moral statements are not accorded authority, but are the stuff of personal creation. MacIntyre, at least at his most despairing, suggests that although cognitivism was a plausible claim historically it is no longer so (loc cit). Mackie (1977)'s startling thesis is that common sense morality is based upon the *fallacy* that there is an objectivity towards which moral discourse strives.

In the next two subsections I will argue that the place to look to evaluate both whether or not we do treat moral (or descriptive) claims as bivalent and whether we are justified in so treating them is in the degree of consensus in our use of the relevant terms. And I will suggest that a differentiated understanding of moral as well as of descriptive discourse is appropriate: a cognitive interpretation is justified - but not for all moral discourse.

First, however, Dummett's third criterion for realism is worth exploring. It is central to his understanding of what a meaning theory should do, and may be regarded as a final a priori test of the appropriateness of a realist interpretation of moral discourse:

"...a workable meaning theory...should enable us to explain in a non-circular manner ... what a speaker's grasp of the sense of any expression is. .... Such an account must be given in terms of how that knowledge is delivered to him, and hence how it is manifested in his observable linguistic and non-linguistic behaviour" (op.cit. pp.340-1)

Here, the moral cognitivist can give an account that involves the speakers of the language learning the use of moral terms in much the same way that they use simple descriptive terms, by following the example of fellow language users, noting similarities, and ultimately grasping (though rarely being able to articulate

their knowledge) what it is that is shared by the actions or situations that correctly bear particular moral predicates<sup>50</sup>.

The non-cognitivist may claim that this is no explanation at all as there is nothing there to grasp. However, the non-cognitivist is here in a much weaker position than the anti-realists in the other disputes of which Dummett speaks. These other anti-realists claim that their counterpart realists (about character-traits, about inaccessible regions of space-time, about mathematics, etc.) can give no account of how the relevant concepts are acquired. This the moral cognitivist can provide: the cognitivist claims that the moral status of actions is open for all to see, whilst the space-time realist for example cannot claim such access for anyone.

Note that Dummett is not asking the realist for an explanation of understanding in other terms. Dummett distinguishes between naïve realists and sophisticated realists. Whilst the latter would reduce descriptions in the disputed domain to claims in another undisputed domain, a naïve realist is someone who, with regard to the disputed class of statements, responds to the question "what is it in virtue of which the true statements in the disputed class are true", with a vacuous answer:

"He has, in fact, no general answer; and, for a specific statement of the disputed class, he can give only a circular answer - .... that the statement, "Andromeda galaxy rotates", if true, is true in virtue of the rotation of the Andromeda galaxy." (op cit p.328)

Now, this response, which Dummett characterises as "lame" may impugn the naïve realist's position regarding statements the truth of which is more or less inaccessible (as his example indicates), for it begs the question how understanding of such an inaccessible condition could have been learnt. But it seems a very reasonable response for singular moral judgements of the form:

---

<sup>50</sup> See Lovibond (1983) section 15 for an account of this process.



"that action was despicable". And indeed Dummett concedes, in attempting to explicate the corresponding notion of "bare truth", that

"any specific way of construing the notion of a statement being true in virtue of the truth of other statements will require that *some* true statements be barely true; and the naïve realist regards the true statements of the disputed class as barely true because he holds the disputed class to be irreducible." (op.cit. pp 328-9)

Dummett need not be taken as making a foundationalist claim that barely true statements are unsupported by any theoretical statements. Merely that we do not normally require support from other parts of the system when using such sentences or when learning their use.

Hence usage of language, and our understanding of its acquisition in no way drives us from a face-value bivalent interpretation of moral language.

This Dummettian approach to the ontological commitments implicit in our use of language should allow the cognitivist to by-pass the objection that she is adducing some moral sixth sense with which value is perceived. Ascription of value like ascription of any predicate requires judgment in interpretation of the evidence of the senses<sup>51</sup>. There is no distinct epistemological problem in understanding how judgements of value are possible.

We now turn to the question whether in practice we are justified in face-value and bivalent interpretation of moral statements: can an answer be given to the sceptic who would assimilate moral judgments to matters of taste? MacIntyre's (1981) claim that the confrontation of different cultures forces us to abandon claims to objectivity, whilst leaving vestiges of objectivist discourse, would explain away the surface moral realism of our ethical discourse. How can we tell if he is right,

---

<sup>51</sup> For a similar view, see Wiggins (1998, p.235) who interprets Aristotle as assimilating the ascription of geometric and moral properties, both being examples of *aisthesis*.

or whether conversely we can and should still sustain an objectivist interpretation?

## **6B A Public Language Test of Objectivity**

The lack of a link between language and reality at the level of the sentence, whilst consistent with equality of objectivity for all domains of discourse, may appear to justify a general scepticism regarding the "world-guidedness" of our beliefs. In ethics, more than in other domains, such scepticism has tended to degenerate into a subjectivism.

But the step to scepticism is misguided: a Wittgensteinian antidote can restore realism to ethics as to other spheres, and can also help to discriminate those statements for which a subjective interpretation *is* appropriate.

It is to the temptation towards scepticism that Wittgenstein alludes with his aphorism: "Not empiricism and yet realism in philosophy, that is the hardest thing" (1978 VI 23)<sup>52</sup>.

Nevertheless, the tendency towards scepticism can appear strengthened by the private language argument:

---

<sup>52</sup> This remark contrasts with "The difficulty is to realise the groundlessness of our believing" Wittgenstein (1969 para 166) - but the context of the latter statement in *On Certainty* is an argument to show that this groundlessness is not itself a ground for incredulity. What follows from awareness of "Grundlosigkeit" is not scepticism but rather realisation that the link with reality can only be articulated **within** the body of beliefs. Truth/justification/assertability conditions for a sentence have no metalinguistic privilege; their link with reality in turn is specified by further conditions that can only be stated within the body of belief. As we have seen, this circularity undermines any claim to meta-linguistic privilege for a particular subset of sentences: any statement that attempts to isolate a set of sentences for privilege is equally a statement within science (in the broadest sense), and hence subject to review and revision, in the light of experience.

"...to *think* one is obeying a rule is not to obey a rule. Hence it is not possible to obey a rule "privately": otherwise thinking one was obeying a rule would be the same thing as obeying it." (Wittgenstein 1953, paragraph 202).

Wright builds a general scepticism upon this foundation:

"The question is whether it is legitimate to think of our mastery of a language as involving anything properly seen as a capacity to recognize what preservation of ... patterns [of use] requires. For suppose that you find yourself incorrigibly out of line concerning the description of some new case. ... you can't, single-handedly as it were, give sense to the idea that you are at least being faithful to your *own* pattern; that is that you recognise how you must describe this new case if you are to remain faithful to your own understanding of the relevant expressions. How then would your disposition to apply the expression to a new case become, properly speaking, recognition of the continuation of a pattern if it so happened that you were *not* out of line, if it so happened that there was communal agreement with you?" [Wright, 1981, pp102-3]

Wittgenstein implicitly provides the answer to this sceptic<sup>53</sup> by restricting his argument to a language-user considered in isolation from her community.

---

<sup>53</sup> An alternative sceptical interpretation of Wittgenstein is provided by Kripke (1982) and can be answered in similar vein. Because Kripke insists (without warrant in my view) that Wittgenstein's sceptical argument is metaphysical rather than epistemic, he cannot avail himself of the solution set out in the text, below, one which builds upon the fact of our success in using language. See p.109 where he eschews this solution. For the metaphysical interpretation of the doubt, see p.21: Kripke slides from the lack of a fact about my mind that will guarantee that I will go on in the same way to assume that there is no fact in the world as to whether I am going on in the same way. Note that if the doubt is an epistemic doubt, there is no threat of regress: we seek only to justify a realistic interpretation of our statements; such justification can persist even if it is possible to entertain a doubt in its regard. (This answers Kripke's worry, op. cit. footnote 87; we can know without knowing that we know.)

Predicating a private descriptor of an object is a paradigm candidate for subjective (and hence non-realist) interpretation: either the statement is meaningless (and thus fails bivalence) or it must be interpreted away from face-value, e.g. as expressing something about the utterer. Use of a predicate like "pleasant" or "fun" for which concordance with consensus is not demanded, can be understood to refer back to the utterer. (Note that thus reinterpreted, there is a different reference to consensus, and a different objectivity, associated with the reinterpreted statement: one goes on predicating "pleasant" or "fun" correctly only if others agree that the utterer is finding an experience pleasing or enjoyable.)

Conversely, when predicating a public descriptor of an object, one for which consensus is demanded, it is the communal agreement itself that both makes meaningful and supports the hypothesis that one is continuing to use the relevant expression correctly - that one is picking out the same objective characteristic.

The absence of an independent criterion to give you comfort (a definition of "game" to assure you that the new activity really is a game, that you are not now going wrong) is made good by the consensus (you think to yourself - it must be a game - they could not all be going on in the wrong way).

"So you are saying that human agreement decides what is true and what is false?' - it is what human beings say that is true and false; and they agree in the *language* they use. That is not agreement in opinions but in form of life.

"If language is to be a means of communication there must be agreement not only in definitions but also ... in judgments. ... what we call 'measuring' is partly determined by a certain constancy in results of measurement."  
(Wittgenstein 1953, paras 241-2)

In this passage, one should read "what we call measuring" as describing our normative practice: we are only willing to construe predicates as genuine

examples of sizing up the world where there is constancy. But when there is consensus, such an interpretation is appropriate.

McDowell (1981) rehearses the same Wittgensteinian argument as Wright, but eschews scepticism:

"[It] is a mistake to construe the argument as making a sceptical point... The aim is not to suggest that we should be in trepidation lest "possibilities" of the 1004, 1008, ... type be realized [as continuation of the "+2" procedure, i.e. possibilities that we will not go on using terms in the same way]. We are in fact confident that they will not, and the argument aims, not to undermine this confidence, but to change our conception of its ground and nature."

However, McDowell, though confident that there is a base for confidence, fails to find one, and quotes approvingly a passage from Stanley Cavell in which he claims that "the view Wittgenstein wants to recommend" is one in which:

"We learn and teach words in certain contexts, and then we are expected and expect others to be able to project them into further contexts. Nothing ensures that this projection will take place (in particular not the grasping of universals nor the grasping of books of rules), just as nothing ensures that we will make and understand the same projections. That on the whole we do is a matter of our sharing routes of interest and feeling, senses of humour and of significance and of fulfilment, of what is outrageous, of what is similar to what else, what a rebuke, what forgiveness, of when an utterance is an assertion, when an appeal, when an explanation - all the whirl of organism Wittgenstein calls "forms of life". Human speech and activity, sanity and community, rest upon nothing more, but nothing less, than this. It is a vision as simple as it is difficult, and as difficult as it is

(and because it is) terrifying." [Stanley Cavell, *Must We Mean What We Say*, quoted in McDowell [1981<sup>54</sup>].

This is unnecessarily mystical and obfuscatory: it takes Wittgenstein's remarks on epistemology and the theory of meaning to address a biological or psychological question about what enables us to continue using concepts. But there is no conceptual conundrum in this area. The explanation for our continuing to use "square" (or "cruel") correctly, is that the new cases are indeed square (or cruel), and we are sufficiently endowed with perceptual equipment to observe the new case, and with memory of the way in which we have learnt to apply the words in the past, to recognise the new case for what it is. The consensus surrounding the application of these terms assures us that we are indeed going on in the same way (even if we cannot specify what that same way is other than vacuously). We have indeed grasped the universal such that we recognise its objective application, despite there being no book of rules codifying in what that grasp consists.

We are thus driven to assert objectivity and entitled to assert objectivity in order to explain how it is that we continue to agree even in new cases "in normal circumstances"<sup>55</sup> (Wittgenstein 1969 paragraph 27). We are able to say that there is some characteristic *of the object* which we recognise (using our perceptual abilities and our memory) in determining how to characterise it.

This is an argument to the best explanation. Hence it may prove to be wrong in the light of new evidence about a particular judgment or set of judgments, evidence that fractures the consensus about the truth value of like judgments, or

---

<sup>54</sup> In a later work McDowell (1994, especially Afterword, Part III) directly confronts Wright's sceptical interpretation of Wittgenstein, and develops an alternative interpretation, termed "naturalised platonism".

<sup>55</sup> This unavoidably vague qualification is crucial: it situates an observation statement within the Quinian web of belief, wherein normality is specified. We will not be able exhaustively to define normality, but our enduring consensus must extend to recognition that the circumstances are normal.

that provides a different explanation of the consensus. A realm of discourse previously thought to merit bivalent face-value interpretation may prove subjective. (Subsequent sections of this chapter will argue that this has not happened to moral discourse, *pace* MacIntyre.)

Although enduring consensus is a reliable discriminator between objective and subjective attributions of predicates, it does not provide an **explanation** of how it is that true objective attributions reach out to the world itself, rather than merely reflecting subjective opinion.

This explanatory anchoring to the world is just what we cannot have: that a domain is objective is just a fact about the world revealed by our consensus. In *On Certainty*, backing Moore's naive realism, Wittgenstein focuses upon the particular, and emphasises that no further explanation of the circumstances in which certainty is possible is attainable:

"One may be wrong even about 'there being a hand here'. Only in particular circumstances is it impossible. - 'Even in a calculation one can be wrong - only in certain circumstances one can't.'

"But can it be seen from a *rule* what circumstances logically exclude a mistake in the employment of rules of calculation? 'What use is a rule to us here? Mightn't we (in turn) go wrong in applying it?'

(Wittgenstein 1969 #25-6)

Wittgenstein then proceeds to declare scepticism inappropriate - but is careful in the parenthetical remark to emphasise that this is not arbitrary: "This doubt isn't one of the doubts in our game (But not as if we *chose* this game!)". This is simply the way in which we are aware of the world.

It may be that the spurious yearning for an explanation for our ability to discern truth, together with the remembrance that the correspondence theory seemed to provide a genuine such explanation but only for "facts", is one cause of the

persistent attribution of metaphysical weight to the contrast between the evaluative and the descriptive amongst purported observation statements.

Even Wiggins (1998) who argues that one of the "marks of the concept of truth" is "the kind of convergence in belief that truth commands where *the best explanation of the agreement in belief that p is inconsistent with any denial on the explainer's own part that p*" (page 151), nonetheless persists in seeing this sort of truth as distinct from an apparently privileged notion of "absolute truth". The latter according to Wiggins pertains only to "purely natural terms (terms that pull their weight in our theoretical-cum-explanatory account of the mechanisms of the natural world)" (pp195-6), whilst "plain truth or truth for the working day" can apply to predicates that "exist solely to denominate properties and thing-kinds of purely anthropocentric importance" (p.343, fn24<sup>56</sup>).

Wiggins' distinction between "absolute truth" and the anthropocentric variety may rest on the concern that the best explanation of continuing consensus about judgments in a particular human domain may not be that there is a quality picked out that persists in the world, but rather that there is something about human beings that consistently responds in a particular way to what is in the world. But this concern is equally applicable in scientific discourse: how would it be if someone said that she thought that shape inhered not in the objects but in the way we perceive them, taking a projectivist stance? We would respond that there must be something about the objects that triggers off our propensity to attribute

---

<sup>56</sup> The latter quote is from a footnote in the 1998 postscript which relates in part to Wiggins' 1987 essay, "A Sensible Subjectivism", the source of the first quote. It is telling that the distinction between natural and non-natural properties is elucidated only parenthetically and in footnotes. It is a conventional distinction that would not, I trust, survive frontal attack from Wiggins. It is absent from Wiggins (1984) treatment of predicates, in which he supports the view (attributed by Dummett to Frege) that the reference of a predicate is a concept (rather than the objects that fall under the concept). The concept thus referred to must be a feature of the world (for that is the function of reference), one that is fundamental to scientific explanation (see section 10 of Wiggins (1984)).



to them a particular shape, else we would not respond in the same way to the same object ... why not call that quality their shape?

Human involvement in recognising properties is the same whether the properties are ethical, psychological, or physical. As Wittgenstein taught us, all language has anthropocentric importance: its meaning stems from its use. But, even granting that there are some terms that are more closely connected with human endeavour, such that (for example) they would have no application in describing a universe without humans, it is unclear why statements involving such terms should be denied "absolute truth". (Note that not only "right" and "wrong", but also "chair" and "man" would fall victim to this discrimination.)

Should we rather avoid invidious discrimination between areas of discourse, and extend objectivity to all predicates passing the public language test?

Well, objectification (such that claims can attain "absolute truth" - i.e. truth simpliciter) is *an option* whenever we have enduring consensus in attribution. But in some cases we may choose to give priority to features of our response rather than to the features of the object that incite that response. This is the case with "fun", as Lovibond notes (*loc cit*), for in this language game it is our response upon which we usually wish to focus, so that you cannot be wrong about what is fun.

Blackburn (1981, pp176) notices that there are exceptions even for these predicates: we sometimes wish to focus upon the features of the world that justify humour, such that we might indeed say "you are wrong to find that funny". I am arguing that whether we can sustain such a critique of individual practice within a domain depends upon enduring consensus in the use of words like fun in core cases. Clearly, as Lovibond and Blackburn both acknowledge, this is not our normal use of the word "fun", but it may be that consistency in use does allow us to recognise an objective property that can be the base of a critique of aberrant practice.

By contrast, Blackburn's projectivism has nothing to say to explain such apparently objectivist critical usage, either in justification or otherwise - it is just a raw fact of our attitudes that we sometimes say that it would be better not to find a particular incident funny. This would be a rather lame response to the immoral scoffer if there is a fact of the matter about the appropriate objects of humour; and an unnecessarily lame response to the killjoy if there is not. We need a theory to tell us whether there is a fact of the matter. The public language test, with objectivity as the best explanation of enduring consensus where it obtains, provides such a theory.

Blackburn himself notes the emptiness of the scepticism that would insist that all qualities are projected onto the world rather than really being of it. However, he sees values as different:

"... the main reason why projectivism might turn out to be no real rival to realism about, say, our description of the world as containing causally interacting particulars in space and time, is that we may lack a conception of the reality upon which this creation is raised; hence we could have no explanation of how this thin reality works on an imaginative mind to give us our thoughts. This difficulty does not afflict the evaluative case..." (p165)

My suggestion is that the fact of enduring consensus, if it obtains in evaluative domains, does give rise to the same difficulty: we lack an explanation of how a value-free reality could act upon the imaginative mind consistently to give us value-laden thoughts. Consensus regarding the extension of a predicate is an indication that it has a determinate projectible sense, a sense that is a mode of presentation of a real feature of the world. In Wiggins' (1984) understanding of Frege, this would (I take it) be a case in which a predicate has a reference, the reference being a concept, the categories that inform our understanding of the world. (See also Sober, 1981, who propounds the anti-nominalist thesis that a full account of the metaphysics of science must include, along with physical

objects, also the properties picked out by the predicates used by the various sciences. I would wish to extend this argument to the science of morality.)

There is no more to objectivity but also no less, than that we say that there is something in the world explaining our continuing agreement if and only if we do go on agreeing.

### **6C Objectivity in Moral Discourse #1: Focus on the Particular**

Imagine that we have consensus in the application of moral terms, and that we categorise new actions without fragmenting the consensus.

It would then be evident that the groups of actions collected under moral predicates were not arbitrarily so collected. We would rightly ascribe the moral predicate to the actions themselves: the characteristic of the actions that ensured agreement on new applications.

The characteristic that allowed us to extend the predicate "right" to new actions, and to continue to agree on its attribution, would be reasonably labelled "the rightness of the actions".

But have we adequate continuing consensus in moral judgment to justify the attribution of objectivity as the best explanation of that consensus?

The prevalence of conflict between competing moral theories is notorious, and as disagreement extends to the fundamentals of moral systems, there seems little hope of constructing a framework for resolution: we appear to have no decision procedures for establishing moral truth comparable to that used in the natural sciences; how then could we expect consensus in moral judgment?

Yet it is at the level of the observation report, upon which all decision procedures depend, and which do not in turn depend upon decision procedures, that

objectivity must be assessed. This rather than high theory is closer to being fundamental (though this needs qualification - see below). For observation statements belong "to the most primitive part of language...: our knowledge of that in which the truth of such a statement consists, of what makes it true, if it is true, really does lie in our knowledge of what it is to *perceive* or *recognise* that it is true." (Dummett 1991, p.345)

The arguments in the preceding sections show that continuing consensus is the only ground available for attributing to ourselves knowledge of what it is to perceive or recognise the truth of observation statements.

At the level of singular moral judgements - moral observation statements<sup>57</sup>, consensus is most evident for terms with heavy descriptive content, what Williams terms "thick" moral terms like "cruel". Other terms too, which we do not normally think of as evaluative, have heavy evaluative implications - words like "man" and "citizen". Indeed there is no firm line between the evaluative and the descriptive: in the right circumstances "The light is red!" can carry strong evaluative and action guiding force. Yet, there will be consensus in its application<sup>58</sup>.

Even regarding the core terms of morality, "right" and "good" there is great consensus over the application of the terms to many new cases. Consider humdrum cases where the evaluation of an action is uncontroversial - the obvious cases that rarely make it into philosophical treatises: particular acts of wanton cruelty to innocent people, or of compassion to the sick and the dying.

---

<sup>57</sup> The phrase "moral observation statements" sounds odd: judgement is required in establishing moral truth; you cannot merely observe it. But so too with descriptions. And in *many* moral cases the degree of judgment required is minimal.

<sup>58</sup> And the consensus is apparently not limited to the descriptive content of the judgment: imagine a traffic-light where the red glass has smashed, but the light bulb shines white - we would still rightly exclaim "the light is red!" without fear of contradiction.

What is required to justify ascription of objectivity is not constant consensus in theoretical discussion, but continuing agreement in reacting to different circumstances, and, further, in adjusting our particular judgments to new information. If it is reality that guides our evaluations, that explains our agreement in going on, then mistakes should be discoverable by new observations. Hence in investigating objectivity, we are asking with Lovibond "whether recognition transcendent moral facts .. can figure in our discourse. Do we know 'what it would be like' for *everyone* to be wrong about a question of morality' - Lovibond (1983, section 19, p.77). If moral judgments are world-guided then an observation that the world is different in a relevant respect should drive us to change our judgment. (A further requirement that we should understand the possibility of *persistent* error is discussed below).

Lovibond's answer is that there is a profound difference between the moral and the natural scientific case:

"The idea of a circumstance which may obtain beyond the awareness of the entire community is bound up with that of a possible experience which would prompt a *more or less uniform and immediate* response among competent speakers, in terms of their becoming disposed to make a change in their assignment of truth values to particular sentences... I am suggesting that the reason it makes sense to suppose we might all be wrong about a question of astrophysics is that *observation* ... might at some time or other *show* that we were.

"This is not what it is like for a community to change its mind about a moral issue... The reason .. is that changes in collective moral outlook .. tend on the whole to happen slowly, by contrast with the immediacy of changes in what we all believe about the physical world."(p.79)

Lovibond illustrates: "What I have in mind is the familiar fact that with regard to the question such as 'Has the milk gone off?' a certain experience will lead

anyone who understands the question to revise his theory from one moment to the next."

Surprisingly, Lovibond compares the sour milk example with a change in "moral outlook", an expression which sounds more comparable to a paradigm change in physics (with which Lovibond says she is not concerned), rather than to a particular judgment. To the judgments of fresh and sour milk it would be more appropriate to compare singular judgments of good and bad moral character or that a particular action is right or wrong. But such judgments are revised uniformly in reaction to new evidence: e.g. we judge that Sue is a good sort until we observe her shooting a passer-by; we revise our judgment again when we see a video tape that shows that the passer-by had drawn a knife. Were she to have compared like with like, Lovibond's distinction between the moral and the scientific would have dissolved.

There is certainly greater consensus regarding particular judgments than regarding moral theories and generalisations (just as there may be consensus regarding the simple descriptive statements that underpin rival scientific theories). The consensus that a particular murder is wrong will be robust against considerable conflict over why it is wrong. Rival theories will source moral opprobrium in - revelation, utilitarianism, or deontological theories, but all will agree that opprobrium is appropriate. The consensus judgment that a particular patriotic act is laudable may be explained either by a nationalist theory or by appeal to the universal value of community.

There may thus be underdetermination of moral theories by the particular moral facts. Can observational concord regarding particular moral judgments count as evidence for moral objectivity even where there is conflict between explanatory theories?

It could just be that value is not susceptible to explanation by theory: underdetermination of theory may be chronic. (However, this is unproven: moral theory may develop to a point at which it is determinate - this is perhaps one goal

of moral philosophy; see below.) It is also possible that no unified theory will emerge even in physics (let alone in social and human sciences). This possibility does not undermine objectivism in physics, any more than the lack of unifying theory in psychology undermines objectivism regarding particular attributions of anger or love.

Albeit, the attainment of a unified all embracing theory would add conviction to the particular judgments that it predicts; but such a theory is not necessary for those judgments to warrant bivalent face-value interpretation.

However, there is a limit beyond which particularism in ethics (the epistemic priority of the particular moral judgment over the general) ceases to be compatible with objectivism regarding the attribution of moral predicates. If the ascription of a particular moral predicate at a particular time is undetermined by previous judgments, it takes on the appearance of an arbitrary decision. For an extreme particularist such as Winch, who appears to take the appraisal of the moral requirements of a particular situation to be private even to the agent himself (see reference and discussion above, in Section 2Aiv), this arbitrariness seems unavoidable. On such a view the cognitivist/non-cognitivist debate would be without content.

I have argued rather that when a term is used anew with consensus, we are justified in claiming that there must be some resemblance to previous cases (that an objective characteristic must be shared even if we cannot define it other than trivially) to explain the coincidence of judgment. Hence although Wittgenstein's scepticism about rules undermines confidence that a private language user is going on in the same way, it does not have that effect in the case of enduring consensus: we do have reason to believe that we are going on in the same way, that there is something about the previous uses of a term that dictate how it is to be applied in a new case.

In a nutshell, concurrence of judgments about individual cases determines whether the ascription of objectivity is warranted (to explain the concurrence). If it is, so that we are correct to say that there is a fact of the matter about core moral

judgments, then (*pace* Winch) it may be possible to develop a theory that integrates and predicts moral judgments in the way that science integrates and predicts judgments in the domain of natural science.

To convergence of moral theory we shall return later. The point here, conversely, is that the failure to discover such an integrating theory does not vitiate claims to objectivity for particular moral judgments, so long as there is adequate consensus. The next subsection considers whether the consensus that we have noted is adequate in view of the extent of conflict that we must acknowledge even at the level of particular judgments.

**6D Objectivity in Moral Discourse #2: a Research Programme for Testing Error Theories and for Establishing which Moral and Descriptive Judgments Warrant attribution of Objectivity.**

Lovibond suggests that morality is a language game with limited intellectual authority, and hence limited objectivity, because the practice which constitutes obedience to a rule "*cannot* be rendered determinate because of conflict within the linguistic community" (1983, section 19, p.81). As we have seen, the appearance of conflict may be exaggerated by concentration upon differences about theory rather than particular cases, but there is certainly no shortage of dispute regarding the correct action in particular cases.

The question is whether the nature of these disagreements is such as to be consistent with moral realism as an explanation of what consensus there is.

Explanations of disagreement compatible with realism come under three heads: hard cases, vagueness, and error. The first two of these explanations would narrow the range of judgments susceptible to a realist interpretation: on the assumption that an anti-realist interpretation for some statements (the hard or vague cases) is compatible with a realist interpretation for other statements within a domain. (We will later have to determine into which category core claims



of partiality fall.) Such differentiation may represent a departure from the conventional understanding of the nature of moral realism. However, if realism and objectivity debates are indeed best understood as debates about the correct interpretation and use of discourse, there is no *a priori* reason for assuming that the same interpretation and use will be appropriate to all statements in a particular domain.

### i      **Hard cases**

There is commonly dispute as to how to act where two recognised values conflict. Perhaps two people each have a legitimate claim to an indivisible good. Or one person has to choose between loyalty to family and loyalty to nation. Disagreement over such cases does not betoken lack of objectivity regarding more humdrum cases involving (legal or moral) property rights or loyalty, so long as it is possible to determine whether a case is a hard one, independently of determining whether there is a consensus. (However, if disagreement is rare in a particular arena of moral discourse, the independence criterion might be dropped: we would still have sufficient consensus to require objectivity as explanation.)

A simple device that would achieve such an independent determination is to take as the criterion of whether a case is a hard one, not the existence of disagreement about its evaluation, but the existence of a consensus that the case is hard (whether or not there is disagreement as to how it ought to be evaluated). In the next sub-section, on vagueness, I will exemplify the complex interaction between this second consensus criterion and the primary consensus regarding the truth value of the moral evaluations. For now, note that it is part of the peculiar role of moral philosophy often to show that cases that appear straightforward are actually hard. For example, where there are disputes about which both sides feel strongly (draft dodging in the Vietnamese war, the legality of fox-hunting or marijuana or abortion, preferment for relatives in private companies), the philosopher's job may be to tease apart the different issues (the conflicting values of personal autonomy, impartiality and of communal or family

flourishing), so as to demonstrate that the issues are difficult and that judgments are not easy.

Attention to hard cases seems to have deterred philosophers otherwise sympathetic to the cognitivist cause. In particular Wiggins (1998) answers his question:

"whether all sufficiently informed intelligences ought to be expected to converge in a sufficiency of cases on a practical judgment of rightness, permissibility, *etc.*, and to converge in such a way that the best explanation of their adherence to the judgment commits even the giver of the explanation to accept it"

by claiming that

"We have no assurance ... that always, wherever there is a practical judgement that survives criticism at the level of the reflective, self-critical agent .. , all sufficiently informed intelligences that understand his predicament .. will converge on that judgment..." (p.174)

Wiggins asks this question and comes to this conclusion following lengthy consideration of a very hard case (Winch's 1965 discussion of Billy Budd's trial). It is this case that seems to lead him to suppose that there can be no expectation of consensus or (in his locution) convergence in the moral domain in "a sufficiency of cases".

Hard cases are not rare, particularly given the indefinite variety and heterogeneity of human interests and concerns to which Wiggins draws our attention. Yet Wiggins himself acknowledges that "certain practical judgments, especially judgments of the morally prohibitive kind" pass the convergence test easily (p.177).

What follows? In a footnote to his question quoted above, Wiggins suggests that the "sufficiency" for which he asks is to be glossed as including

"... all or most of the cases where there are practical judgments that we as agents do after criticism *etc.* persist in wishing to affirm, do not regret afterwards, and which other people concur in both at the time and later".

The equivocation between "all or most" (which I have emphasised) is critical. If the requirement is for "most" cases to pass the threshold of consensus, presumably statistical significance is at issue rather than mere majority. Enduring consensus must be sufficiently commonplace to justify the argument to objectivity as the best explanation. The force of the demand for explanation of coincidence of judgment, and hence the strength of the claim to objectivity, will thus be determined by statistical theory: the greater the size of the domain, and the more frequent is coincidence of view within it, the less likely the hypothesis of chance coincidence becomes. It is clear from this that total consensus is not required; and it is plausible that greater cogency will arise if the domain within which sufficiency is assessed is restricted by excluding cases acknowledged to be "hard" - a legitimate move if, as suggested, we use a second consensus about hardness to determine which cases to exclude. Hence, it is clear that the existence of hard cases does not in itself vitiate the argument to objectivity from enduring consensus.

Why then might Wiggins in the alternate demand that *all* cases should excite consensus? Wiggins' problem with a statistical approach may be a concern with the interpretation of the hard cases excluded, and the implications of that interpretation for our understanding of morality in general.

For such cases, an anti-realist interpretation would be appropriate. We might say that hard cases, where desiderata conflict, should be treated as we treat judgments in other fields where evidence underdetermines truth. For example, when judging someone's character to be courageous or not, we may say that where there is evidence there is bivalence (vagueness aside), and where there is no evidence, there is no fact of the matter. So also in moral cases, we may say that where a case is not "hard" (and there is sufficient consensus in like cases), a

bivalent interpretation is appropriate; but for hard cases (and in areas of discourse lacking consensus), there is no moral fact.

The problem here is that a moral realist in humdrum cases seems to be committed to seeing hard cases, where there is no fact of the matter, as different in kind from straightforward cases that are otherwise similar. And if the absence of moral fact in such cases is interpreted as a total lack of moral colouring, one might conclude that it would be inappropriate to strive or struggle. Yet as Wiggins says, one would never "give up the search for that which one would oneself be satisfied was the right thing for one to do (here and now) *simply* on the strength of becoming convinced that no answer to the question 'what is it right for him to do' would ever command principled consensus".

From this fact that we do struggle to find the right thing to do where there are moral dilemmas (the stuff of tragedies), i.e. in hard cases, Wiggins concludes that moral deliberation is not after all about a search for truth: for if it were, deliberation would come to an end with a conclusion that there is no fact of the matter.

Here are two responses:

First, we might claim that cases in which it is clear that there will never be a consensus are vanishingly rare: no cases are that hard that they need a separate metaphysical treatment. This approach would allow one to maintain the statistical model for determining whether consensus is adequate to force attribution of objectivity, searching for consensus first amongst prima facie straightforward ones, but then amongst prima facie hard cases too. This approach would proceed on the assumption that Wiggins eschews, that we might after prolonged deliberation "always" reach consensus. The struggle is a struggle to discover what will in the

full light of all the consequences of the actions, and with full maturity of judgment, turn out to have been the right course.<sup>59</sup>

Second (to echo the discussion of Raz's doctrine of incommensurability in Section 2Aiv), recognising that there is no fact of the matter does not imply that it does not matter: what is hard about a hard case is that either choice will involve sacrificing a good that is objectively recognised as good. In a hard case, it is not that moral colouring is turned off, but rather that the colouring is not black and white. Although we struggle in vain, we struggle in the hope that some way out will materialise; but ultimately deliberation may indeed give out. This response recognises that such cases are different in kind from straightforward cases, as moral evaluations lack truth value, but denies that that is inconsistent with the phenomenology of such cases - the sense of struggle.

Furthermore, the phenomenon of struggle may involve an element of wishful thinking; when tragic choices have no right answer, a confessedly arbitrary choice may be better than a strained pretence that the decision represents exercise of moral autonomy (Raz) or definition of character (Wiggins). Such an arbitrary choice is nonetheless appropriately accompanied by distress that it is necessary.

---

<sup>59</sup> Against this Blackburn argues that: "... unfortunately it is wrong to infer, from the proposition that routine consensus is sufficient to justify the notion of correctness in new application of a term, the startling conclusion that correctness may exist even when consensus does not." (1981, p.173). One might argue that in some circumstances, if there is reason to expect that developments will secure convergence, correctness may exist even in the absence of consensus. Nevertheless Blackburn is surely right to doubt whether this can provide a general solution to the problem of conflict. Yet he is wrong to conclude against realism for all ethical claims, for we may rely on the second understanding of hard cases presented here.

Conversely, to abandon a realist interpretation would leave unanswered the question of how to account for consensus where it does endure.

Wiggins suggests that the alternative to moral realism is non-realism rather than anti-realism (footnote 25, p. 165). I suggest, following the second understanding of hard cases, that once we recognise that anti-realism is the appropriate interpretative response, the appeal of non-realism overall is diminished, for a realistic interpretation of straightforward cases is facilitated. Hence non-realism is avoided by combining moral realism for straightforward cases with anti-realism for hard cases.

As the following discussion of vagueness should illustrate, failure of bivalence for some uses of expressions is by no means limited to moral discourse; hence allowing anti-realism in some uses of terms must be compatible with taking a realist stance against non-realists.

## **ii Vagueness**

There is consensus on which actions are wrong, and upon which objects are brown, but there are also many actions and objects where classification of an action as wrong or indifferent, or of an object as brown or black, is difficult. Here too neither difficulty nor disagreement about such cases compromises the objectivity of the normal case.

How shall we interpret the vague penumbra?

The public language test of enduring consensus must be determinate if it is to force us to objectify value as its best explanation. Hence, it would be better to avoid a vague characterisation of vagueness: the statistical significance of agreement will be diminished if within the relevant domain of judgments are included many disagreements relating to this border territory.

Were wine tasters to agree that some particular wines are "fruity" and others "not fruity", but to disagree over a substantial majority of cases, we would be sceptical of claims to objectivity of judgment even about the wines for which there was a consensus. After all, random selection would yield agreement about some wines. A claim that all the disputed cases were at the borderline between two objective categories, and hence should be excluded from the sample, would be derisory. We would insist that our statistical tests to determine whether agreement was explicable as random coincidence should include all cases. We would only allow the allegedly borderline cases to be excluded were there some antecedent test that differentiated them, akin to the consensus test of a hard case discussed above.

A parallel general test for a bottle of wine being borderline would involve the tasters independently deciding whether a bottle should be considered to be borderline. With two tasters, we will then have three classes of wine: that which both consider to be core; that which both consider to be borderline; and that which only one considered borderline. The most restricted objectivity test would include within the sample only the bottles which both tasters identified as core.

Were the sample for the restricted test large enough and agreement common enough, we might on its strength conclude that agreement could only be explained by an objective quality of the wine. The interpretation of any consensus regarding the fruitiness of the wine classified as borderline by one judge and core by the other would be more problematic. Could we use more limited consensus here to reinforce the view that these bottles really are closer to the borderline? What would we say if there were consensus here, or even regarding consensus-borderline bottles, but not in the first class? Presumably in that case we would discount the discrimination of whether bottles were "borderline" as subjective, and explore whether there was significant coincidence in judgment of fruitiness amongst the whole sample in order to determine whether fruitiness was an objective or a subjective quality.

The position becomes more complex if we have a multitude of tasters all expressing views not only on the fruitiness of the wine (and this should really be a continuous rather than a discrete variable) but also on whether the judgment is core or borderline. We would need to model the relationship between winetasters' judgement of the coreness of a wine, and the reliability or objectivity of the linked judgment of fruitiness.

Let us assume that there is some correlation between judgement of coreness and consensus over fruitiness. Results of statistical tests would place each wine on a continuum:

at one extreme would be bottles of which we would conclude with great confidence that their fruitiness was a matter of objective fact - nothing else could explain the extent of agreement amongst tasters who rate this a core example of fruitiness (or alternatively of not-fruitiness - or even of medium fruitiness)

at the other would be the borderline bottles, about which there is consensus that they are borderline - i.e. judgment is very difficult; and about which there is no consensus regarding their fruitiness

and in the middle would be bottles that are more or less borderline: about which there is more or less consensus regarding whether they are borderline and more or less consensus regarding whether they are fruity. About such bottles we would have varying degrees of confidence about whether there is a fact of the matter regarding their fruitiness. We would be unsure whether the extent of consensus required us to say that there was a fact of the matter regarding their fruitiness, or whether on the contrary the extent of consensus could be explained by mere coincidence of subjective judgment.

Similar treatment is appropriate in the attribution of almost all predicates: not only colour predicates but also shape and moral predicates.



iii **Conclusion regarding Objectivity in the Presence of Hard Cases and Vague Penumbrae**

In the absence of statistically robust empirical research, we attribute objectivity on the basis of quotidian experience of consensus. For example, our shared understanding of a particular action (or object) may serve to display consensus (or lack of it) in the use of several terms. If I promised to deliver you a square brown table, coincidence in judgment that I am now discharging my obligation with this particular table, will rely upon the objectivity of moral, colour and physical judgments. Consensus in the use of any of these predicates in core circumstances, and consensus whether current circumstances are core, and consensus regarding the nature of the present action, will jointly determine whether we should say that there is a fact of the matter whether I have fulfilled my promise.

(Note here that it is empirical study of the use of terms - i.e. a study of their extension - which is epistemically determinate of whether they have a conceptual reference in the sense of Wiggins (1984). Meaning is use.)

I conclude that a nuanced response to the related disputes between realists and anti-realists and between ethical cognitivists and subjectivists is required. We should not expect a simple answer to these issues to be applicable across a field of discourse (however delineated). Each judgment will be accorded a bivalent face-value interpretation if there is consensus both that it is a core case of the use of the terms involved and regarding the truth value of core cases for those terms. (Of course, this does not preclude dispute regarding the truth value of the particular judgment under consideration.) Where one or both of these are lacking to a degree, a realist interpretation becomes less likely to be correct.

Where a realist interpretation is not plausible, we plump either for an anti-realist or a subjectivist understanding. Subjectivism is possible where we can interpret a claim to be focused not on its apparent object but rather on to the judge (i.e. the

speaker). (Hence, where consensus is lacking for a particular wine, a judgment that the wine is fruity can nonetheless be interpreted as an expression of approval - assuming the taster is partial to fruitiness.)

In the case of morality, I hypothesise (in the absence of empirical research) that there are cases for which a realist interpretation is appropriate, e.g. that there is statistically significant coincidence of judgment in the predication of wickedness to actions for which there is consensus that they are core cases - neither hard nor vague - for classification as wicked or otherwise. There will also be more or less borderline actions along a continuum ending with actions about whose merit we would conclude that there is no fact. But the existence of a large number of judgments about which there is no consensus in the moral domain (whether hard or vague cases<sup>60</sup>) is quite compatible with near complete confidence in the correctness of a realist interpretation of core moral claims.

What is the role of the moral philosopher in this enterprise?

I mentioned above the activity of philosophers in pointing out the complexity of many judgments. This might undermine consensus within groups who previously were quite convinced of the correctness of their judgments - whether individualists or authoritarians, impartialists or communitarians. Yet the somewhat surprising overall impact of philosophising in this way is to reinforce the objectivity of **core** moral judgments: by increasing the pool of cases for which there is a consensus that they are hard and hence to be excluded from analysis, consensus regarding the remaining judgments is enhanced, and the conclusion to their objectivity as the best explanation of consensus is reinforced.

Parallel work to influence the consensus regarding the hardness, or the centrality of particular judgments, paradigmatically those surrounding the evaluation of particular works or performances, is presumably also part of the core business

---

<sup>60</sup> There need be no clear distinction between the two: for example a hard choice of mercy over justice in a particular case may *ipso facto* be on the borderline between being right and indifferent.

of aestheticians and critics. (So also in many other disciplines: judgments of proficiency and style in sport, in crafts, in taste in food, etc etc).

Second, a neglected role of moral philosophers and aestheticians is to undertake the empirical research to determine the degree of consensus pertaining to core cases, and hence the soundness of the argument to objectivity.

#### iv Error theory

Williams (1985, chapter eight) argues that moral claims are not anchored in objectivity partly for the lack of an error theory to explain disagreement:

"An ethical theory...cannot do something that explanations of perception can do, which is to generate an adequate theory of error and to account generally for the tendency of people to have what according to its principles, are wrong beliefs."

This is a legitimate challenge. I have suggested a statistical approach to determining whether the coincidence in judgment in the predication of moral terms is sufficient to force the explanation of objectivity. But once we hypothesise objectivity we are left wondering why *anyone* fails to make the correct judgment. Vague and hard cases aside, the model demands that the specific moral property can simply be perceived, as the colouring and shape of objects can be perceived.

The sort of error in perception that Williams is considering is no doubt that relating to poor eyesight, for which there is a physical explanation.

Yet, poor eyesight (or weak memory) not only explains error in perception but also (*pace* Williams) in moral judgment: e.g. when a witness testifies to the guilt of an innocent woman because he mistakes her for another. And there are numerous other error theories that can be credibly be used to explain failure to express moral truth: disingenuousness; miseducation; an unwarranted

attachment to a general ethical theory. At an extreme, madness may be attributed to someone professing disregard for, or acting against, the right. All these theories are likewise needed to explain errors in descriptive judgments.

A word about miseducation and about madness.

That education enables finer discrimination of the good is consistent with moral cognitivism and moral particularism: as Burnyeat (1980) expounds Aristotle, "by doing the things you are told are noble and just you will discover that what you have been told is *true*. What you may begin by taking on trust you can come to know for yourself. This is not yet to know *why* it is true, but it is to have *learned that* it is true in the sense of having made the judgement your own..." (p.59). This fits well with the model developed above in which learning the meaning of moral terms involves coming to recognise that which the particular actions properly described by each term has in common, at which point you can yourself make appropriate judgments in new cases. (Aristotle and Burnyeat argue additionally that the moral character of an action can often only be properly appreciated by oneself doing the action, or like actions; a point I have not emphasised but which is certainly consistent with the foregoing discussion.)

Conversely, of course, someone lacking such training will be unable to discriminate appropriately, and this provides an explanation for their errors.

It might be argued that if coincidence in judgment is the product of education, then the inference to objectivity as the best explanation fails, for the coincidence of judgment can merely be attributed to the common education. However, a shared education could not on its own explain how the group members continue to agree on new cases. A real feature of the world must be being recognised. (It is possible in theory that what is recognised is not a moral feature of the world but some descriptive feature. This objection is discussed below in section F.)

Attribution of madness to those disputing our moral judgments may seem a dangerous practice - akin to the utilisation of mental hospitals for political

dissenters. But what is wrong with putting *political* dissenters in mental hospitals is that the judgments which they are disputing are usually ones for which we have *no* enduring consensus. Indeed it is the lack of consensus that renders manifest dissent dangerous to the establishment. By contrast, who would fail to doubt the sanity (or the ingenuousness) of someone who applauded the actions of a serial murderer?

There are many people with incorrect beliefs about the physical and moral world whose opinions we simply discount: witch-doctors, astrologers, para-psychologists, psychoanalysts, racists or UFOlogists. We will rely upon error theories to explain how they can be wrong: and if we cannot find an error theory, we assume that there must be one were we to look hard enough.

There is a limit beyond which we should not push such assertiveness. It is a matter for investigation whether errors are so baffling as to require revision of our understanding of correct judgment. This is a further field warranting empirical research, albeit one in which dispassionate assessment uncoloured by the moral outlook of the researchers would be very hard to substantiate. However, pending this research, it is not obvious that errors are more problematic for the objectivity of judgment in ethics than elsewhere.

In principle, where one can identify who is inappropriately educated and who is insane independently of their moral judgments (perhaps on the basis of their non-moral judgments), one can conduct the empirical research into moral objectivity only upon well educated sane people. Where we lack such criteria of good upbringing and sanity, however, we are forced to include all in the research base. Nevertheless, even were those capable of fine moral discrimination few, statistical significance might accrue to coincidence in judgment across the whole population, so long as a large enough number of judgments were sampled.

#### **v. Global Error**

Can error theories for cognitivist ethics explain global error? A realist interpretation of moral language should be consistent with the consensus itself proving mistaken, for the realist hypothesis is that there is a moral reality that transcends our recognition of it. Global error is possible for colour terms and for physical predicates more generally. For example, we might discover that a sunspot has systematically distorted our vision.

Could we envisage similar mistakes about moral judgments?

Blackburn (1981) alleges that "Wittgenstein regarded consensus as a necessary as well as perhaps sufficient foundation for the notion of correctness." But then: "If Wittgenstein leads us this way, however, he fails to allow for the possibility that goose-stepping along with everyone else can yet lead to moral error."(p 173)

Lovibond (1983) has a similar problem. She explicates the notion of correctness in the application of moral terms as founded in "the coercion exerted by a linguistic community upon its individual members", but then finds that there is "no one outside the *total* speech community who could exercise the necessary coercion". (Section 13, para 57)

As a consequence, Lovibond finds herself forbearing from criticising the practice of infanticide in a tribal community (pp 53-4). She takes it that there is no perspective from which their practice can be criticised; they cannot all be wrong. How much more difficult to establish an independent perspective where a practice is judged uniformly across all linguistic communities.

Yet this problem begins to dissolve if we take sufficiently seriously the language of realism where adequate consensus is established<sup>61</sup>. By our continuing

---

<sup>61</sup> Note that the position I am advocating is full-blooded not quasi- or internal- realism, or any other qualified version, for I am suggesting that the **only** proper foundation for a realist interpretation in any domain is enduring consensus about particular judgments as described; and I am suggesting (in the absence of empirical evidence to the contrary) that this is as common in evaluative as in descriptive domains.

consensus we discover that there is such a thing as rightness amongst actions and of malignancy amongst tumours and of validity amongst proofs. The world might turn out to be otherwise than we have perceived it to be in each of these domains - we might discover through observation and investigation that we had made significant and pervasive mistakes. (In each area we would be judging that the hypothesis of fundamental past error was more credible than the hypothesis of current error.)

There are two ways in which seemingly objective moral claims, underpinned by collective consensus, could prove in error.

First, the statistical model that underpins the judgments of adequate consensus itself allows for the possibility of error. The coincidence of judgment regarding the use of core terms, or more narrowly that a particular case is a core one, might prove to be mere coincidence. If consensus thus dissolved, we would conclude either that the particular case is a hard case (e.g. what we took to be a simple murder is a complex case of self-defence the morality of which is not clear cut) or at the vague penumbra of the application of the relevant terms; or, where break in consensus is pervasive amongst judgments using particular terms, that all judgments using those terms are to be interpreted either anti-realistically or subjectively.

Second, error may materialise that reverses the consensus on a particular case: it may turn out that Jones was not the murderer. The possibility of error may not be small even in the straightforward consensus cases: for the case may turn out to be other than we thought.

Could we all be in error not about the nature of the case but about its morality?

Many cases of moral error may be hard to disentangle from descriptive errors. For example, a group may be misled regarding the nature of a different ethnic group. Think of the numerous racist theories that adduce alleged factual differences between racial groups. Their racial beliefs apparently justify

discrimination: are they victims of moral or nonmoral error? Either way, it can be corrected, and recognised as a mistake. The theories can be exploded with empirical research in obvious ways. There is no conceptual problem in understanding the possibility of error here.

Can we conceive of finding ourselves misled in considering that hair colour does not in itself justify discrimination?

What would we make of someone who claimed that black haired people should be treated less well - without seeing the need to justify this claim in any way by reference to the character of those with black hair? Could we make sense of this error? If not, how can we envisage coming to see this as correct?

But similarly outlandish scientific beliefs would be equally hard to understand. In Moore's example, we cannot conceive of being convinced that the earth was created within the last century. And as Wittgenstein comments on this case:

"The *truth* of certain empirical propositions belongs to our frame of reference." (Wittgenstein, 1969, para 83)

I guess Wittgenstein uses "empirical" in this context with irony.

I suggest that the proposition that discrimination purely on the basis of hair colour is wrong belongs to our frame of reference. Foot (1958) made a similar point about the inappropriateness of "commending the clasping of hands as the action of a good man". In both cases, it would be paradoxical to claim that these moral insights are not objective because we cannot imagine coming to the conclusion that they are false.

Yet, if the cause of error is profound and durable enough, then a whole community may be misled for some time. Only when the cause is uncovered will we discover that we were in error with regard to some class of applications of terms. This is perhaps the account to be given of the enduring consensus that persisted until recently in believing that racial differences were grounds for discrimination. Suppose that historical research suggests that this consensus



was not based upon any separate mistaken scientific assumptions. The mistake was simply that skin colour was taken to be relevant. We might explain the error as the product of bias and lack of sound moral education. (The moral error here can be traced to a description inappropriately imbued with moral character - hence there is no enduring consensus lacking explanation.)

Some such error theory may also allow us to understand how it is that the infanticidal tribe of which Lovibond writes could have gone wrong so profoundly. (In principle it would be possible for a community to be deluded collectively into seeing things as they are not, without subsequent enlightened generations being able with certainty to pinpoint the cause of error. Yet this lack of explanation would be a cause for concern.)

Might we now be collectively deluded in regard to some class of judgments? If we might be, of what worth is our enduring consensus? Two points may mitigate this scepticism in its particular application to ethics.

First, scepticism on such a global scale is also possible in physics (its appeal waxes and wanes with the greater and lesser integration and consensus in science, just as in ethics). In physics, of course, error incurs penalties, which should in due course lead to its discovery. Can we say the same about ethics? Lovibond allows that "a change in social practice may incur certain *natural* penalties", but oddly does not explore whether this corrective might steer the community towards moral correctness. Whether such a mechanism enforcing correct moral judgment exists is discussed below (subsection G on convergence) where we note that there is no absolute assurance that we will be led to truth in physics. Nevertheless, if error has a cause that has distorted perception of value or of nature, we would expect discovery of error to follow from removal of that cause.

Second, examples of profound and global mistakes subsequently uncovered are not as common as sceptics sometimes imply. Enduring consensus makes mistakes less likely. The extent of moral error in other communities is perhaps

exaggerated by biased application of the principle of charity in translation - discussed further in the next sub-section. But further, errors rarely seize an entire linguistic community (defining a community as broadly as possible - as is appropriate in this context - to include all those aware of each others' judgments). Blackburn's allusion to the Nazis ("goosestepping along with everyone else") fails to recognise the near global condemnation Nazi practices excited (even if that condemnation was often suppressed).

To sustain a claim of objectivity for a realm of discourse, or, better, for the core application of the key predicates within it, does require us to provide an explanation of error. However, there is no *a priori* (or, so far, *a posteriori*) reason to believe that this cannot be done in the ethical domain as in the scientific.

### **6E Objectivity in Moral Discourse #3: How Charity in Interpretation should Limit attribution of Immoral Judgments to Other Cultures**

What of radically different moral systems? It is the occurrence of conflict between cultures that drives MacIntyre (1988, p168-70) - and many others - to despair of moral realism.

Wittgenstein (1969, #80-81) notes that "The *truth* of my statements is the test of my *understanding*. // That is to say: if I make certain false statements, it becomes uncertain whether I understand them."

Davidson (1973) inverts Wittgenstein's insight to claim that the members of one group attempting to understand an alien people are bound to interpret their actions and their speech in a way that maximises consistency of judgment across cultures. Were we to find ourselves attributing to the alien group inexplicably false beliefs, we would question whether we had properly understood them.

"We get a first approximation to a finished theory by assigning to sentences of a speaker conditions of truth that actually obtain (in our

opinion) just when the speaker holds those sentences true. The guiding policy is to do this as far as possible, subject to considerations of simplicity, hunches about the effects of social conditioning and of course our common sense, or scientific, knowledge of explicable error" (Davidson, 1973, p19)

If it is not possible to generate a translation manual that attributes true beliefs (or explicable error), we are unlikely to attribute rationality to the group at all. Hence, it may be incoherent to hypothesise a radical confrontation between cultures (*pace* MacIntyre).

This does not eliminate disagreement. Rather establishment of the core of agreement is necessary "to make meaningful disagreement possible". The difficulty then is to decide which sets of statements to consider core to our rationality, statements that should be rendered true if at all possible, and of which statements is rational debate about their truth conceivable.

**Ethical relativists implicitly assume that in interpreting others we should preserve truth regarding descriptive observation statements, even if we thereby find ourselves attributing gross immorality. Such a policy begs the question against ethical cognitivism whilst apparently undermining its plausibility by generating the appearance of intercultural moral conflict.**

Yet we may well have a choice between the attribution of a weird science and a wicked morality.

A people sacrifices its members in intermittent wars: is it wicked, or is it possessed of a set of social scientific beliefs that justify such acts?

If we have rather more conviction in *our* judgment that a *purposeless* sacrifice is wicked than we have in our judgment that there is (in this case) no social benefit justifying the sacrifice (or if we have a better error theory to explain their deviant

social science than to explain a deviant morality), we may choose to attribute to the people a morality like our own, but a different implicit social science.

Charity in interpretation is also relevant in more subtle ways in evaluating apparent conflict even within a language group. It is necessary to define the group of common language users before the significance of disagreement can be assessed. Take a discourse about astronomy. The language peculiar to astronomy is given reference and sense by all those who are mutually acknowledged to be communicating with each other about astronomy. There may be many different non-overlapping groups, each using a different set of terms perhaps in different natural languages to refer to the heavenly bodies, or the groups may overlap, so that there is really only one large group (if the groups span more than one natural language, then the specialist terms must be unambiguously translatable). Ascription of objectivity to the application of a term for such a language-using community will be justified if there is a continuing consensus regarding its reference (or at least on the appropriate decision procedure). If there is general disagreement over the use of a term, and there is no decision-procedure, then the application of the term must be deemed subjective.

However, disagreement may be attributable to there being two different uses of the term, within two different groups, which do not after all share a common usage for the term in question. In other words the best explanation of the phenomena observed may be that there are two different objective qualities referred to respectively by the two communities using the same phoneme. Even if Californians use the same phonemes as the Scots, if a substantial section of their beliefs appears to differ, it may make little sense to assume they speak the same language. "Patriotism" may be considered a virtue in one country but not in another, and yet both may agree that taxes should be paid and enemies repulsed.

The appearance of conflict in such cases would dissolve were a charitable interpretation to eschew homophonic translation. But care is needed. The

statistical cogency of coincidence of judgment would be slight were all cases of conflict relabelled as semantic discord. An independent criterion of semantic disagreement would be needed to justify this approach. (It is not just that the Scots and the Californians disagree about whether to be patriotic, but they also disagree about what being patriotic comprises.)

Apparent conflict may also mask underlying consensus where the social meaning of action varies between different groups: breaking a taboo may be wrong merely because it is disrespectful to ignore a taboo - this is consistent with the unexceptionableness of the same action in a different society. In such cases a full description of the social circumstances of a particular case would excite cross-societal consensus. Even conflicts within a society may be explained in this way - an individual who works in a factory, drinks in a club, has a family life, goes to church may be at the intersection of different groups within each of which different actions or words will carry different social meanings.

An example: the alleged prevalence of infanticide in China. This seems abhorrent. How can we claim moral objectivity on the basis of consensus if there are such differences in moral sentiments: how can they have got it so wrong? But the possible error theories outlined could be applicable here: perhaps the non-evaluative aspects of the actions are different from how we imagine; perhaps we should recognise that infanticide is not so different from abortion; disingenuousness - they do not really believe what they do is right. Or perhaps those who commit the infanticides would not justify their actions; perhaps they see themselves caught tragically between a positive injunction to provide for their families and a negative injunction not to harm a child. Are these error theories applicable? This is in part an empirical question: but only in part - the principle of charity may force us to give some such explanation to avoid attributing to them abhorrent moral views.

Parallel error theories are required to explain disagreements regarding the application of most descriptive terms; and there is parallel reason to avoid attribution of crazy views.

Is there continuing consensus sufficient to justify ascription of objective truth to singular judgments of duties of loyalty to family or to state? It would seem so, at least for some classes of action: notions of treachery and loyalty seem pervasive and would be out of place in an ethics that did not recognise the importance of human relationships. There are discordant judgments, but perhaps these may be accounted errors, and explained as misguided Godwinesque triumphs of incorrect ethical theories over singular moral judgements.

#### **6F Moral Objectivity is not Reducible to Descriptive Content**

Might the objectivity of moral judgments relate merely to the implicit descriptions of group norms - might recognition of these descriptions be what explains our ability to continue agreeing in the application of moral terms? In which case, the prescriptive force of moral judgments might have no objectivity, representing mere subjective opinion, or perhaps a culturally-induced consensus?

The sceptic suggests that our ability to carry on agreeing in collecting actions under moral categories may be explicable by some method for deriving evaluative conclusions from a descriptive base. This might involve a way of describing in neutral terms what it is that makes each action we call "right" right. Critically, someone without moral sensibilities, morally blind, having mastered the reductive theory, would be able to make new evaluative judgments accurately on the basis of mere descriptive discrimination.

Williams (1985, p.141-2) discusses McDowell (1978)'s contention that there may not in fact be a descriptive counterpart to an ethical concept. Williams finds this most plausible for the "thick" concepts (e.g. cruelty) which appear to be world guided, and hence for which there is likely to be consensus as to their application, but which nonetheless may include an ineliminable prescriptive content. The same argument, however, can be made for thin moral terms, like right or good, if it is accepted that there is a continuing consensus regarding their use in core cases. Williams concedes that "we should not assume that we could

see how people "go on" if we did not share the evaluative perspective in which this kind of concept has its point." He goes on to suggest that "moral philosophy" should "consider what must be said" if McDowell is right.

My claim is that if McDowell is right about a particular moral concept, if enduring consensus (in core cases) is only achievable with the adoption of a committed ethical perspective, then that moral concept should bear a realist interpretation - such that objective status is granted to the evaluative and prescriptive claims made using the concept.

What sort of evidence would count in favour or against the contention that you need to be able to recognise an imperative feature of the world to be able to use some moral concepts appropriately?

A reductive theory, linking evaluative judgments to descriptive ones, would give the lie to the McDowell hypothesis. Such an explicit reductive theory that could explain our continuing consensus in the application of a moral term, would be epistemically prior to the ethical judgements that it explained. Moral statements that were not supported by the theory would simply be mistaken.

For example, even if there were consensus about a Ghandi that his civil disobedience was justified, yet the reductive theory categorised such behaviour unequivocally as wrong, then we would have to conclude that the consensus was wrong about Ghandi. Of course, modification of the theory in the light of recalcitrant judgements of this sort would be legitimate, (e.g. by stipulating a principled exception to the ban on civil disobedience in the case of colonial governments); but ultimately there would be no room for a claim that a particular judgement was correct, however powerful the consensus behind it, if it did not find a descriptive anchor through the reductive theory.

If this approach was not credible - if the explanatory theory became too convoluted with exceptions, or if too many consensus judgments were ruled

wrong - and the reductive project could not be repaired, then we would have grounds for abandoning the hypothesis that there is such a theory.

Utilitarianism presents a reductive algorithm of this sort. One might take the abhorrent particular consequences of utilitarian theories, e.g. the legitimacy of sacrificing innocent individuals to the greater good, as evidence that happiness maximisation is not the descriptive base upon which ongoing moral consensus is founded.

More generally, the failure of ethicists, even within a single tradition, to come up with a set of rules etc that coheres with the range of our settled moral intuitions suggests that there is no systematic reduction of moral judgements to descriptive ones.

This is of course consistent with the view of particularists like Dancy (1993), who give epistemic and metaphysical priority to the particular moral judgment.

It is true that a reductionist interpretation of moral judgements need only have local application; for it would be seeking to explain local behaviour. Indeed, such a position is likely to be motivated by a view that the application of moral terms is learnt through local custom and cultural indoctrination. Inconsistencies between ethical systems would not require explanation. Nevertheless systems of moral rules have failed to deliver intuitive results even within singular moral traditions. (At least this seems true of both current Western mores, and of traditional Jewish morality - within which there is latitude to suspend the rules to fit the Rabbinic authorities' understanding of the exigencies of the moment.)

Although the reductivist does not have to explain disagreements between communities, consistency between moral judgments across different communities would need further explanation. I have argued that there is such consensus (at least at the level of the particular). If different reductive theories supported the moralities of the two communities, but both agreed at the particular level, then that would be grounds for doubting the priority of the



reductive theories. Suppose a religious community and a utilitarian community both agree that particular cases of almsgiving are laudable. The reductivist claims that "right" is reducible to "what is implicit or explicit in revelation", or to "what is utility maximising", respectively. The agreement of this and other communities about a range of examples of almsgiving then becomes a coincidence begging explanation, an explanation that might involve reference to the fact that almsgiving is in fact laudable.

Of course, consensus about moral facts across different communities might be a consequence of their sharing a reductive theory. But in this case too, the coincidence of different communities happening upon the same sets of non-moral properties to bear their approval would beg an explanation. What might this be? Suppose utilitarianism were the universal metric for determining how to act. Mightn't we conclude that it was true? Reductivism is thus compatible with realism.

We would avoid the realist conclusion in these circumstances only if some other explanation were available for the prevalence of the reductive scheme. In the case of utilitarianism an evolutionary explanation might be credible. Communities thrive if they attempt to maximise utility. Communities that thrive, expand and are emulated. None of this would prove that utilitarianism was true.

However, were there enduring agreement on a set of moral judgments that were not obviously conducive to survival, the evolutionary hypothesis would be less plausible, compared to the hypothesis that agreement is explained by the truth of the moral claims inspiring agreement.

The proponent of the view that all moral judgments are founded upon a descriptive base is not forced to claim with the reductivist that there is a system of laws or generalisations that relate moral statements to descriptive counterparts. In principle, she could hypothesise that there is some prescriptively inert feature of the world that is picked out by each moral term, a feature that may

defy definition (as so many phenomena do) but which is nonetheless the basis for the attribution of moral terms, explaining continuing agreement in their use.

The consequence would be that even the amoral could learn how to use moral terms as the morally committed do, but with this difference: always they would maintain their distance from the moral commitments that others would claim to be implicit in the moral judgments.

Someone claims that she understands moral terms, and indeed she uses moral terms and makes moral evaluations as we do. Yet she claims not to be committed, and indeed she acts immorally. If such a case occurred, would we be forced to acknowledge a descriptive base to our moral claims? We might instead claim that she was lying: that she did feel the prescriptive force of her moral judgments, merely choosing to ignore it. Wittgenstein tells us that "it is our *acting*, which lies at the bottom of the language-game" (1969, #204) but in this case moral backsliding is part of the game too.

Suppose someone were to claim that they were colour-blind but that they could nonetheless discriminate colours from their visual texture and their feel. Would we not suspect that they were not colour-blind at all - that they were merely deluding us (and perhaps themselves)?

In both cases the subject would be claiming that they could perceive distinctions without the phenomenological content that we consider central to our ability to discriminate: in the one case the action-guiding character of moral judgment<sup>62</sup>, in

---

<sup>62</sup> Mackie's concern with the strangeness of this character of moral judgments, his "argument from queerness" is best addressed by Platts' (1980) contention that "The queerest thing about this as it stands is the claim that it is an *argument*. ... I find neutrinos, aardvarks, infinite sequences of objects and most pertinently impressionistic paintings peculiar kinds of entities; but I do not expect nuclear physics, zoology, formal semantics or art history to pay much regard to that." We might add that from a Wittgensteinian perspective, with its focus upon the use to which we put language, it is the judgments that are not action-guiding that should be regarded as peculiar.

the other the phenomenon of colour in all perception. In both cases, we would surely require some independent verification of the existence of a different explanatory base: in the case of morality we have seen that such an explanation<sup>63</sup>, requiring a reduction of moral to non-moral terms, is lacking.

## **6G The Possibility of Convergence of Moral Theories**

Notwithstanding the primacy of the particular judgment, there are two reasons for seeking convergence upon a moral theory, i.e. upon generalisations that explain our particular moral judgments. A general theory would

- provide reassurance that our particular judgments are picking out genuine features of actions
  
- help us to decide hard cases.

Theory fulfils precisely these roles in science. In addition, in science, theory yields predictions, that in turn act as a test of theory. This self-correcting mechanism steers theory towards truth, and leads (in principle) to convergence of theories from different starting points.

Is there a parallel process in ethics? Williams:

"If a wider objectivity were to come from all this, then the reflective ethical considerations would themselves have to be objective. ... If this is understood

---

<sup>63</sup> For individuals, we would require a reductive account of how they are able to use moral categories though lacking moral sensibility. However, the sceptic's case might be a general one: that we all really have only descriptive judgments about the world which we kid ourselves into thinking imperative. Here too an explanation would be required - as for any other claim of mass delusion. In the next subsection, a possible global evolutionary explanation for our mistaking descriptive judgments for imperatives is considered.

as our coming to have propositional knowledge of ethical truths, then we need some account of what "tracking the truth" will be. ... If it is construed as convergence on a body of ethical truths which is brought about and explained by the fact that they are truths - this would be the strict analogy to scientific objectivity - then I see no hope for it. In particular, there is no hope of extending to this level the kind of world-guidedness we have been considering in the case of the thick ethical concepts. Discussions at the reflective level, if they have the ambition of considering all ethical experience and arriving at the truth about the ethical, will necessarily use the most general and abstract ethical concepts such as "right", and those concepts do not display world guidedness." (Williams 1985, Chapter 8)

I have argued above that there is probably sufficient convergence in the particular and concrete use of "right" and "good" for us to be sure that they are "world guided". If this is accepted it is not difficult to imagine how ethical theorising could converge.

A theory might contain the moral maxim that "honesty is the best policy in personal relations". The decision procedure that underpins the objectivity of this claim would involve testing it against particular cases (as well as for coherence with other generalisations). Should someone who has had an affair confess to her husband? The maxim would imply that she should. But experience - particular moral judgments exciting consensus - might show that in such cases we judge otherwise. The theory will need modification: we might say that the simple honesty maxim has to be replaced by a judgment whether the damage done both to the relationship and to the institution of marriage outweighs the general value that we accord to honesty, and the particular therapeutic and cathartic value of confession.

Furthermore, moral theories often themselves specify some form of general test of success, and allow for modification in the light of results. So, for example, if a general aim of an ethical theory is the creation of and sustaining of a well ordered, happy, productive society (a high level ethical desideratum akin to the

pursuit of simplicity and predictive power in scientific theories<sup>64</sup>), then the failure to generate such a society must reflect back upon the detailed moral code being followed - suggesting the need for modification. The parallel with the development of scientific theory, particularly in the light of modern philosophy of science (from Kuhn through Lakatos), seems very close.

At bottom Williams' argument is question-begging, for the key question in deciding whether theories about the rightness of actions are objective knowledge is whether the individual judgments that they summarize are world-guided (whether they "track the truth"); but that moral judgments are not world-guided Williams simply asserts.

My claim that we can imagine what it would be like to see such convergence is founded upon the claim that there is objectivity amongst singular evaluative statements: and that ongoing consensus is evidence that these judgments are world guided. Convergence would involve moral theories explaining increasing varieties of singular judgments, with the consequence that an increasing consensus would attach to the moral theories as well as to the individual judgments.

Williams overlooks the consensus that exists in evaluating individual actions, and focuses instead upon the diversity of moral theories. But disagreement even over theory should not be exaggerated: moral theories can diverge in complementary (as well as in inconsistent) ways, eg in support of different, but equally valid, ways of communal flourishing. In such cases apparently conflicting theorems may each be objectively true for the different communities - in the same way that apparently conflicting laws can consistently be obligatory on different communities (e.g. right-hand drive; left-hand drive laws). MacIntyre's own moral theory, which sees the good for man as defined by his own social environment (and in particular by the practices within that environment that carry benefits internal to their pursuit),

---

<sup>64</sup> Desiderata which in both cases are taken as consistent with and secondary to the pursuit of truth.

would quite clearly allow different actions to be correct in different communities without inconsistency. And the notion of pluralism in morality - without relativism - is not new with MacIntyre - stretching back at least to Herder (see Berlin 1976).

Convergence that is driven by prediction of correct particular judgments assumes that collectively we are concerned to arrive at a true theory - one that underpins true particular judgments and decides hard cases correctly. In the case of science, this concern is built into evolutionary explanations of the development of scientific ideas - for those who seize upon correct scientific theories are likely to prosper. Above we referred briefly to hypothetical evolutionary explanations of a spread of utilitarianism. In that example one would expect convergence but convergence would not indicate truth - for the reason that utilitarian societies prospered would not relate to the truth or otherwise of the morality. In general, only if evolutionary fitness depends upon the truth of belief will we have reason to hope that convergence will be upon truth. This same caveat applies in the case of science, too: it could be that societies prosper that have an incorrect but nevertheless successful science. (Was Newtonian physics successful because it approximated to truth, or because it deviated from it in a way that made it usable?)

Such doubts are global and intractable. They rest upon hypotheses regarding the nature of evolutionary fitness that are hard even to state in a form susceptible to proof: hypotheses that societies systematically prone to moral or scientific error might do better, and that we are hence in the grip of systematic self delusion regarding the nature of our decision criteria, and perhaps the cogency of our conclusions for action. What can be said is that it is not obvious that the singular moral judgments upon which there is consensus are those that are most conducive to evolutionary success. (Perhaps a utilitarian morality **would** meet that test; in which case the failure of utilitarianism to explain our judgments is a proof that moral judgments are founded upon observation of moral truth rather than upon evolutionary fit.)

The alternative to scepticism is to take our commitment to truth as primitive and to assume in both the moral and the scientific sphere that fit with singular veridical observation has been what has marked theories as worthy of adoption and promulgation. If this is right, then were convergence to occur, it would imply convergence on truth.

Dancy (1993, p.68) has it that "ethics is empirical. Moral principles are learnt in and from particular cases." Meta-ethics is also empirical: where we observe enduring consensus in the application of moral terms regarding particular cases, we can conclude to the truth not only of the individual judgments but also of the moral principles upon which, guided by the singular judgments, we converge.

## **6H Implications for Partialism**

Williams (1985) claims that in matters of morality we have reached a "reflective" stage of analysis in which society is so aware of conflicting moral systems that it is unable to consider those with opposing moral views to be just wrong. Williams thus dismisses the possibility of ethical knowledge:

"No doubt there are some ethical beliefs, universally held and usually vague ("one has to have a special reason to kill someone"), that we can be sure will survive at the reflective level. But they fall far short of any adequate, still less systematic, body of ethical knowledge at that level."

Implicit in this reckoning is the assumption that epistemic privilege is accorded to generalisation. The conclusion of our discussion of meaning theory is that there is no reason to assume that ethical knowledge will differ fundamentally from other forms of knowledge. In which case, absolute epistemic privilege cannot be accorded to any statements; whilst relative priority must be accorded to the singular judgments, the observation statements whose truth conditions lie in the mere recognition that they obtain.

Hence we will not overturn observation statements, or our conviction in their objectivity, upon discovery that other societies, sharing those observations, have different theories to explain them.

Thus it is not true that all that survives reflection and comparisons between modern societies (and including some traditional ones), are vague and general moral theorems. The body of ethical knowledge is vast - and consists mostly of precise and specific judgments that this particular action is right and that is wrong, that this is a good result and that a bad one. Many such judgments survive reflection and inter-societal comparisons; I hypothesise that consensus is too common to be explained without appeal to their truth.

We came to discuss moral epistemology in order to assess the claim of the neutralists and sceptics amongst political philosophers that the sheer diversity of conceptions of the good obliges political institutions to allow or even to encourage each to pursue her own conception of the good, whilst all must respect the rights of others to follow their own star.

The foregoing discussion of moral objectivism has I hope served to weaken the appeal of this simplistic distinction between the right and the good. Moral truth is to be uncovered piecemeal by considering the extent of consensus regarding first the objectivity of particular judgments, i.e. whether or not they represent core applications of the relevant moral terms, and second their truth.

How do obligations of partiality stack up against these empirical tests. Should Jack be loyal to Jill? This of course depends upon the circumstances. If Jack is Jill's brother, and she is in need, there would be consensus both that this is a core case of loyalty and that (assuming there are no complicating factors) Jack ought to succour her: a moral fact.

Political institutions can recognise such obligations, rather than pretending that it is purely a matter for Jack to decide if he wants to make a commitment to Jill. If



Jack's obligation of loyalty is objective fact, it can and should be taken into account in designing the structure of law and incentives.

And society does of course recognise and enforce familial obligations from parent to child and between spouses. We evidently do not consider it a matter of indifference whether a woman's projects include concern for her own offspring. Recognition of wider partialist obligations is also widespread - though rarely do we think it politic to enforce them with negative sanctions.

As discussed above, to discover the extent of such obligations would require empirical work. How much consensus across societies is there regarding the extent of partialist obligations? Does it survive full understanding of relevant facts in particular cases?

The reliance upon consensus in moral epistemology does not distinguish moral questions from descriptive questions. If we wish to discover whether a particular chemical increases the smoothness of whisky, we will have to establish first whether there is objectivity in this matter. Is there consensus (at least amongst those whose consistency in judgment provides an independent confirmation of their status as cognoscenti) in judgments of smoothness in core cases? If there is, is there statistically significant correlation between judgments of smoothness and the presence of a particular ingredient? (Note that we might conclude that there is such a thing as smoothness of whisky, but one that does not correspond to any particular combination or set of combinations of chemical ingredients: just as we might conclude that justice or cruelty cannot be reduced to a set of descriptions.)

For many partialist moral judgements, including, arguably, the existence of obligations to vote or to pay taxes, objective truth may seem unimpaired by the lack of consensus amongst political philosophers.

Similarly, the objectivity of the question whether Cyril should pay his taxes will depend upon consensus regarding the decision procedures for determining

whether paying taxes is a universal obligation. This may in turn depend upon theorems relating tax-obligations to the representativeness and benignity of government, and to theorems establishing the dependence of order and the preservation of society upon obedience to the law and contributions to the Treasury. Ultimately all these decision procedures for determining truth and falsity of the hierarchy of theorems will be underpinned by statements that approach bare truth or bare falsity (in Dummett's locution): "such and such a contribution would constitute paying a tax"; "such and such an anarchic state of affairs (the hypothesised consequence of non-taxpaying) would be intrinsically bad".

I have suggested that many evaluative statements approximate to bare truth, or are closely supported by bare truths. But many others are embedded more deeply in the web of belief, dependent for support on other warranted statements extending both inwards and outwards. Such statements are less exposed to recalcitrant experience, and by the same token less susceptible to confirmation in their objective status by empirical investigation into continuing consensus in new applications in like circumstances.

Nevertheless, if we uncover a nexus of particular moral judgments, all endorsing acts of partiality, we may seek an explanation in the shape of an understanding that independent value resides in the flourishing of groups and relationships. This is the conclusion to which we moved in Chapter Four. I hope the foregoing discussion provides an indication of the sort of research programme that would provide empirical justification for such a theory.

## 7 PARTIALITY TOWARDS LARGE GROUPS

In this chapter I explore the peculiar difficulties associated with establishing obligations of loyalty to large groups, absent an explicit social contract, given that an individual's actions in support of the larger group appear to have negligible influence upon the group or upon its projects.

I suggest (in Section 7A) that individuals' actions in a social context are best understood analogously to the way a single person understands a string of her own actions that will collectively issue in a desired result, but none of which separately appears significant. It is our epistemic dependence upon the group (discussed in Chapter Six), that underpins the conclusion (i) that the rightness of an action should be assessed within its social context, just as an individual views her acts within the context of her plans over time, and therefore that individuals are bound by obligations of loyalty to large groups.

This analogy also helps to explain why the contribution of single actions to a social end seems insignificant. One explanation is (ii) that the (questionably rational) non-linearity of risk aversion leads us to understate the cost of taking small risks. Another is that (iii) the non-linearity of the function from individuals' incremental contributions to a project to the resulting collective benefit ensures that only in a collective context do individual actions achieve significance.

However, the strength of the resulting obligations to groups is dependent upon two further questions: the value of the larger group and of its projects; and the importance of acts of loyalty to sustaining it and them. The members of a group must be careful to determine obligations of loyalty so as to minimise the occasions on which loyalty is superfluous. (Subsections 7B and 7C).

The nature of the intrinsic value embodied in the group and of the obligations of individuals to the group will also determine the scope of group obligations - that is membership criteria, and criteria of entry to and exit from the group (Subsection 7D).

I conclude (subsection 7E) with some observations on the persistent tensions between partialist and impartialist perspectives.

## **7A The Surprising Significance of Individual Acts of Loyalty to Large Groups**

In Chapter Four we established the intrinsic value of good groups (in addition to their obvious instrumental value), groups being constituted of individuals bound to each other by historical connexions. The recognition of good groups as meriting support explains how partialist practices can compromise fairness legitimately. Individuals in such groups can be justified in favouring other members of the group, providing that the groups concerned are amongst those with intrinsic value, because such partiality is constitutive of the flourishing of such groups, their flourishing being an aspect of their intrinsic value.

I suggested (in 4Cv) that recognition of this intrinsic value must underlie the associative obligations that Dworkin (1986), for example, acknowledges in both personal and political life. And it presents a partial answer to the first of the two classical questions that motivate Dworkin's consideration of these issues in political philosophy:

"whether the state is morally legitimate, in the sense that it is justified in using force against its citizens, and whether the state's decisions impose genuine obligations on them" (Dworkin 1986, p218)

For a state that qualifies as a good group, the use of force to preserve itself or to strengthen itself can be justified even in the absence of consequential gains, either to its members or to others, notwithstanding the unfairness of its actions or the compromises to other goods and rights (including the compromise with fairness). This legitimacy follows immediately from the recognition of the intrinsic value of the group.

However, the second question - whether the state's decisions (and indeed its existence as a group of intrinsic value) create effective obligations on citizens, in particular obligations strong enough to override impartialist demands upon the individual, remains unanswered<sup>65</sup>. Although recognition of the intrinsic value of good groups and relationships provides justification for many actions of partiality towards friends and relatives, given that the benefit to the family or the friendship is often tangible and often exceeds the direct and opportunity costs in fairness or other goods, the same is rarely obvious of acts of partiality towards populous groups. (Nevertheless, *pace* Simmons (1996)<sup>66</sup>, to the extent that partial actions do contribute to the good of the state, they are *ipso facto* justified.)

Large groups rarely appear to benefit significantly from actions of individuals, nor do the collective projects of states appear much affected by individuals' failure to participate; yet the direct and opportunity cost involved in actions of loyalty to the state may be tangible and significant, and may include failure to achieve other goods, like happiness or fairness. Hence the Dworkin argument, that associative obligations can in some circumstances extend to the wider political community, hits a sort of free rider objection: the group's flourishing will not be perceptibly diminished for the lack of my cooperation, so I will rarely be justified in favouring the group.

---

<sup>65</sup> The notion of an "effective" obligation needs explication. In a world of competing ends and constraints, almost every obligation is defeasible. But the insignificance of the individual within the large groups ensures that communal obligations will almost always appear to be trumped by some more pressing personal commitment. If that is right, such communal obligations, even if recognised as obligations, are ineffective.

<sup>66</sup> Simmons questions both the Burkean claim that we are bound by a "permanent standing contract" to whichever state we happen to be borne into, and communitarian suggestions that our identity or our morality depends upon obedience. However, he does not address the possibility that the local community embodies a good that its members are in a unique position to foster. I suggest that this fact is itself sufficient to create an effective obligation, subject to the discussion in the text.

Some examples:

- The gain in both happiness and fairness achieved by an individual giving money to the indigent rather than paying taxes, may exceed the cost to the group of this one flouting of the law. There is some risk that their marginal tax payment, rather than enabling government to undertake a worthy project that would otherwise go unfunded, will support excess administration or a less desirable programme.
- The gain in happiness and fairness from visiting the sick or the lonely on election day may exceed the cost to the group of one vote less. The chances of an individual vote determining an election result, or securing the number of votes needed to establish a government's democratic legitimacy are vanishingly small.
- Even in a just war, the grounds, utilitarian or otherwise, that justified the waging of the war will not appear to justify individual acts of killing: the certainty of the loss and distress caused by the killing of the enemy soldier will appear to outweigh the tiny chance that this action will secure significant benefit to the war effort.

Similar arguments will apply to most other laws stipulating duties to the state; although universal non-obedience would be crippling to the state, individual law-breaking may often appear likely to do more good than harm.

Insignificance may be illusory - akin to the sorites paradox: the contribution of a single pace or grain of sand is obscured by the length of the journey to be traversed or by the scale of the heap to be created. But there is often a genuine paradox. The contribution of the individual will rightly be seen as highly likely to be either inadequate or superfluous: the rest of the group will either act to secure the end without the individual's participation, or they will not, in which case the

individual's contribution would be vain. The chances of an individual's contribution proving significant seem negligible.

Only in exceptional cases, isomorphic to the construction of a pyramid in which each individual must lay a brick, will an individual be able confidently to assert that her contribution was vital to the realisation of the project. Even in such rare cases of proportional contribution, individuals may need assurance that the rest of the group will play their part. Although this assurance can in some cases be provided by the force of law (as we discuss below), in other cases, we will need a moral mechanism to force the individual to view her action in the context of the group.

Hence, it seems that in order to justify individual contributions to collective projects, we need to justify loyalty. This loyalty must generate an obligation to obey that is conditioned, like the assessment of the right course for the group, by the benefits accruing from general conformity to the law, and that is therefore blind to the minimal benefits arising from the individual's acts or to the possibility that others will not act in support.

Raz (1994, Chapter 15) argues that the difference that the law makes to one's moral obligations can in most cases be explained without recourse to "a general obligation to obey". Only just laws should be enacted, and such laws should be obeyed because they are just, not because they are legislated. (It is nonetheless right that they be legislated with appropriate sanctions -- in order to motivate the weaker members of the community to act well.)

Raz accepts that some laws alter one's moral obligations. In particular, the government may use laws to generate benefits of coordination for its citizens. Raz uses the example of a law forbidding barbecues in certain beauty spots: without the law personal restraint would give little or no benefit; with the law, collective restraint will preserve the beauty spot. Nonetheless, he argues that it is the fact that the law is in existence, rather than an obligation to obey the law, that changes one's moral duty:

"I would not have had any reason to avoid having barbecues in beauty spots ..., but for the introduction of the law which gives rise to the expectation that the widespread but damaging practice will come to an end, or at least that it will be sufficiently reduced so that my self-restraint will make a difference, however little." (Raz, 1994, p.348)

This argument would be valid only if the barbecue case is one of **proportional contribution**. That is, if individual acts of restraint make a proportional contribution to the recognized environmental benefit sought. This is not plausible, however: once sufficient other people are obeying the law to render it effective, my own contribution will probably be as nugatory as it was when everyone was despoiling the countryside, for it is likely that the environment can recover from occasional depredations. The "however little" of the benefit of the individual's restraint is likely to be insufficient to justify it against the pleasure that would accrue to myself and to others from infringing the law; as a result, absent an **obligation to obey the law**, few will be obliged to conform to it.

A general argument from the benefits of the state will not provide the requisite obligations upon individuals. Consider, for example, a utilitarian justification of the state. Wolff (1996) boils such a justification down to a simple deduction:

- "1. The morally best society is the one in which happiness is maximized.
2. The state promotes happiness better than the state of nature.
3. The state and the state of nature are the only alternatives we have.

Therefore

4. We have a moral duty to bring about and support the state."

Wolff goes on to discuss how the different premisses can be questioned. However, I wish only to note that the conclusion is couched correctly in the plural. Hence a subsequent deduction from "we" to "I" is required; but this is problematic. **We** may have an obligation to support the state in circumstances in which **our** support will be effective, but it does not immediately follow that **I** must support the state given **my** genuine doubts regarding the efficacy of my support.



Thus, justifying partiality towards the state is more of a challenge than justifying partiality towards friends, relatives and lovers, as in the latter cases individual actions of partiality tangibly confer benefit and hence constitute the flourishing of the valued relationships.

To fill this gap, political philosophers have often posited some form of social contract, binding all members of a group to abide by group rules notwithstanding their apparent sub-optimality in a particular case.

Would a social contract do the task set for it?

If we assume a pro tanto obligation to keep contracts (as discussed in Chapter Four), then any contract to pay taxes or otherwise to be loyal to the state would help to justify loyalty to the state. However, a general contractual obligation to keep the law would be a blunt instrument. If what binds an individual to keep the law is simply the existence of the contract, it might apply equally to all laws irrespective of their importance. Defeasible it will still be, if the cost of obedience is too high, for the obligation to keep contracts is itself defeasible. Yet, not only is it difficult to determine when the cost would be too high (i.e. how much disutility to suffer in order to fulfil contractual obligations), but less tractably still, the contractual obligation apparently does not distinguish between more and less important acts of obedience.

Consider two of the examples cited above: obligations upon a soldier and upon a taxpayer. We would probably wish a soldier to take his obligations of obedience more seriously than a taxpayer, as more is at stake. Yet grounding obligations in contract allows no distinction between the cases. Part of what is required of a credible theory of political obligation is discrimination between duties to the state that allows them to be overridden in appropriate circumstances.

In any case, as often remarked, we have no contract, only a *quasi-contract*, which cannot serve to bind citizens at all. It is true that the quasi-contractualist can retort

to the free-rider -- what if everyone did that? -- for the quasi contract is precisely designed to optimise the rules for the group, the rules by which most people must abide to ensure societal stability. But this rhetorical question directed at the free-rider is weak in the absence of the mechanism we are seeking, a mechanism to force the individual to view her act in a social context; for she can merely assert that she knows that everyone won't opt out, so it would be wrong of her to let that hypothesis influence her optimising calculus. And in any case, if they all did opt out, her action would make no difference.

Implicit or tacit consent, which may perhaps be imputed in most western societies to most of their members, will not do either -- for we need an argument to overturn the assumption that tacit consent to abide by laws and partialist duties is contingent upon there being no alternative action by which more good will be achieved.

Perhaps **obligations of reciprocity** will serve to bind individuals, or at least to encourage them, to keep the law, even in the face of apparently more cogent conflicting demands. In Chapter Four (section B: Debts of Reciprocity) we concluded that there were obligations of reciprocation, and that these went some way towards explaining loyalty in some circumstances. We also saw, however, that reciprocation is ill-suited to provide the base for obligations towards the state: the cogency of such obligations is not appropriately proportionate to benefits received. Nevertheless it is instructive to consider whether reciprocation could justify the unconditional loyalty of an individual who has received benefits.

On the basis of services received, someone might favour the state even if she knew that her action was unnecessary to the state's ends, and that her effort could be better spent. For it might be claimed that the obligation to the state engendered by an individual's receipt of benefits from the state is often sufficient to outweigh the general obligation (*ceteris paribus*) to maximise happiness, or otherwise to pursue the good.

Simmons (1979, p.186) considers such an argument only to dismiss it with a quote from M.B.E.Smith claiming that "when a person owes a debt of gratitude towards another, he does not necessarily acquire a prima facie obligation to display his gratitude in the most convincing manner"<sup>67</sup>.

Simmons notes Nannerl Henry's response that "The reason obedience is the coin in which this particular debt must be paid is that political services cannot be provided unless subjects obey their governments."<sup>68</sup>, but he rejects the conclusion by pressing the analogy with gratitude to an individual:

"If Jones was a benefactor to whom I owe a debt of gratitude, and Jones needs to be plugged into a kidney machine all day in order to stay alive, it does not follow that I can discharge my debt to him only by locating, operating or paying for kidney machines."

Simmons here confuses the question of the extent of the debt with the question of the kind in which repayment should be made. Jones's generosity may not have been sufficient to merit such sacrifice. But whatever the extent of the debt, it would surely be perverse, or worse, to repay it in any other way than that most needed by Jones. Hence Henry is surely right that any debt to state should be repaid in the coin most needed by the state - and a case could be made that that is collective obedience.

However, this does not overcome the free rider problem from the individual's perspective. For the analogy with an individual is not compelling: in the case of the state there is little genuine benefit conferred. There would be no point in giving Jones a bit for a kidney machine if you know that none would provide him with the rest; and even if Jones asked for the bit, your obligation to give it him

---

<sup>67</sup> Simmons cites Smith's paper "Is there a Prima Facie Obligation to Obey the Law?" Yale Law Journal 1973

<sup>68</sup> Quoted from "Political Obligation and Collective Goods" in Pennock and Chapman (eds.) Nomos XII: Political and Legal Obligation, Atherton, 1970

must be greatly mitigated by your knowledge that it would in practice be useless to him. Similarly, there is no point in voting if you know that insufficient others will vote to achieve democratic viability (or to achieve a change in government, if that is required). Conversely, you would also be relieved of any obligation to provide a bit for Jones' kidney machine, even if he asks for it, if you know that someone else is providing the complete machine. Why vote if your vote will be superfluous?

In such cases you would surely be right in using your discretion to give to Jones what you reckon he most needs. Similarly, an individualist acknowledging a debt to society, could discharge it by judging for herself whether obedience is the appropriate coin, or whether some other mode of bounty is more beneficial to the state than her marginal contribution to the vote or to tax revenue or to a war effort.

What if it is argued that a sense of fair play would require that you contributed in the same way as your fellows? If you have eaten with friends, and it is time to do the washing up, you might claim that more utility would accrue from your going to the cinema whilst they wash up. Such an attitude, even if based upon a fair assessment of relative utility, might breed resentment. Even if you found an altruistic occupation, or one that contributed to the welfare of your circle of friends as much as washing up, your actions might still be condemned: you would be failing "to pull your weight" as you should.

There is a parallel with attitudes towards someone who wishes to contribute in civilian life rather than to fight on the frontline in a just war. The complaint is not that there is inequality of benefit, but rather that there is inequality of cost: we are all expected to make equal sacrifice for the common good, and doing the same as the others is one way of ensuring such equality.

However, there is a limit to the number of people who can do the washing up at once: in circumstances of sufficiency, discretion to find a substitute contribution to the group, like sweeping the floor, may be acceptable, so long as there is

equal cost. Someone might refuse to kill because she reckons the costs of her actions as a soldier will outweigh the benefits. She will not be condemned for failing to pull her weight so long as the alternative contribution that she makes puts her at equal risk with her fellows; she might perhaps join the frontline medical corps. (There will however remain the issue of whether she is sharing the moral and psychological cost involved in necessary killing; what would match this sacrifice?)

Considerations of reciprocity, of fair play, do therefore demand some contribution, perhaps in kind, whilst others vote or fight (and this contribution should be for the benefit of the group as a whole and at a cost to oneself similar to others' sacrifice). However, these considerations will not force one to vote, to pay taxes or to fight when one is confident that these activities are nugatory.

How then is loyalty to the group sufficient to obligate participation in collective action to be justified?

### **i      Assessing Single Acts in Collective Contexts: Smoking and Voting**

How is it that the individual act of obedience can appear so insignificant that it can scarcely trump any alternative good deed, whilst the joint obedience of the many can generate sufficient benefit to justify significant collective sacrifice? We may judge that the benefits of at least a forty per cent electoral turn-out (in empowering the people and legitimating government) exceeds the cost of the time and effort of all those voting; yet for an individual amongst them there appears virtually no benefit<sup>69</sup> to set against the opportunity cost.

---

<sup>69</sup> There is pleasure in exercising the rights of citizenship, but this pleasure would be undermined by awareness of the vanity of the individual vote - were it not for the recognition of the obligations of loyalty which we are here attempting to justify.

Consider an analogy with a series of actions by an individual. Take someone who is considering giving up smoking. She may argue that one extra cigarette will do little harm: the potential harm is outweighed by the pleasure enjoyed. Yet if she considers say a dozen years of smoking a packet a day, and the consequential damage to her life expectancy, she may well determine to give up.

How does the quitter quit? She sees that every cigarette is part of a single chain. She makes a single decision to cease smoking. She turns down every subsequent offer of a cigarette not with a separate decision but as part of the same decision.

It is not that she considers the act of smoking a single cigarette as a different act when it is under a different description. That would be sophistical. Smoking a cigarette is the token act under consideration, an act individuated by its causes and consequences<sup>70</sup>.

What distinguishes the quitter is that her *decision* relates to the next dozen years of her life rather than merely to the next cigarette. The decision itself then makes a difference: the current act of restraint has different causes and different consequences than it would have had without the decision. It is caused in part by the decision and contributes to the causation of similar acts of restraint extending over future years, so that each is part of the greater whole that will win the smoker significantly greater life chances. The current act of restraint is the embodiment of a decision that governs not only this act but the agent's life over the next twelve years. It is therefore to the decision, and hence to the acts governed by the decision, that rationality can be predicated or withheld. Similar bodily movements can be rational as an act implementing one decision (e.g. to give up smoking), and irrational as an act implementing another (e.g. to deny herself this cigarette).

Our ex-smoker has not curtailed her free will by her long term decision: she could decide to smoke again. But by deciding to give up for a dozen years, she has

---

<sup>70</sup> See Davidson (1969) on the individuation of actions.

transformed a future choice to smoke again from a balancing of the probabilities of harm and pleasure arising from that cigarette to a choice as to whether to vitiate her own earlier decision. Such a decision might constitute an indictable failure of will. Whatever our view of the initial decision to give up smoking (whether we view it as morally<sup>71</sup> imperative or not), we may conclude that the weakness that would be instantiated by subsequently taking up smoking again would be reprehensible. In such a case, the cost of the cigarette might well exceed the pleasure accruing. The cost would comprise: a) the small damage to life chances that it would inflict; b) the increased likelihood of future recidivism and associated health costs; c) the intrinsic evil inherent in the failure of the will. The act would rightly be characterised in a way that linked it to the earlier acts and decisions such as to invest it with greater significance, akin to that of a broken promise.

Yet the quitter's commitment remains defeasible. She has not bound herself never to smoke regardless of the balance of costs and benefits. For example, if her life expectancy is foreshortened from some other cause, then she can change her assessment of the balance of pleasure against the disutility of the risk incurred by a dozen years' smoking, and decide to smoke again now without deserving censure.

Turn from the individual considering the consequences of a string of her own acts into the future, to the members of a large group considering a set of acts to be executed at one time.

For the group at election time, too, the risk of calamity can only be averted and the prize of a flourishing democracy can only be secured, if the group is able to make a collective decision encompassing all the individual acts of voting, a decision that views all the separate votes as part of a collective act, that of sustaining democracy through voting.

---

<sup>71</sup> "Moral" connotes "all things considered". Selfish or prudential or hedonistic acts can be moral in this sense if they are right, all things considered.

Such a decision might result from a collective decision making process, such as a democratic process in which everyone does participate, and in which each individually and hence the community collectively determines to vote in all future elections.

The collective decision would then affect all the individual actions of the group members, not by curtailing their free will, not necessarily by imposing any sanction upon them, but merely by changing the nature of their individual decisions and the nature of the act of voting. No longer is it merely a decision to vote, it is now a decision to sustain the collective decision. Each vote now represents a part of the collective decision to vote and must be morally evaluated in that light, just as the individual acts of restraint of the quitter have to be seen as part of the chain of actions that will produce a healthier life.

Failing to vote would now have additional costs. On top of (a) the tiny direct impact cost to sustaining democratic process, would be (b) the chance of emulation in this and future elections, and (c) the vitiation of the collective agreement. Alternative pursuits would no longer be justified for a passing gain, any more than taking up smoking again can be rationally entertained, by someone who has quit, merely for the sake of a few moments pleasure. In both cases the single act of restraint or of voting is viewed as a part of a larger project; the cigarette or the failure to vote would flout the overarching decision and risk further and complete failure; and the moral stakes are accordingly raised.

What if there is no explicit decision? For the community lacking a decision, but nonetheless standing before a project that can only succeed if a substantial number participate, the problem is like a prisoners' dilemma: cooperation is required to maximise benefits for all. An agreement across the voters that they will all act in the cooperative manner to secure the best outcome may not be feasible. (In any case such an agreement would share the other drawback of the social contract noted above: compliance would not obviously admit of proportional defeasibility.)



However, the integration of a set of individual acts into a collective act need not be the result of a singular fully-participative collective decision or explicit social contract. Rather the convention of cooperation may emerge over time, from whatever source. Prisoners' dilemmas can be solved for repeated games by the emergence of a pattern of cooperation that rationally affects expectations.

Consider the ex-smoker again: if she takes up smoking on inadequate grounds, she will be at fault for breaking her resolve. At this stage, it is not so much the original decision that transforms the nature of her current action, it is the actual history of non-smoking. She might be reckoned at fault even if she gave up earlier not as a result of a positive decision, but merely as a result of lack of access to cigarettes; her action now might nonetheless be regarded as that of an ex-smoker resuming a habit she has quit rather than as that of a smoker having one more cigarette.

So also the community may be correct, given for example a history of participative democracy, to judge that its members ought now to vote (absent significant cost to the individual in so doing, on which more below), even if there has been no explicit collective decision to institute elections.

Rather, there is a collective judgment that voting is required, that each vote is a part of a greater whole. Each must do his part, must vote, if he accepts this collective judgment. The community now jointly recognises or judges that a certain form of behaviour is correct.

What if an individual distances himself from the collective judgment. Is he not justified by his own lights?

The argument to objectivity from continuing consensus developed in Chapter Six here comes into play. The individual is not in a position simply to dismiss the judgment of the group, particularly a collective judgment composed of myriad independent individual judgments. The best explanation of that coincidence of

individual judgments, judgments that one should vote if the cost of so doing is within certain limits, is that those judgments are correct, that in those circumstances there is an on-balance obligation (a moral obligation) to vote.

In short: it is evidently optimistic that all citizens without compelling excuses should vote; the community judges that it is right that they should vote; I have no compelling reason not to trust the communal judgment as to the right action in this case; hence it is right that I should vote. Epistemic dependence upon the judgment of the group neatly circumvents the lack of explicit agreement, at least where there is some convention or precedent to make it reasonable to assume that people will act jointly. No further step is then required to obtain a decision to act collectively, and one that binds the individual. "What action is right", language theory dictates to the agent, "is what the community of those who share with me an understanding of the term, reckon is right."

The collective judgment cannot be dismissed, but it can be questioned. The individual citizen might take the view that the system has become corrupt, and that it is no longer worthy of being sustained, that the best explanation of coincidence of judgment is not correctness but systematic bias.

However, such a challenge cannot be defended lightly, for it is the group decision that has to be challenged rather than the consequential obligation upon the individual. What the individual voter cannot rationally do, is to challenge his obligation to vote once he recognises both the legitimacy of the group's collective judgment that voting is desirable and his membership of the community. That his contribution is negligible to the end that is sought by the community does not prevent his act from being rightly judged as falling under the ambit of the collective judgment.

Similarly, the ex-smoker might rationally decide to take up smoking again if her circumstances changed relevantly, but cannot rationally smoke again if she recognises both the wisdom of her having given up smoking and that this act would constitute a resumption.

Nevertheless, for this explanation of the obligation of the individual to be convincing, we must probe the apparent paradox that the community judges that I should vote notwithstanding that my individual vote appears of less value to the community than its opportunity cost. The parallel with the individual considering the costs and benefits of a string of actions subject to a single decision is again instructive.

Once the relationship between individual and collective action is clarified, we can explore the defeasibility of the obligations upon individuals.

## ii      The non-linearity of Risk Aversion

The expected cost of this cigarette is 6 minutes<sup>72</sup> off the smoker's life, 40 years hence. The expected benefit is 6 minutes of pleasure now. Such statistics are unlikely to persuade her to forego current pleasure.

The expected loss of life from a dozen years of smoking will be (assuming, *arguendo*, that the relationship between smoking and life expectancy is linear at two hours for every day of smoking) one year.

Is this calculation a more compelling reason for giving up? If 6 minutes immediate pleasure more than compensates for six minutes off the end of life, why should not a year's smoking pleasure, spaced out over a dozen years, compensate for a year off the end of one's life? If each cigarette's pleasure offsets its costs, how can it be right to give up?

The parallel argument regarding voting starts from the premise that if the gain from an alternative action exceeds the loss consequent upon failure to vote, one ought not to vote. And if for many individuals the social gain from alternative

---

<sup>72</sup> All statistics that follow are merely illustrative. For simplicity, I also assume a zero personal and social discount rate.

action will often enough outweigh societal loss, then why should the balance be any different when summed across the electorate? Perhaps we had all better do something more immediately and certainly gainful than voting.

Yet we do reckon both giving up smoking and voting to be sensible and often correct behaviour.

We have assumed the health cost of a cigarette to be 6 minutes off life expectancy. Let us further suppose (more realistically) that this expected six minutes represents not a certain 6 minutes life loss but rather a tiny increase (one part in 876,000) in the *probability* of losing *ten years* off your life. This too may not appear worth worrying about.

The situation is now very different when contemplating the consequences of a dozen years smoking on similar assumptions. As we calculated, a forty year old reduces her life expectancy by one year to 79 years if she determines to smoke for another dozen years. Suppose (in line with the assumption in the previous paragraph) that that "expected" 79 years decomposes into a 90 per cent chance of health till 80, and a 10 per cent chance of death at 70. A one in ten chance of losing 10 years is a statistic that might drive a change in behaviour, that might warrant some sacrifice.

That a ten per cent chance of losing ten years is reckoned a worse outcome than a certain loss of one year is a standard manifestation of risk aversion. The simplest explanation for risk aversion in general is the diminishing marginal utility of consumption. It is plausible to suppose that (at least in anticipation) not only consumption but life itself has diminishing marginal value: the value of ten years from age seventy is more than ten times the value of one year from age 79 (even if health is constant across the decade).

This model is of course simplistic in its assumption that both pleasure and risk of smoking increases linearly with the duration of the habit. The probabilities are not independent. But different dependencies work in different directions. The second

cigarette will have less impact on the probability of death than the first, for you were more likely to have been already doomed. (Agreeing to play Russian Roulette twice is not twice as risky as playing it once, as there is a one in six chance that you will be invulnerable to the second trigger-pull). On the other hand, the second cigarette may have greater impact because it impedes the body's recovery from the first one. Further the pleasure from the second cigarette may be less than that of the first. I assume that these effects balance each other out to leave the costs and benefits of additional cigarettes constant<sup>73</sup>.

In a nutshell, the smoker is willing to take (roughly) a one in a million chance of losing her ten years off her life for a small gain (six minutes of pleasure), but yet is unwilling to take a one in ten chance of loss of life for a gain some one hundred thousand times greater. Is this rational?

Suppose we know that the chances of death on crossing the road at a zebra crossing are one in a million. For how large a reward would you cross the road. Perhaps £1? The expected gain would be £0.999999 (the reward times the probability of survival) less one millionth part of the value of a life (the probability of death times the value of life). If you would cross for £1, you implicitly set the value of your life at less than £1m. This is in fact in the ball park of figures derived from observed risk taking behaviour, figures that are used in costing road safety schemes in the UK. Most would cross.

The chances of death on crossing a motorway are, say, one in three. There is a million pound note on the other side. The expected gain from crossing is thus £666,666.67 less one third of the value of a life. With a life valued at £1m, the expected gain is £333,333.34. Apparently one should cross, but most would refuse. In this case we put a far higher implicit value on life.

---

<sup>73</sup> Although as mentioned, the statistics are illustrative, they must nonetheless be plausible if the intuitions to which I appeal are not to be distorted.

It is surely rational to refuse to cross the motorway. My own intuitions are less forceful with respect to the £1 coin across the zebra crossing. Perhaps it is irrational to collect the coin after all. Perhaps we should consider that if we all cross roads regularly for pound coins, some of us will die, for an average gain (across all those who cross) of a million pounds, and none of us would be willing to sacrifice ourselves or a loved one or even a stranger for that sum (and perhaps not for any sum).

So perhaps after all it is irrational for the smoker to smoke even a single cigarette.

Yet we can say that we are only crossing the road the once. The smoker may be only smoking the one cigarette at this time, so she will say that her rationality in making the judgment about taking this tiny risk for this certain gain cannot be impugned. On this logic, the gain rationally required to justify taking a risk is not proportional to the scale of the risk.

Risk aversion is clearly non-linear. To make sense of our preparedness to take small risks with catastrophes but not larger risks despite compensating larger pay-outs, we would have to escalate the costs of failure as the risks rise by adding to the costs a term that tends towards infinity as the probability of catastrophe approaches one. The alternative is to maintain that the value of life is very high (perhaps infinite) but that we are irrational in our willingness to take small risks.

These conclusions derive not only from the contrived example given, but also from the generally observed rescue principle that sanctions extraordinary expense to save a particular threatened life.

So the smoker's dilemma is a real one (whether rational or not). She is willing to take the tiny risk that this cigarette will damage her. She is rightly unwilling to take the much larger risk that smoking for a decade will damage her. Yet repeatedly taking the tiny risk will ensure that she takes the unacceptable large risk.

The parallel dilemma for the would-be voter is this. Each person's failure to vote reduces the probability of obtaining the substantial prize of a flourishing participative legitimated inclusive democracy by a small amount, and increases the probability of political catastrophe by a small increment. In exchange, the non-voters gain the small but certain benefits of their alternative actions (cf. the pleasure from the cigarette, the pound across the road).

If three fifths of the populace fail to vote, then the threat to democracy is substantial. The scale of this threat might lead us all to agree that a large majority ideally should vote, notwithstanding the significant opportunity cost involved (the aggregated individual gains from alternative actions).

Yet each individual may judge his own contribution to reducing the risk of disaster and to increasing the chances of securing the democratic prize, to be inconsequential. Again, I am not sure whether this estimation is rational: perhaps if avoiding the threat to democracy justifies the collective sacrifice, avoiding a proportional fraction of the risk justifies the individual sacrifice.

Either way, the individual's epistemic dependence on the group should ensure that in deciding what to do, an individual will view her act in a collective context, and assess the risk accordingly.

### iii      **The Non-Linearity of the Social Gain Function**

However, for collective goods there is an additional doubt that interferes in the calculations of the would-be loyal citizen.

Imagine that the contributions of individuals to democracy from voting are scored along the x-axis, and the consequential benefits of those contributions are scored along the y-axis.

The benefits curve (the social gain function) will rise sigmoidally: at first very slowly (the first few votes establish no legitimacy) then rapidly (as democratic legitimacy is established) and then very gently again (the difference between a 90% and a 95% turn out is much less significant than the difference between say 30% and 35%).

Hence, if democracy is flourishing an individual may be confident that he is one of the superfluous excess voters. Nevertheless, as we established, the decision as to whether to vote must be viewed in a social context. The individual should judge whether the whole community should vote. In this context, his vote might as well be in the more valuable centre section as at the extreme. And the decision to vote can be made if the collective gain exceeds the collective cost.

For the community, the moral fact that it is correct for the community to undertake a particular joint project, or to act in a particular way, given the likelihood that the whole enterprise will yield or constitute something valuable, determines the moral evaluation of the relevant acts of the individuals of which the group is composed. The individuals can recognise from the coincidence of their judgments what is the right and the good for the group. Having recognised the nature of the good for the group, they are then forced to evaluate their actions as parts of the collective whole rather than in isolation.

Such an evaluation is not an optional add on, as if one could say that the individual's act (of voting, paying taxes, going to war) can be viewed either from the perspective of the state as a necessary contribution to the well being of the whole, or from the perspective of the individual as a nugatory addition to the efforts of so many others and one easily trumped by concrete gains accruing to alternate actions. Rather, if the group project is correct, then the parts stipulated for the individual are necessary contributions to a valuable whole. There is no alternative perspective from which this fact can be omitted.

If the group's projects, or the very flourishing of the group, has value, then that which sustains the group or in which its flourishing consists will have value. This



value devolves upon the relevant individual actions of the group's members, and issues in obligations of loyalty.

From this we can deduce the nature of individual obligations of loyalty, Dworkin's second question, if we understand first the importance of the intrinsic value of large good groups, and second the degree of collective loyalty necessary to ensure that they survive and flourish. The nature of the intrinsic value embodied in the group and of the consequential obligations of individuals to the group will also determine the scope of group obligations - that is membership criteria, and criteria of entry to and exit from the group. It is to these questions that we now turn.

## **7B The Value of Large Groups**

The cogency and the nature of obligations of loyalty to large groups depends upon the group's peculiar intrinsic value. This will determine what loyal actions will serve to sustain and to energise the group.

Large groups can embody two forms of intrinsic value: large groups are fora in which i) moral and scientific debate can flourish, and ii) interpersonal virtues, particularly the virtue of justice, can be exercised.

Although loyalty is unfair to those outside the group, and although for that reason amongst others loyalty will compromise some moral ends, the value of the group is often sufficient to justify the imposition of powerful obligations upon the individual.

### **i The Group as Moral Forum**

The test of objectivity discussed in Chapter Six reveals an epistemic dependence of individuals upon the group. Though it is an epistemic rather than an metaphysical dependence, individuals nevertheless find that they can bear

neither a belief system in general nor a morality in particular except as members of groups. The group's morality, like the group's science, becomes, by default, the guide as to what is right.

It is because he recognises that to ascribe objectivity to moral claims requires participation in a society, that Aristotle can compare the solitary man to a solitary draughts piece - values do not apply to him - none of his statements have objective meaning (the impossibility of a private language).

Thus, the value of the group to man lies in part in providing a test for objectivity for moral (and other) judgments.

"Speech ... serves to indicate what is useful and what is harmful, and so also what is just and what is unjust. For the real difference between man and other animals is that humans alone have perception of good and evil, just and unjust, etc. It is the sharing of a common view in these matters that makes a household and a state." Aristotle, Politics, lii

One might think that the value of the group is to determine truth. But if the group or community is the epistemic source of moral truth, how is progress or justified rebellion ever possible?

Lovibond explores this issue at length (1983). On page 170 (and elsewhere - see eg pp 195 and top of p. 198) Lovibond recognises the desirability of an account of "Kantian Moralitat" to provide a perspective outside that of the group from which the group can be criticised, to provide a space for the moral reformer. But Lovibond's own understanding does not provide this metaphysical space. She would locate the development of morality entirely within the group: Hegelian *sittlichkeit* unguided by Kantian moralitat. Hence, as we saw in Chapter Six, she cannot criticise the group's morality. The community is the locus of a freely developing morality, but this freedom is empty, for it is a freedom undirected by truth.

True freedom must be a freedom to choose what is right, and to work to discover what is right. This is available only to a more full blooded realist than Lovibond. But if the group's morality and science is guided by the truth, what then is the peculiar moral value of the large group?

The value of the community for a full blooded realist resides not in its invention of moral truth but in the indispensable role that it, with other communities, plays in discovering it. The community acts as a test bed for moral innovation: Can the rebel convince at least a fragment of her community of the error of their ways? If she cannot she must doubt her own insights and even her own sanity; if she can, then her revolutionary discoveries are to that extent vindicated.

Even Hegel, the chief exponent of the large group, the state, as the bearer of moral truth, does not imagine that the community or even the state is an infallible guide to the right. Hegel differs from Rousseau in having less faith in the ability of the democratic constitution to deliver a body of law conforming to the right. This is no contradiction within Hegel, for the epistemic dependency of morality upon consensus within society does not require that any particular constitutional form will be able unerringly to produce laws in accordance with that moral code. Indeed it must be admitted that the constitutional form advocated by Hegel is also not guaranteed to do so - and he evidently was aware of the possibility that even fundamentally sound states might stray. Note the concession at the end of this famous passage:

"The state in and by itself is the ethical whole, the actualization of freedom... In considering freedom, the starting point must not be individuality, the single self-consciousness, but only the essence of self-consciousness; for whether man knows it or not, this essence is externally realised as a self-subsistent power in which individuals are only moments. The march of God in history, that is what the state is. The basis of the state is the power of reason actualising itself as will. In considering the idea of the state, we must not have our eyes on particular states or on particular institutions. Instead we consider the idea, this

actual God, by itself. On some principle or other, any state may be shown to be bad ..." (Hegel, *Philosophy of Right*, 1821, addition to paragraph 258).

It is the task of individuals to strive for a state which embodies the true rationality as they conceive it.

Hegel explicitly stresses that the obligation of the individual to the group is contingent upon the group being composed of individuals identifying with the group (otherwise only an association, of which membership is voluntary, has been achieved). If, as I suggest, the value of the group is in part determined by the epistemic foundation that it provides for moral discourse, then the parallel conclusion is that obligations to the group are contingent upon individuals concurring in the ascription of moral terms, such that it is clear that their community is one that uncovers moral truth. Only such a moral community is worthy of support. A similar criterion might identify a scientific community worthy of support; and conversely, a group of scientists who could establish no consensus on the application of descriptive terms would find it difficult to sustain the enterprise of science, and would not merit loyalty.

"If the state is confused with civil society, and if its specific end is laid down as the security and protection of property and personal freedom, then the interest of the individuals as such becomes the ultimate end of their association, and it follows that membership of the state is something optional. But the state's relation to the individual is quite different from this. Since the state is mind objectified, it is only as one of its members that the individual himself has objectivity, genuine individuality, and an ethical life. Unification pure and simple is the true content and aim of the individual, and the individual's destiny is the living of a universal life. His further particular satisfaction, activity, and mode of conduct have this substantive and universally valid life as their starting point and their result."(Hegel, *op. cit.*, remark to paragraph 258.)

## **ii The Group as Forum for Exercise of Relational Virtues**

The second value of the group is as an arena in which to be good. It is only in the state that some virtues can apply: "The virtue of justice is a feature of a state; for justice is the arrangement of the political association."(Aristotle, loc.cit.)

These two forms of value borne by the group provide a theoretical underpinning to the bare cogency of duties of loyalty. Indeed, between the two passages quoted, Aristotle asserts that "the state has a natural priority over the household and over any individual amongst us", indicating that the state *derives* its priority from its role as a focal point of the community's moral values and an enabler of the exercise of the virtue of justice.

The groups that bear value, and thus can demand loyalty, are diverse. For instance an individual living in a Western democracy would have myriad overlapping ties to the various institutions and groupings which make concrete her moral universe, from her local chess club or village or church, to the whole of Western democracy (and indeed, if the argument in Section 6E is correct that charitable translation preserves moral truth across cultures, to the whole of mankind).

Many of the smaller groupings will play only a small role in the moral life of the individuals who make them up, and there may be only a small class of evaluative terms for which continuing consensus in application is secured by that group. Such groups will be able to call upon only limited degrees of partiality: an obligation imposed by the chess club may easily be overridden by obligations, whether partial or not, imposed either by more embracing groups that are more fundamental to the development of a moral tradition, or by closer groups which create greater moral opportunities.

Is there any reason why more embracing groups should demand greater loyalty?  
Aristotle, again:

"As all associations aim at some good, that association which is the most sovereign among them all and embraces all others, will aim highest, i.e. at the most sovereign of all goods. This is the association which we call the state" (Politics li).

The value of the state, and the extent of obligations to it, depends in part upon its success in creating a coherent moral tradition with which almost all of its members identify, and which sustains the moral enterprises of the smaller communities of within it. A state that is successful in this grand aim may have the priority that Aristotle (and Hegel, in the passage quoted above) ascribe to it; a looser association of individual groups, perhaps a pluralistic multicultural society which has failed to establish dialogue between smaller groups in the moral enterprise, or which fails to provide a forum for the exercise of relational virtues between subgroups, would command less loyalty.

### **7C How Much Partiality Should the Group Mandate?**

The cogency and legitimacy of loyalty obligations depends also upon what it is reasonable for a group to mandate given the particular value of the group. A weighing of obligations according to the genuine needs of the state is therefore required. In this section, I contrast this view with those of MacIntyre and of Walzer, and draw conclusions regarding the defeasibility of obligations of loyalty.

This is not, however, the view of MacIntyre:

*"If first of all it is the case that I can only apprehend the rules of morality in the version in which they are incarnated in some specific community; and if secondly it is the case that the justification of morality must be in terms of particular goods enjoyed within the life of particular communities; and if thirdly it is the case that I am characteristically brought into being and maintained as a moral agent only through the particular kinds of moral sustenance afforded by my community, then it is clear that deprived of*

this community I am unlikely to flourish as a moral agent. Hence my allegiance to the community and what it requires of me - even to the point of requiring me to die to sustain its life - could not meaningfully be contrasted with or counterposed to what morality required of me. Detached from my community I will be apt to lose my hold upon all genuine standards of judgment." MacIntyre (1984, pp.10-11)

MacIntyre here emphasises the value of the community to the individual, and uses that as a foundation for a virtue of patriotism strong enough to warrant self sacrifice at the group's behest. But his argument is implausible. It relies upon two additional suppressed premises:

- First, the argument assumes that it follows from an individual's dependence upon X that X should be supported. The dependence MacIntyre envisages is dependence for moral sustenance upon a community. But consider an analogy. Suppose someone is totally dependent upon an evil man to the extent that she could not live without him, it would not follow that she should be loyal to him. In such circumstances, the virtuous woman might be called upon to sacrifice her life to **harm** her protector. Of course, MacIntyre is assuming that the community that is succouring the individual is one that embodies value. But in that case, the community is worthy of support regardless of the individual's dependence upon it.
- The second premise is that the individual's loyalty is needed by the relevant community. It does not follow from the individual's dependence upon the community that the community is dependent upon the individual's loyalty. And even if the community embodies great value, it will not legitimately demand partiality from its members if it can thrive without it, or if other considerations outweigh likely benefits from partiality.

Absent these premises, MacIntyre's argument collapses to an *ad hominem* argument against someone who claims that he need not support the community because he is independent of it. But rather than persuade such a person that they are indeed dependent upon the community, the appropriate response would be to try to convince them of the value of community, and of his obligation to contribute to that good. MacIntyre emphasises aspects of the intrinsic value of the community (the support to the individual; the group practices of intrinsic merit) but misleadingly implies that their worth is contingent upon the perspective of the individual; and he leaves unanswered the question: how much loyalty can a community properly demand?

MacIntyre also leaves undefined the scope of the community to whom loyalty is required. One test he gives is that deprived of the life of that community / would be "unlikely to flourish as a moral agent". One part of the risk relates to the epistemic dependence upon the community to enable me to "apprehend the rules of morality". We have argued that this epistemic dependence upon the community for morality is structured: there is a dependence upon the widest human community for the basic furniture of the moral universe - the particular judgments about the clearest cases. Obligations of loyalty related to these virtues of community will relate in part to this widest of all communities.

MacIntyre would emphasise that dependence upon the human community as a whole for moral context is supplemented within smaller communities by particular moral institutions and practices that others may find difficult to evaluate, and by locally expounded moral theory or theories to explain the moral facts, which may issue in particular moral judgments about which there is no inter-community consensus. In these dependencies resides some of the value of these communities, which could be analysed as above as providing both epistemic environments for the local enterprise of moral discovery, and for the exercise of the moral virtues.

The strength of obligations of loyalty will, however, be commensurate with the value of the respective communities. These obligations of loyalty will have to fight



for their place against more general obligations, even including those generalist obligations that are the peculiar insight and discovery of the local community. There is no contradiction in a local community being responsible for the discovery of a set of universal moral truths that in practice oblige the members of the community to focus their concern exclusively on peoples outside their own moral community. The intrinsic value of the community may be great, but not so great that it overrides the pressing needs of those in other communities, even where the home community has the extraordinary merit of being the first to recognise such outward-looking (perhaps international) moral obligations.

That this is no contradiction follows from the objectivity of moral truth: even if morality is epistemically dependent upon such morally-sensitive communities, its metaphysical base is not dependent upon any community, and there is no inconsistency in a community judging that even self sacrifice, with the loss of moral knowledge that would accompany the destruction of a community, may be required in some circumstances for the greater good (e.g. to avoid great suffering).

Yet this also need not be so: in other circumstances support for one's own community notwithstanding competing calls on one's time and resources will indeed be justified by the fact that the home community is the bearer of important moral truth and insight; partiality may equally be justified by the community's production of another good (perhaps aesthetic); or finally one may simply be obliged to support the group because the flourishing of healthy groups is intrinsically valuable, and flourishing consists in such mutual support (as discussed in Chapter Four).

In sum: the community must justify the demands it makes upon its members. Goods that morality deems worth pursuing are both the goods particular to the community, including the community itself, and the goods common to all communities.

Hence, *pace* MacIntyre, it does not follow immediately from one's dependency upon one's community for one's intellectual apparatus that one need be patriotic, or indeed that the community itself should commend patriotism in a practically relevant sense -- a sense in which patriotism involves discriminating in favour of one's own community to the detriment of others.

A similar communitarian fallacy is found in Walzer (1983), who appears to take it for granted that the particular virtues that are developed most fully in the institutions of a community are *ipso facto* limited in application to the members of that community.

Walzer's book explores the social meanings conferred on morally-laden terms within western communities, and the implications of such meanings for the scope of moral obligations. The argument to best explanation developed in Chapter Six above provides an epistemic foundation for Walzer's view that the "social meaning" of a particular morally-laden term is significant; that when usage implies that a particular form of action is right, then that judgment is probably correct and binding. (The "social meaning" of a term is what is implicit the enduring consensus in its application to new cases; and the best explanation of such enduring consensus is that the community severally and jointly apprehends moral fact.)

But that a community does embody in its judgments and practices moral truth does not imply that the proper range of these judgments and practices should be limited to the community's members.

For instance, Walzer convincingly argues that, within the social meaning of 'medicine' in modern western society is a prescription that doctors ought to be swayed in the meting-out of treatment by nothing other than medical need. Walzer suggests that this rule has application only to fellow members of the nation-state sharing this social meaning. But there is nothing in our understanding of the moral foundation of the practice of medicine to support such a limitation.

Rather we must conclude that the extent of obligations of loyalty depends upon moral judgment in the context of local circumstances - judgment that recognises that the promotion of the particular value that is this flourishing community is not the only end.

Western moral theory, *our* understanding of the moral truth, is averse to discrimination - hence the tension between impartial theories and partialist moral facts discussed in Chapter One and the premium upon tolerance explored in Chapter Five. This aversion is justified by a desire to express mutual respect and a sense of equality of human worth, to sustain individuals' self respect, to promote diversity of moral enterprise and to promote fairness.

Especially if we are right about the value of these moral ends, partiality to the group **must be justified**, in one of the following ways:

- loyal practices may be constitutive of a group's flourishing, and the latter may be a good of such intrinsic value as to offset loss of fairness etc
- loyal practices may be necessary instrumentally to enable the community to pursue other ends which have sufficient intrinsic value as to offset loss of fairness etc
- loyal practices may be necessary to preserve the community itself from disintegration, and the community may be of sufficient intrinsic value as to warrant incurring other costs to sustain it.

How do some of the paradigm acts of loyalty to large groups, particularly to states, stack up against these criteria. In particular, should we divert time and resources from other pursuits in order to vote, to pay taxes, to fight, at the behest of the governments of western democracies? A full treatment would be particular to a specific community; nevertheless the following thoughts are illustrative of the types of considerations that might be relevant.

i) Voting. Voting is justified both as a practice that is constitutive of the group's flourishing and as instrumental to decision making in the interest of the group. In

few democracies is voting compulsory. The reasoning perhaps is that only voluntary voting is genuinely constitutive of a flourishing democracy.

ii) Taxpaying. Tax is required instrumentally for the prosecution of a community's projects. Its value, and the appropriate degree of compulsion will be dependent upon the objective value of the projects. However, even if the particular projects of a particular government are not judged meritorious, the provision through taxation of funding for government might be judged necessary to the effectiveness of the democratic process and hence to the flourishing of the community.

iii) Fighting. Ultimately the very life of the community, may be in jeopardy. But fighting will put at risk the lives of individuals on both sides. In Chapter Five we argued, following Rawls, that the good is what gives life point. If part of the good is embodied in community, then is possible that life should be sacrificed to community.

In all cases, obligations are dependent upon the value of the particular community offsetting the direct and opportunity costs of the sacrifices demanded on its behalf. Hence the rebel can argue that a particular community is not inclusive or cohesive enough, not sufficiently conducive to good practices, to merit sacrifice. However, it is also important that the very act of making sacrifices to the community can strengthen the community, reinforcing the justification of the sacrifice.

Obligations of loyalty must be proportional: loyalty need not be blind to the costs that it would impose. So obligations should be proportional to the different sacrifices required of different individuals. Such proportionality is delivered by allowing each member of a community to determine whether she is caught by the communal obligations judging her own position from the group's perspective.

Recall the analogy in section 7A with the smoker considering a string of acts of restraint that would yield a joint health benefit. How should the ex-smoker view an

offer of a cigarette accompanied by the suggestion that just one will have no effect? She should ask herself: is this consistent with my original decision, or with the decision implicit in my long abstention? She may respond that yes this is an exceptional circumstance. (Perhaps she is an actress required on this one occasion to smoke for the performance.) But in response to the challenge, she will have to judge the individual case as part of the longer term behavioral paradigm that she has set for herself.

Similarly the voter must view his case as part of the collective whole. If there has been a collective decision to perform a certain act, then exemptions may be ordained centrally. Thus conscription waivers may be set for those in vital jobs or newly wed. These rules implicitly assess the individual's circumstances in the relevant public context.

However, such an assessment can be accomplished even in the absence of an explicitly agreed set of rules. Individuals can judge their own circumstances in the context of whatever imperative or project is governing the group as a whole. They must judge whether their circumstances are such that exemption would be consistent with the success of the overall project. A decision not to vote may be appropriate even whilst recognising the importance of the collective enterprise: proportionality to costs of voting might well be built into the correct assessment of the obligation. So for example, it might be judged that only those within easy reach of a polling booth on the day of the election were so obligated.

Individuals' dependence upon the collective judgment of others to confirm their moral judgments, may act to restrain people from judging their own circumstances in too favourable a light. In judging their position to be exceptional, agents should seek to echo the continuing consensus regarding the correct way to sustain the valuable institutions of their community<sup>74</sup>: would others agree that

---

<sup>74</sup> Though it is the institutions of his group that the agent seeks to sustain, the value of those institutions will be objective rather than relative to that group. In principle, therefore the group whose endorsement he seeks in confirming the correctness of his own

they are right not to vote; would others agree that their case is exceptional?

Agents may not be able to state a general rule under which their position falls, but they must have confidence that others would recognise its character: if not, they owe an explanation of the group's error.

What now if the individual claims that what is exceptional about his case is that he is one of the few that will have the verve to act independently, to devote his election day or his tax money or his conscription labour to some other purpose. As a result the state's objective will be met without his participation.

Others, we may guess, will not view his action favourably, for they will judge his action in the wider group context and will find his willingness to break ranks rather a cause for censure than commendation. They will assess that a community in which some free-ride on the loyalty of others is not achieving the cohesive ideal for which they strive.

The argument is once again posed at the level of the community rather than at the level of the individual, and this is correct: for the individual is willy-nilly a member of the community; and in that context a decision to act other than on the group's project must be assessed as an act of disloyalty. (Even if the individual's circumstances are such as to justify him - and even if this is confirmed by the group's assessment - he may not be able to avoid a sense of disloyalty. To wit: the mixed feelings of those whose jobs were starved during the War.)

It could be however that in a community of natural conformists, the maverick is judged *ipso facto* exceptional, and justified in breaking with the norm and doing his own thing. But even if that is the judgment, it is one that does not undermine

---

judgment will be the wider community who understand his language. This will be important for a reformer sustained by the support from enlightened members of other groups. Nevertheless, she will have to show that her group has become corrupt or otherwise subject to error: otherwise one would expect that the members of a group are best able to judge how valuable are its institutions and what is necessary to sustain them.

collective institutions for it does not generalise: mavericks are rare. An individual can exempt herself, however, if and only if her case merits exemption when judged in the collective context; and her best guide to her own case will be the collective judgment of her fellow citizens as to legitimate grounds for exemption.

## **7D Membership Criteria for the Group**

This justificatory structure has consequences for the question of membership: who are members, who should be allowed to enter the group (enjoying the partiality of the members) and who to exit the group (escaping the obligations towards the members)?

Membership criteria can be justified according to the three justificatory formulae above:

### **- partial actions constitutive of the groups flourishing**

Participation in group activities by those who have historical ties - through birth for example - is a form of flourishing. This is MacIntyre's focus when he emphasises that only a Frenchman can be loyal to France (op. cit.). By the same token, loyalty obligations are strongest to "Frenchmen" who have participated sufficiently to have built historic ties with France. Reinforcing such ties through acts of loyalty is in part what constitutes the flourishing of the French nation. This criterion should inhibit exit by individuals from a group: if they leave a group one particular form of contribution to the good is foregone. However, these considerations do nothing to justify denial of entry to one who wishes to build ties.

### **- partial actions necessary to enable a community to pursue its ends**

Those who would prejudice achievement of a group's ends should be excluded by this criterion; but exit of the talented, who understand the group, and are in a position to contribute to the production of the group's culture, might justifiably be constrained. This does not imply that forcible restraint on exit is justified - merely

that exit should be reckoned to have moral costs. If these are great enough, the community may apply moral pressure, or even in extreme cases physical penalties upon emigrants. Conversely expulsion of those who impede the realisation of the group's goods may be justified (banishment or exile).

**- viability of group threatened**

Those whose entry would threaten the very existence of state (gross overcrowding, civil unrest etc.) may be excluded.

Walzer (1983, Ch.2) argues that limitation of entry into a group is a fundamental right. But whence this fundamental right he does not say. Indeed, as we have indicated, it stands at odds with some of the conceptual arguments he uses elsewhere. Regarding the allocation of offices, Walzer says that "the right to equal consideration .... is one of the things the members owe to one another", but he does not say why the social meaning of office, which demands that a job must be given to whoever is best qualified, should be limited to members of a given nation. Opening offices and other public contracts to international competition might be more supportive of group viability.

**Arguably membership of groups should be granted to anyone who is willing to participate in the production of the good.** However full membership might await the building of historic ties through contribution to the group and to its projects. Entry requirements may be strict or lax, according to the nature of the group. Even if someone is excluded from group membership on the grounds that she does not wish to participate in production of the group's particular good, it will still not follow that the group's doctors should not treat her, or that the group's armies should not defend her. Such denial in the presence of the inclusive demands of fairness and of pain minimisation will only be justified where necessary for the promotion of the group's good.

Conceptions of the nature of the intrinsic value embodied in a society or group, although vexed, must be brought to bear on the settling of questions of



immigration or membership. If a golf club's value relates exclusively to promotion of excellence in sport, then it is legitimate to discriminate against potential entrants who are lacking in skill or enthusiasm for golf. It would also be reasonable to limit membership to prevent such overcrowding as would adversely affect the pursuit of good golf. But it **would not** be legitimate to discriminate on grounds of race, sex or religion.

Discrimination on grounds of religion is sometimes legitimate, however, because a person's religion -- unlike his race or sex -- may well be relevant to the good being pursued by the group. A Roman Catholic school which wishes to foster an unquestioning faith in its pupils is justified in excluding children of other faiths, but not of other races or backgrounds. The crucial test in these circumstances is whether a group allows membership if the candidate espouses the values of the group. Judaism is in many ways an elitist religion: it is not evangelical, and yet has a strong belief in the rectitude of its way of life, discouraging excessive social intercourse between its devotees and gentiles, disallowing inter-marriage. But it is not racist, for the restrictive practices are adopted in pursuit of the good of promoting a way of life and, crucially, it is open to converts who are also willing to adopt and pursue this way of life.

What can we say about immigration into a Western democracy like England? To be categorical would require a more precise definition of the good or goods embodied in the community or communities that are England, but the lines of argument should be clear. For instance, forbidding immigration of the indigent on the grounds that welfare payments would be increased would not be legitimate unless it were suggested that the resulting impoverishment of (the rest of) the community would be so great as to threaten the social fabric. Even then, the highly-uncertain nature of such an economic forecast would have to be weighed against the certain relief of distress that could result from the immigration. *Pace* Walzer, I see **no** grounds for disregarding the unhappiness of the potential immigrant.

On the other hand, to the extent that a society is cohesive, and is promoting a particular way of life, the immigration of those with a very different outlook may legitimately be restricted. In the case of England, however, it is arguable that one of the goods that it fosters is diversity and tolerance; such a good will tend to be promoted by allowing immigration to those of differing beliefs, even including evangelists for different values. The stronger a society is in its own beliefs, the more it is likely to be able to tolerate and even benefit from the admittance of outsiders. Hegel (1821), Para 270, note 3, makes the point in this regard that although states do, strictly, have the right to exclude those who do not accept their legitimacy and authority, if they are strong they can rely on the appeal of their own institutions gradually to lessen the gulf in outlook between members of the state and outsiders. He further makes the point that manhood itself represents the common "root from which the desired similarity in disposition and ways of thinking comes into being" so that the exclusion of even those of different race and custom (like the Jews) "is the silliest folly".

Intertwined with questions of membership of a group are questions of access to that which the group controls, whether golf links or comradeship. In the case of nations, historically it has often been access to land which has triggered disputes. Groups or nations claim rights to land over and above the property rights of the members of the group. These are rights to police, tax and vet the ownership, or even compulsorily to purchase the land under jurisdiction. On what basis can a group claim such rights? Walzer claims that "politics is always territorially based" (1983, p. 225). For him, this cuts both ways: he allows nations to exclude others from what they consider their territory, but he also feels very uncomfortable with the idea of a nation sharing its land with another group, or with individuals who are not granted full membership (for instance, guest-workers).

But the cogency of the linkage between groups and their territory should be dissected. Others should only be excluded from settling in a group's land if such settlement is likely to impair the value embodied by the group; for excluding sale

or rent of living space to other groups is gross partiality, liable to compromise efficiency, fairness, and self-esteem.

The conventional objection in the case of guest workers arises from the fact that often such workers would be very pleased to be given full citizenship. It was argued above that there will often be no good grounds for denying them citizenship, if they will not impede the attainment of the goods strived for by the society. But let us suppose that there are good grounds for excluding someone from citizenship but not from physical entry. If a society has developed an advanced welfare system which functions harmoniously, excessive new membership could strain the system and might legitimately be ruled out, given the difficulty of achieving and sustaining such a system, and the great loss of welfare that would follow its collapse<sup>75</sup>.

Alternatively, a society may require a high average skill level amongst its members so as to generate sufficient resources to devote to artistic endeavours -- new entrants who would be a drain on resources may prejudice the success of these endeavours.

In both cases arguably, but in the latter case certainly, equity and compassion make a counter-claim upon the society to keep its resources open, but the narrower aims of the society may on occasion be deemed to prevail.

In cases where full membership of a group is legitimately ruled out, is it really wrong to allow entry to guest workers as Walzer contends -- occupiers of the land without citizenship rights? As both the society and the guest workers benefit from such arrangements, the onus must be on those who would forbid them.

---

<sup>75</sup> Here we are justifying myriad partialist acts of discrimination on the grounds of their collective impact (cf section 7A above): an individual refusal of membership is more likely than not to diminish welfare; but refusal might nevertheless be justified on the basis of the risk to the collective good consequent upon a general open immigration policy undermining domestic harmony. Of course the plausibility of such a claim would have to be sustained.

However, a problem with this approach to guest workers is that rich citizens may want to adopt this semi-independent status. Such exit would have to be resisted. Here a judgment of whether the group could retain its cohesion and bonds whilst entertaining guest workers would be required. If the guests are different - eg in religion - differentiation might be more viable.

One further objection to this arrangement, and indeed to that of a permanent tourist, is that to the extent that societies offer welfare services, it will offend the society if residents do not have access to the services. Will the society be able to stand by idly while its unemployed guest workers starve or fall ill ? Such arguments extend to education: can the government allow the gypsies to educate their children after their own fashion? But the force of these arguments surely derives merely from our uncertainty as to how much partiality to members is justified. If we feel guilty about the poor of non-members who reside near us, this may indicate that we are not justified in excluding them from our succour. Poor resident aliens might excite more compassion than the poor of distant nations, but this is more an indictment of our lack of compassion for the distant poor than a justification for keeping them distant!

In sum, the position reached on immigration and emigration (entry and exit) is the reverse of that which prevails in most modern countries: entry to the group should be open to all who are willing and able to contribute to the flourishing of the group (unless the strain to the group of accepting new entrants exceeds the total benefit, **including** the benefit to the immigrant), exit from the group may be stigmatised in order to ensure fulfilment of obligations to the group.

Note that the approach here advocated could result in radically different rights and obligations being imposed upon citizens of different states, according to the nature of the particular goods embodied by their communities. Tan (1998) indicts Rawls for inconsistency between his treatment of individuals within a liberal society, and his approach (in his Law of Peoples) to those within a hierarchical state. Rawls wishes to accord respect to the hierarchical state, not to interfere in its internal affairs providing it meets certain criteria. But as Tan points out, Rawls

finds it hard to justify this from within his conception of Justice as Fairness. I suggest that to make sense of the two level determination of the rules of justice that Rawls apparently now advocates (one at the level of the state, one within liberal societies), it is necessary to recognise that flourishing groups of many sorts may be of value - indeed that there may be value in diversity. But then it follows that discrimination against individuals within those societies, and between societies, may also be justified. It also follows that the justification of the liberal society must be on the basis of the positive values that it represents, rather than by construction from some limited overlapping consensus.

## **7E Concluding Remarks**

What direction should conceptual and empirical enquiry take if it is to help to resolve the tensions between the partialist and the impartialist perspectives? There is a substantial and enduring tension in much political philosophy between strivings on the one hand to develop substantive theories with practical application, and on the other, to achieve universality and a claim to truth.

This tension is well illustrated within Rawls' work. As O'Neill (1997) points out, although "from a communitarian perspective, Rawls' account of reasoning about .. justice assumes too slender a shared sense of identity", from a wider perspective "Rawls assumes too much in building his account of public reason on an account of citizens' shared sense of political identity" (p.422). Thus on the one hand he does not allow his view to be informed by the particularities of the community that he addresses, and hence seems to ignore some of what imbues that community with moral vitality; but on the other, universal application is not assured since he is forced to assume some minimal commitment to the very liberal values that justice as fairness seeks to establish. (O'Neill herself advocates a yet more abstract Kantian account of a reason that will reach "the

world at large" (p.424<sup>76</sup>). But this approach cannot resolve the tension but only shift the balance; and communitarian values will still go unrecognised.)

The answer I have suggested here is to embrace an objectivist pluralism. Universal political truths may be few, but values and good actions that can be universally recognised are many. Yet, recognition that values that are genuine can nonetheless conflict, means accepting that there is no hope of a single correct political system reconciling all tensions; politics necessarily involves struggle.

If that is the correct framework, what role can the political philosopher play within a particular ethical tradition?

First, she can explore and highlight the tensions between the different values that a particular tradition is attempting to embody, and the costs of ignoring those it is downplaying. This tension is real and can be fraught. Highlighting these value conflicts should ensure that debate is well informed.

When the values of community are allowed to confront those of impartiality, the priority of the right must be compromised, and discourse must recognise that the taboo on viewing individuals as means as well as ends in themselves is broken. The discussion above regarding immigration illustrates this. Yet whilst asserting that community has value, communitarians should recognise that there is a broader perspective, and that the values of loyalty are necessarily discriminative ones, compromising our correct desire to show equal respect to all.

Conversely, liberals need to confront the opportunity costs of liberalism. Waldron (1993, Chapter 15), for example, argues that an underpinning of liberal rights is necessary to provide individuals with the freedom to choose their identities and their commitments. We can accept Waldron's point that this freedom is of value. But too often liberals elevate such values to a point of priority, failing to recognise

---

<sup>76</sup> O'Neill here quotes Kant's 1784 essay "What is Enlightenment".

their costs. There is a trade off. Establishing rights to public welfare has clear benefits but can also undermine family support networks and local communities; rights to marry who you wish can create social harmony (Waldron cites Romeo and Juliet) but can also undermine the preservation of a culture (when someone marries outside the faith).

There are good arguments in defence of impartiality particularly in the design of political institutions, but they are not meta-ethical arguments to be wielded behind a veil of ignorance with priority over conceptions of the good. Rather they concern the objectives of underpinning self-esteem, of recognising that in one important sense all have equal worth, of maximising fairness in different senses, of bolstering personal autonomy. These goals must compete with goals of community.

The appropriate balance will be determined by the nature of the good that is realised in different communities. In cohesive societies that realise goods of community, highly discriminatory practices may be justified. Societies that emphasise liberal individualism must be confident that the cost in community values is justified.

Donald Franklin  
December 1999

## **BIBLIOGRAPHY**

- Anscombe, Elisabeth** (1971) "Causality and Determinism", reprinted in Soza and Tooley, Eds, "Causation", Oxford Readings in Philosophy, 1993
- Aristotle**, The Politics, translated Sinclair and Saunders 1962/81, Penguin
- Baron, Marcia** (1987) "Kantian Ethics and Supererogation", Journal of Philosophy, Volume LXXXIV, No.5 (May), pp 237-262
- Barry, Brian**, (1995) Justice as Impartiality, Clarendon Press, Oxford
- Berlin, Isaiah** (1976) Vico and Herder, The Hogarth Press
- Blackburn, Simon** (1981) "Rule-Following and Moral Realism", in Holtzmann and Leich, eds, "Wittgenstein: To Follow a Rule", Routledge Kegan Paul
- Blum, Lawrence** (1980) "Friendship Altruism And Morality", RKP
- Brink, David O** (1986) "Utilitarian Morality and the Personal Point of View", Journal of Philosophy, Volume LXXXIII, No.8 (August), pp 417-438
- Brink, David O** (1989) Moral Realism and the Foundations of Ethics, Cambridge University Press
- Burnyeat, Myles** (1980) "Aristotle: on Learning to be Good" reprinted in Philosophy Through its Past, ed. Honderich, Pelican, 1984
- Cocker, Dean and Oakley, Justin** (1995) "Indirect Consequentialism, Friendship and the Problem of Alienation", Ethics 1995 pp 93-133
- Cottingham, John** (1983) "Ethics and Impartiality", Philosophical Studies 43, pp 83-9
- Cottingham, John** (1991) "Ethics of Self-Concern", Ethics, Vol 101, No.4, July 1991, Symposium on Impartiality and Ethical Theory, pp 798-817
- Cottingham, John** (1998) "The Ethical Credentials of Partiality", Proceedings of the Aristotelian Society, Vol XCVIII, part 1, 1998, Presidential Address
- Davidson, Donald** (1963) "Actions Reasons and Causes", reprinted in Philosophy of Action, Ed. Alan White, OUP
- Davidson, Donald** (1973) "On the Very Idea of a Conceptual Scheme" Presidential Address, American Philosophical Association, Dec 28, 1973
- Dummett, Michael** (1991) "Realism and the Theory of Meaning", Chapter 15 of "The Logical Basis of Metaphysics"



**Dworkin, Ronald** (1977) *Taking Rights Seriously*, Duckworth

**Dworkin, Ronald** (1981) "What is Equality", *Philosophy and Public Affairs*, Vol 10 pp 185-246 & 283-345

**Dworkin, Ronald** (1986) *Law's Empire*, Fontana

**Foot, Philippa** (1958) "Moral Beliefs", *Proceedings of the Aristotelian Society* 1958-9, pp 83-104, reproduced in *Ethics*, ed. Thomson and Dworkin, Harper & Row, 1968

**Gewirth, Alan** (1988) "Ethical Universalism & Particularism" *Journal of Philosophy*, June 1988

**Gödel, Kurt** (1931) *On Formally Undecidable Propositions*, English Translation, New York Basic Books, 1962

**Habermas, Jurgen** (1995) "Reconciliation through the Public Use of Reason: Remarks on John Rawls' Political Liberalism", *The Journal of Philosophy*, March 1995

**Hare, Richard** (1982), *Moral Thinking: It's Levels, Methods and Point*

**Hegel, (1821)**, "Philosophy of Right", quotations from TM Knox translation, 1952, OUP

**Honderich, Ted** (1969) *Punishment, The Supposed Justifications*, Pelican Books

**Hooker, Brad** (1995) "Rule Consequentialism, Incoherence, Fairness", *Proceedings of the Aristotelian Society*, 1995 Part 1

**Hume, David** (1757) "On Taste"

**Kekes, John** (1981) "Morality and Impartiality", *American Philosophical Quarterly*, Vol 18 pp 295-303

**Kripke, Saul**, (1982) *Wittgenstein On Rules and Private Language*, Harvard

**Kymlicka, Will** (1989) *Liberalism, Community and Culture*, Oxford

**Kymlicka, Will** (1990) *Contemporary Political Philosophy*, Oxford

**Lovibond, Sabina** (1983) *Realism and Imagination in Ethics*, Basil Blackwell

**MacIntyre, Alasdair** (1984), "Is Patriotism a Virtue", Lindley Lecture, University of Kansas, 26th March 1984

**MacIntyre, Alasdair** (1981), "After Virtue, a study in moral theory", Duckworth

**MacIntyre, Alasdair** (1988), "Whose Justice, Whose Rationality?"

**Mackie, J.L.** (1977), *Ethics: Inventing Right and Wrong*, Harmondsworth, Penguin

**McDowell, John** (1978), "Are Moral Requirements Hypothetical Imperatives", *Proceedings of the Aristotelian Society: Supplementary Volume*

**McDowell, John** (1981) "Non-cognitivism and Rule-Following", in Holtzmann and Leich, eds, "Wittgenstein: To Follow a Rule", Routledge Kegan Paul

**McDowell, John** (1994) *Mind and World*, Harvard

**Mason, Elinor** (1997) "Can an Indirect Consequentialist be a Real Friend", *Ethics* 1997 pp 386-93

**Mendus, Susan** (1996), "Some Mistakes about Impartiality", *Political Studies* XLIV, pp 319-27

**Mill, John Stuart** (1859), "On Liberty" in *Utilitarianism, On Liberty, and Considerations on Representative Government*, edited by H.B.Acton, 1972, Dent Dutton

**Nagel, Thomas** (1986), "The View from Nowhere", OUP

**Nozick, Robert** (1974), *Anarchy State and Utopia*, New York, Basic Books

**Nussbaum, Martha** (1985) "The Discernment of Perception: an Aristotelian Conception of Private and Public Rationality", Chapter Six in *Proceedings of the Boston Area Colloquium in Ancient Philosophy, Volume I*

**O'Neill, Onora**, (1996), *Towards Justice and Virtue*, Cambridge University Press

**O'Neill, Onora**, (1997), "Political Liberalism and Public Reason: A Critical Notice of John Rawls, *Political Liberalism*", *Philosophical Review*, vol 106, July 1997, pp 411-428

**Parfit, Derek** (1984) *Reasons and Persons*, Clarendon Press, Oxford

**Piper, Adrian** (1991) "Impartiality, Compassion and Imagination", *Ethics*, Vol 101, No.4, July 1991, Symposium on Impartiality and Ethical Theory, pp 726-757

**Platts, Mark** (1980) "Moral Reality and the End of Desire" in *Reference Truth and Reality*, Ed Mark Platts, RKP

**Quine, W v'O** (1951) "Two Dogmas of Empiricism", reprinted in *From a Logical Point of View*, Harper 1953

**Railton, Peter** (1984), "Alienation, Consequentialism and the Demands of Morality", reprinted in Scheffler (1988), pp 93-133

**Rawls**, John (1972), "Theory of Justice", Clarendon, Oxford

**Rawls**, John (1985), "Justice as Fairness: Political not Metaphysical",  
Philosophy and Public Affairs 14 no 3 (Summer 1985)

**Rawls**, John (1988), "The Priority of Right and Ideas of the Good", Philosophy  
and Public Affairs 17 no 4

**Rawls**, John (1996), Political Liberalism with a new introduction and the "reply to  
Habermas, Columbia

**Raz**, Joseph (1986) The Morality of Freedom, Oxford University Press

**Ross**, W.D. (1930), The Right & the Good, Hackett Publishing Company,  
Indianapolis/Cambridge, 1988

**Russell**, Bertrand, "Notes on PHILOSOPHY January 1960", in Philosophy vol  
35, 1960

**Sandel**, Michael (1982), "Liberalism and the Limits of Justice", CUP

**Scheffler**, Samuel (1982), "The Rejection of Consequentialism", Clarendon,  
Oxford

**Scheffler**, Samuel (1988), ed., "Consequentialism and Its Critics", OUP

**Simmons**, A. John (1979) Moral Principles and Political Obligations, Princeton  
University Press

**Simmons**, A. John (1979a) "The Principle of Fair Play" in Philosophy and Public  
Affairs 8, Summer 1979

**Smiley**, Tim (1995) "A Tale of Two Tortoises", Mind, Vol 104, Oct 1995, OUP

**Sober**, Elliott (1981) "Evolutionary Theory and the Ontological Status of  
Properties", Philosophical Studies 40 (1981) pp 147-76

**Stocker**, Michael (1976) "The Schizophrenia of Modern Ethical Theories" in The  
Journal of Philosophy 73 no.14 August 1976

**Tan**, Kok-chor (1998) "Liberal Toleration in Rawls' Law of Peoples" in Ethics,  
Jan 1998

**Thomson**, Judith Jarvis (1997) "The Right and the Good", Journal of  
Philosophy, vol 94, pp273-298

**Vlastos**, Gregory (1962), "Justice and Equality", reprinted in J. Waldron, ed.,  
"Theories of Rights", OUP, 1984

**Walzer**, Michael, (1983) Spheres of Justice

**Waldron, Jeremy**, (1993) *Liberal Rights, Collected Papers 1981-91*

**Wiggins, David**, (1984) "The Sense and Reference of Predicate Expressions", *Philosophical Quarterly*, July 1984

**Wiggins, David**, (1987), "A Sensible Subjectivism" in *Needs, Values, Truth*, Blackwell

**Wiggins, David**, (1998) *Needs, Values, Truth, Third Edition*, Clarendon Press, Oxford

**Williams, Bernard** (1973), "Utilitarianism for and Against", with J Smart, CUP

**Williams, Bernard** (1976), "Utilitarianism and Moral Self-indulgence", reprinted in *Williams* (1981)

**Williams, Bernard** (1976a), "Moral Luck", reprinted in *Williams* (1981)

**Williams, Bernard** (1981), "Moral Luck, *Philosophical Papers 1973-80*", CUP

**Williams, Bernard** (1985), *Ethics and the Limits of Philosophy*, Fontana

**Winch, Peter** (1965), "The Universalizability of Moral Judgements", *Monist*, Vol.49, reprinted in *Ethics and Action*, 1972, RKP

**Wittgenstein, Ludwig** (1921), *Tractatus Logico Philosophicus*; quotes from Pears and McGuinness' 1961 translation, RKP

**Wittgenstein, Ludwig** (1953), *Philosophical Investigations*, translated by GEM Anscombe, Blackwell

**Wittgenstein, Ludwig** (1969), *On Certainty*

**Wittgenstein, Ludwig** (1978), *Remarks on the Foundation of Mathematics*, translated by GEM Anscombe, 3rd Edition, Blackwell

**Wolff, Jonathan** (1996) *An Introduction to Political Philosophy*, OUP

**Wright, Crispin** (1981), "Rule-Following, Objectivity and the Theory of Meaning" in Holtzman and Leich (eds) *Wittgenstein: to Follow a Rule*