INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

THREE STYLES OF IMPARTIALITY

by

LYDIA HUCKER

A dissertation submitted to the Graduate Faculty in Philosophy
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy, The City University of New York

2000

This manuscript has been read and accepted for the Graduate Faculty in Engineering in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

9/19/00
Date

Chair of Examining Committee

9/19/00
Date

Executive Officer

Douglas Lackey
Bernard H. Baumrin
Steven M. Cahn

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

CONTENTS

CHAPTER I

BACKGROUND

Introduction

Being moral seems necessarily to involve being impartial; according to at least one philosopher, morality is even identical with impartiality.[1] Impartiality is at its foundation a preferred attitude to take toward resolving moral problems that involves a stepping-aside from sole consideration of one's own personal desires, biases, and interests in a moral decision-making situation in order to adopt a perspective that is either more objective or more inclusive, one that takes into account the desires and interests of others as well as essential material facts of the situation when warranted. However, impartiality is not a monolithic concept; it is used in moral philosophy to reflect differing attitudes in moral deliberation according to differing theories; this can result in outcomes of moral deliberation that vary significantly from one another.[2]

Furthermore, a survey of various characterizations of this attitude shows that there is a wide variety at least in its verbal expression; for example, these are all words and phrases used in the literature of moral philosophy to denote an impartial attitude: the "God's-eye view," the "original position," the "ideal observer," the "sympathetic benevolent impartial spectator," "objectivity," "inter-subjectivity"; being "disinterested," "neutral," "unbiased," "dispassionate," "interpersonally intelligible," "independent," "even-handed," "detached," "fair," and more. Such a survey suggests that these verbal

expressions also can mean different things, and in cases where there might be significant differences in meaning, it further suggests that their implementation in moral decision-making would lead to different outcomes in the same deliberative situation, depending upon which style of impartiality is being drawn upon.

There are other considerations as well, some of them psychological. Impartiality has been construed at various times as a non-voluntary dispositional, or even natural, response of a certain kind of moral deliberator. At other times, it has been assumed that it must be a deliberately willed, self-conscious stance. These also suggest the possibility of differing outcomes because, if it is true that there are significantly different modalities or styles of impartiality, we must ask whether one chooses a particular style consciously, based on an assessment that this is the best style, for whatever reason, or whether one adheres more or less habitually to a certain style, based perhaps upon longstanding and unconscious biases and assumptions, including embedded presuppositions about personhood, relationships, the proper constitution of a moral community, and other things.

Again, there is a tendency to consider impartiality not only as monolithic, but as quantitative; one speaks about "more" or "less" impartiality. Thomas Nagel's continuing examination of the role of the impartial stance in everyday life, in epistemology and the sciences, and in the design of social structures is a case in point.[3] How are we to reconcile, he asks, the two often-conflicting perspectives within which we must constantly live and between which we must constantly navigate? Our personal perspective tells us one thing about our being-in-the-world; our impersonal perspective may very well tell us something entirely different, something that our more-personal selves probably will not always be best pleased to hear. He places this fluid and

sometimes uneasy relationship between our personal and impersonal selves upon a continuum of perspectives from any point of which we might address a moral problem, presumably with results varying in being more or less impartial, – a quantitative approach utilizing a geographic metaphor.

Francis Bacon cautions scientists against the "Idols" of the mind that can sway even our presumably objective scientific exploration; these "Idols" must be ever more active in moral deliberation, the rightness of whose outcomes are not in principle as transparent as are those of science.[4] Between the influence of "Idols" and conscious, dispositional, or habitual choices of impartiality-style, it is likely that moral deliberation will sometimes result in outcomes that to the immediate perception of ordinary bystanders (with their own different styles), seem to be not impartially arrived-at at all, although steps toward impartiality may have been taken in good faith and in all sincerity.

For example, assumptions about what a moral community ought to look like may be embedded in a particular impartiality style, whether chosen consciously or unconsciously; adopting an impartial stance in moral deliberation may lead to a perspective that is either more or less objective or inclusive. A determinedly objective stance, one relying for example on ideas of detachment and impersonal equality, may generate (or be generated by) a moral community far different in character than one that would be generated by (or would generate) a consciously inclusive stance relying on ideas of sympathy and benevolence.[5] And yet, moral deliberators using either of these styles are likely to be absolutely persuaded that they are deliberating impartially.

On the other hand, since steps toward impartiality were in fact taken, and assuming a functional style of impartiality was in fact engaged, can it be argued that although outcomes may be different for different styles, perhaps even mutually exclusive

or contradictory, and minimizing to the greatest extent possible the subversive influence of "Idols," nevertheless these *were* in fact impartial decisions, the best that we can hope to accomplish? But in that case, our understanding of what it means to deliberate impartially is going to need some reexamination, if a goal of philosophizing about morality is to defend the possibility of an outcome that "any reasonable person" could agree to or one that would be objectively right (requiring in its turn a specification of what is meant by "objectively right").

I shall agree with its defenders that impartiality is not in fact just another "Idol," that it is indeed the *sine qua non* of moral deliberation; but I shall identify three broad styles of impartiality, each of which can lead to a different outcome in moral decision-making: sympathetic impartiality, detached impartiality, and Kantian impartiality. These three styles are distinct but not mutually exclusive; although moral theories taking these styles of impartiality as their hallmarks may vary greatly among themselves in their structural emphases, it will be secn that outcomes of moral deliberation within these frameworks are not dissimilar. Moreover, similar psychological mechanisms are used to attain the targeted states of impartiality.

I will show that there is indeed a continuum of impartiality, but that it is not quantitative but categorical in nature; one begins at pretty much the same point to attain impartiality and travels along the same psychological pathway toward its attainment, but the "stopping places" along the way yield different *categories* of impartiality, not different "amounts." In moral deliberation, the place along the continuum where one stops is emblematic of the deliberator's theoretical commitments and will determine what decisions he ends up with in a deliberative situation. Certain assumptions may be identified underlying each style – assumptions about the nature of the human person,

what it means to be in relationship to other persons, how best to structure social systems to enhance the differing moral ideals that are thus generated, and so on.

Operating from a premise that a moral community ought to be as inclusive as possible and targeting examples of moral deliberation to that goal, I shall argue that each of these three broad styles has certain problems in its own way, but that two of them lead to outcomes of almost certain exclusion from the moral community,[6] and only one of them holds the promise of overcoming the obstacles to a community that is broadly inclusive. I make the initial assumption that an inclusive community is desirable compared with an exclusive one, using as my justification for this assumption the explicit or implicit justification of many existing moral systems: that one is urged to love one's neighbor as oneself and that the particular moral system under discussion (e.g., Hobbes' or Locke's) satisfies this demand.

Is this a case of taking a Biblical injunction as the foundation of morality, *as* a Biblical injunction, and then trying to shape discussion to satisfy it because of a partisan bent toward theological strictures? It is not; the injunction to love one's neighbor as oneself is, however, a useful kind of shorthand that expresses the requirement that all moral theory should take account of both inclusivity and equality as foundations of morality.

Talking about impartiality makes this clear. It is clear that before one can even use any modality of impartiality, one must understand in general what is being asked for. Given that the ordinary use of impartiality in moral decision-making occurs between or among persons and their situations, the default position of the deliberator at the inception of impartial deliberation should be one that includes *all* participants in the deliberative situation, and includes them *as equals*. In discussions of larger moral problems, ones that

may influence social structures, for example, deliberators should also have the same default position: to include as a starting-point *all* the parties affected, as well as those who are not in a position to deliberate, *as equals*. Otherwise, as a matter of logic, impartiality loses its function, which is to treat like cases alike, from a disinterested perspective (however that is to be understood or established), and to give reasons if they are not to be so treated. Excluding someone from consideration *ab initio*, by treating some interested party as insignificant to deliberation – as a moral non-equal – then makes "impartiality" incoherent; one might as well not bother.

This does not mean that the *outcomes* of deliberation will be or ought to be either completely inclusive or render persons completely equal in other kinds of status, for example, economic or social. What it does mean is that impartial deliberation must start from that assumption in order to satisfy the purpose of initiating impartiality in the first place. The problem of some of the styles of impartiality that will be discussed is that they are prone to end up with unequal, exclusive outcomes not because of the outcomes of straightforward impartial deliberation, but because deliberation did not in fact begin from an impartial place.

In addition to the logic of impartiality, there is a practical matter as well; for impartial decision-making to yield the very best decisions possible (however "best" is to be interpreted), it stands to reason that there ought to be as much information as possible available to the deliberator. Eliminating potentially interested persons from consideration eliminates a great deal of information as well, so outcomes of deliberation may well reflect this weakness.

It may seem tendentious or trivial to think that interested persons will be, or even can be, excluded or treated as less than equal in impartial deliberation, because the term

"impartiality" already carries a great deal of noble weight. One is aware, in striving for impartiality, that one is searching for a "higher" frame of mind than one concerned strictly with one's own business. But for someone who is aware of trying to attain an impartial frame of mind, this very awareness of its nobility can mask the prejudices and biases that render him unconsciously partial or exclusive from the outset. It is these hidden biases that can subvert attempts at impartiality, from the commonplace moral decision-making of the ordinary person all the way to the construction of elaborate moral or social theoretical structures and systems that rely on conceptions of impartiality that unwittingly exclude and do not accept the equal moral status of persons, even while an impartial frame of mind is sincerely sought. While it is certainly true that as humans, we may never be entirely free of such partialities and biases, and a perfect impartiality, however it is understood, is in principle beyond us, the point is that there are different ways of being impartial, some of which are more prone to keep biases such as these well hidden from the deliberator, and some that are more likely to force a confrontation with those biases in order at least to acknowledge them, if not to overcome them, which may be the best we can do. Differing understandings of impartiality then are prone to yield differing outcomes of deliberation.

That differing outcomes are in fact commonplace occurrences depending upon the style of impartiality in play can easily be seen by considering an ordinary example. The outcome of a jury deliberation in a criminal trial is likely to be different depending upon whether detached or sympathetic impartiality is the dominant deliberative modality in force. Detached impartiality, in its broad outlines, involves consciously separating oneself from personal interests and emotions in order to achieve the status of a more "general" person, for the purpose of being able to assess a situation more objectively, removed by a

greater distance from one's own biases and partialities. Sympathetic impartiality, broadly speaking, involves on the other hand seeing or knowing the world from another individual's partial-to-self perspective. Typically, it is a specific person whose perspective is borrowed, but it can be a class or group of persons as well. A jury is asked to consider the facts presented in evidence and to make reasoned interpretations of how and whether these facts will be reflections of truth; these interpretations are ideally to be arrived at using methods of interpretation such as deductive and inductive logic and cognitive techniques of analysis, synthesis, assessment of probability, and so forth – deliberative styles characteristic of detached impartiality. However, when the outcomes of inductive and deductive processes appear to favor one side over another, the losing side may well try to elicit the jury's sympathetic identification with either the defendant or the victim. In a situation such as this, a subtle advocate might even be able to influence the outcome of deliberation simply by having correctly assessed the psychologically preferred deliberative modality of the majority of the jurors ahead of time, perhaps even during jury selection, tailoring presentation of the case to that modality. Sympathetic identification with a defendant, or with the group of people that a defendant is seen as representing, for example, may easily trump detached assessment of evidence, as was seen recently in a highly publicized criminal trial.[7]

I shall explore these three distinct styles of impartiality, and I will argue that the choice of style (which is typically not even consciously experienced as choice) depends in the first instance upon pre-existing cultural, personal, or intellectual biases which tend by and large to lead one to an outcome already decided upon, pure impartiality being relatively rare, if it can be said to exist at all (and undesirable if it does exist, as will be seen). I will further argue that different styles of impartiality are likely to lead to differing

outcomes in moral deliberation, and that furthermore, within the same modality, different outcomes are also possible, depending upon psychological capacities and hidden philosophical presuppositions. I will argue that nevertheless, in the final analysis, it is the best we have, and with due caution and care for the influence of the "Idols," a reasonable approximation to objective moral decision-making is indeed possible, even in spite of the fact that outcomes may differ. Last, I will argue that two of the three of these styles of impartiality –sympathetic and detached impartiality – render us particularly prone to "Idolatry" and that the remaining style, "Kantian" impartiality (including an example of a sympathetic utilitarian theory), offers us the best hope of achieving reliably decent outcomes, given the essential roughness and imperfection of the moral decision-maker.

The command that one must love one's neighbor as oneself is not intended to stand for or encompass all elements of moral philosophy, but it is true that this imperative is either explicitly adduced or at least implied by a great many of the writers on moral issues whose theories I will be discussing. At the same time, it can and does highlight a number of the problems of impartiality, and thus of moral deliberation, and provides a reasonably small and narrow framework within which to examine these problems in outline without becoming distracted.

I will take "moral deliberation" in its most general sense and not concern myself with specific levels or purposes of deliberation. My concern is to look at kinds of impartiality across varieties of instances and to try to discern how ordinary, conscientious moral deliberators might best direct themselves to achieving a workable state of impartiality. Such a workable state might be one that satisfies three conditions: it permits a sense of having arrived appropriately at a decision; it recognizes and respects the equality of persons; and it takes due account of the unsurprising fact that no human

person can hope to achieve "perfect" impartiality and objectivity.

Styles of Impartiality

What follows is a brief outline of what is involved in these three broad categories[8] of impartiality and which moral theories will be discussed in order to highlight some of the difficulties of each one.

1. Sympathetic Impartiality

In the dance between personal and impartial perspectives, sympathetic identification is an attempt to take in the personal perspectives of others, whether one other or a whole society of others, in order to "see" the world as another person or group of persons might see it. It is a process of impartiality that is predicated on the belief that such sympathy is already present in the first instance between naturally benevolent persons and that it is to be cultivated in order best to regulate the differences between one's own and others' personal perspectives. It is here that we find concepts such as the God's-eye view and the benevolent sympathetic ideal observer. In its ideal form, it depends upon the concept of a kind of "super-person" who is omniscient about all the aspects of a situation and all the interests of the persons involved (and therefore, I will later argue, outside of morality altogether). In the historical survey that follows, I will examine sympathetic identification through the writings of Butler, Hume, Smith, and Firth, and I will reserve discussion of a contemporary theory of sympathetic utilitarianism, that of Hare, for the chapter on Kantian theories. It will be seen that in this modality there is a strong tendency to end up with exclusive, not inclusive, outcomes.

2. Detachment

Detachment is an attempt to separate and detach oneself from one's own personal perspective in order to achieve a more impersonal view; it is often the style of impartiality called upon when the goal is to establish structures of just practices in various human institutions such as the design of basic social systems, adjudication in a courtroom, or third-party arbitration in personal disputes. It is predicated upon the presumed increasing ability of persons to agree with one another about controversial issues the farther away they are removed from their personal interests, presuppositions, biases, and desires – there is an implicit assumption of the existence of an objectively right decision which rational persons may approach and agree upon if they act in good faith to attain detached impartiality. "Objectively right" in these contexts is often defined as that which most reasonable persons in the proper frame of mind would agree to. In its ideal form, it can abstract an impartial deliberator into a kind of non-personal entity separated from anything even remotely personal in moral deliberation. The original goal of justice can then become thinned to not much more than assessment of the optimal effectiveness of institutions to garner at least tacit assent from their participants; in those cases, rationality can be construed as merely the most effective means to self-interested ends, even if self-interest is broadly conceptualized to include benevolence and justice. It will be seen that the participants in deliberation, particularly in this ideal form, tend to have been tacitly preselected to assure such assent and that there is a tendency to play upon prejudices and fears concerning the "other" in order to facilitate this preselection by rational persons. I will explore these and other ideas in two chapters on the writings of various contract theorists; one chapter will survey impartiality in what I will call "egalitarian" contract theories (Hobbes, Rousseau, and Rawls), the other will look at how

impartiality plays itself out in what I term "Lockean" contract theories (Locke, Gauthier, and Nozick).

3. Kantian Impartiality

Kantian impartiality is difficult to describe, but a convenient way to start is to recognize it as an effort to step aside from points of view altogether in an attempt to tap into the self-legislative property of persons. It seeks that commonality among humans and other rational creatures[9] that constitutes us all as necessarily part of a moral community by virtue of our rationality and seeks as its goal to describe and actualize fully that commonality in order to stimulate our recognition of ourselves and others as part of that community; in Kant's terminology, as part of the Kingdom of Ends. It can be seen as the underpinning of detached impartiality, where detachment would go if it were to explore its own foundations, in that the implied objectively right answer to a moral problem is not to be recognized solely through reasonable agreements but can be pursued further, to be uncovered through the disciplined understanding of the essence of moral personhood. The moral law is within, not without, and is objective nonetheless. Theories of sympathetic impartiality would also find similar assumptions underlying them, with the commonality in question being the human feeling nature, that which leads persons to "do by nature the things contained in the law."[10] In an ideal form it can easily be misconstrued as verging on aridity and formality and can encourage a kind of rule-rapture or fanaticism that ends by being antithetical to the original goal of membership in the moral community. Kant is particularly prone to being misunderstood in this way. Rawls' "Kantian Interpretation," although it uses the language of reasonable agreement, nevertheless emphasizes a conception of moral personhood that takes the interpretation

beyond simple agreement under conditions of detachment. Several interpretations of Kantian ethics will be discussed, as will Hare's utilitarian theory. It may seem peculiar to include a utilitarian theory in the same place and for similar reasons as Kantian theories, particularly since it is a theory based on sympathetic impartiality; but Hare himself calls attention to the similarities between Kant's ideas and those of "a certain kind of utilitarianism,"[11] including his own.

4. "Idols" Again

Finally, turning the analysis of styles of impartiality back to the individual moral deliberator in everyday life, I shall offer the view that both sympathetic and detached impartiality are particularly prone to activate various "Idols," with the dangerous additional problem of making these idols even more invisible to introspection than usual, since the moral deliberator is self-consciously aware of having made the required effort at impartiality and believes himself *ipso facto* to have overcome most significant idolatrous tendencies. Kantian impartiality, which relies neither upon the sympathy and benevolence of the deliberator nor upon his ability to separate adequately from his personal interests, offers the best hope for as true an impartiality as essentially self-interested but genuinely reasonable and moderately well-meaning humans can aspire to, particularly when interpreted through the necessary "situatedness" of human persons. I shall argue that this is because not very much in the way of cognitive or emotional effort is required to achieve an adequate level of this sort of impartiality, provided one has had a reasonably normal upbringing in a reasonably decent society.[12] The attitude towards persons and situations generated by the primary demand of this kind of deliberation, to treat persons as ends, allows for appropriately impartial decision making in situations

structured so that impartiality is explicitly expected, such as sitting on a jury, as well as for different decision-making procedures in situations where a formal, overt use of impartiality would be out of place, such as in family situations. Both, however, are informed by respect for persons, and so both reflect the kind of impartiality that such a respect engenders, one based on the recognition of persons' moral equality. Such a level of impartiality is within the reach of most ordinary people, regardless of their level of education; it neither encourages nor requires perhaps-unreachable motivations of sympathetic benevolence or detached disinterestedness, and tends to accommodate in a perfectly natural way most of the idolatrous tendencies that are as present within us as is the moral law.

NOTES

Chapter One

[1] e.g., M.G. Singer,*Generalization in Ethics* (NY: Russell & Russell, 1971 [1961]), 49.

[2] It may also reflect differing levels and purposes of deliberation, according to Cynthia Stark. The demands of impartiality in justification procedures may differ greatly from those of decision procedures. "Decision Procedures, Standards of Rightness, and Impartiality," *Nous* 31:4 (1997), 478-495: 479.

[3] e.g., Thomas Nagel, *The View from Nowhere* (NY: Oxford University Press, 1986). Also from Nagel: *Equality and Partiality* (NY: Oxford University Press, 1991).

[4] Francis Bacon, *The New Organon* (NY: The Liberal Arts Press, 1960 [1620]), 44-62. The Idols of the Tribe reflect common human cognitive and intellectual tendencies: tending to impose more order upon things than there actually is; liking affirmatives rather than negatives; appreciating rapidity and simultaneity; resisting intellectual limits; the mind not liking what it can't see, yet reaching for abstractions. Idols of the Cave influence what and how we study the world by imposing our favorite frameworks around pieces of it. Idols of the Market Place reflect our tendency to use words as common currency, joining names and things in alliances whether the alliances are proper or not; naming things that may not even be there simply because they are framed to the capacities of the "vulgar." Finally, the Idols of the Theatre reflect the received wisdom in residence in our minds, whether from philosophical systems, superstition and theology, or insufficient empirical observation leading to a rush to judgement.

[5] However, we shall see later that sympathy may actually work against inclusion, and a too-great attraction to equality may lead to undesirable totalitarian outcomes.

[6] "Inclusion" and "exclusion" in this exploration refer to the recognition or non-recognition, respectively, of the equal moral status of persons *as* persons. This clearly has concrete outcomes for specific individuals of actual inclusion in or exclusion from whatever community is under consideration; but for the purposes of looking at various styles of impartiality these terms will remain at a metaphysical level. "Community" is to be understood the same way – analogously to a Kingdom of Ends.

[7] The O.J. Simpson trial.

[8] The term "category" is more suitable when discussing moral theories, while "style" is more apt when discussing what individual deliberators actually do; however, for general purposes these terms will be used interchangeably with "modality."

[9] Kant, for one, explicitly refuses to limit rationality to humans alone: "... the command, 'Thou shalt not lie,' does not apply to men only, as if other rational beings had no need to observe it." I. Kant, *Foundations of the Metaphysics of Morals*, trans. Lewis White Beck (Indianapolis, IN: The Bobbs-Merrill Company, Inc., 1959 [1785]), 5.

[10] Joseph Butler, *Five Sermons* [1726], ed. Stephen Darwall (Indianapolis, IN: Hackett Publishing Company, 1983), Sermon II, 34.

[11] R. M. Hare, *Moral Thinking: Its Levels, Method, and Point* (Oxford, U.K.: Oxford University Press, 1981), 4 – 5.

[12] See, e.g., Rawls, *A Theory of Justice* (Cambridge, MA: The Belknap Press, 1971), §69.

CHAPTER II

SYMPATHY, SYMPATHETIC IDENTIFICATION, BENEVOLENCE, AND THE IMPARTIAL SPECTATOR

A minimal effort at impartiality, and, it is argued, a natural one,[1] is the attempt to see a situation from the perspective of another; to try to adopt the point of view of someone else involved in a situation and then to note, or even enter into, whatever perceptions, emotions, and ideas follow from taking that perspective; these then become material for a more broadly considered judgement of the situation and offer the opportunity for potentially conflicting perspectives to be mediated impartially, in a way that is not available from a singly-oriented point of view. In the psychological mechanisms of sympathy, however, interesting questions of identity and self-deception can come into play, which can influence the outcomes of moral deliberation in one direction or another.

Taking on another's perspective, or "role-taking" in Kohlberg's term,[2] is accomplished through a modality of sympathetic identification. One takes on and takes in the perspective of another in order to "see"[3] the problem as it it seen by the other. John Rawls explains how, in a well-ordered society, role-taking may develop at an increasingly complex level as the number and variety of one's roles in the society expand.[4] But seeing-as-other is not as straightforward as it sounds. The very idea of seeing "as" another might see is difficult to understand. In addition, there are different *ways* of seeing-as-other and there are different conceptions of who the other might *be*. A variance in

interpretation of any of these elements alters perception and can lead to differing decisions by the decision-maker within the modality of sympathetic identification itself.

When one looks at the world from another's point of view, what exactly is going on? Is it still oneself, but looking mentally out through the other's eyes? In that case, one must still be oneself, with all one's own personal interests, passions, and beliefs intact, but having added in, to the extent of one's ability and willingness, the other's interests, passions, and beliefs. Leaving aside the trivial truth that no one can wholly see the world through another's eyes, no matter how sympathetic or empathetic, it is still important to wonder just how far into another's perspective one actually can or wants to or ought to get, for ability and willingness present yet other problems. For these reasons, one can make the effort to interpret the world from within the other's cognitive universe, but one might not ever be able to be sure of the extent to which two world views are being intermingled and the extent to which one's own motivations are not the operative set.[5]

On the other hand, one may certainly make a conscientious effort to leave oneself out of it in order more fully to take on and take in another's world view. For Adam Smith, the matter is straightforward; it is not I who is in your place, but I-as-you; our persons have been exchanged:

> ...though sympathy is very properly said to arise from an imaginary change of situations with the person principally concerned, yet this imaginary change is not supposed to happen to me in my own person and character, but in that of the person with whom I sympathize. When I condole with you for the loss of your only son, in order to enter into your grief I do not consider what I, a person of such a character and profession, should suffer, if I had a son, and if that son was unfortunately to die: but I consider what I should suffer if I was really you, and I not only change circumstances with you, but I change persons and characters.[6]

However, this suggests that sympathetic identification can only occur if the other's world view is somewhat familiar (if one in fact knows what the other's "person" actually is – his interests, passions, beliefs, and character), and thereby casts doubt upon the reliability of this exchange of persons. There are also contemporary ideas along those lines, for example that in principle sympathetic engagement can be possible only if one shares the other's cultural and cognitive assumptions.[7] This suggests that even assuming this exchange of persons to be the case, the two persons can become confused with one another because of their mutual familiarity and easy exchangeability, thus leaving the door open to self-deception; moreover the exchange can be only temporary within the frame of sympathetic identification and one must return fairly quickly to oneself in order to proceed with decision-making, perhaps not even consciously aware of having made the return. Having previously either imported the other's world view into one's own and set one's own aside, or having "entered by sympathy" into the sentiments of the other either as oneself or as the other,[8] now one must try to weigh and balance the two perspectives, incorporating them somehow into a decision that one makes as an agent oneself, not as the substituted person of the other.

For these and other reasons, it would be important to mark out carefully who the other is, indeed who the other *must* be, in the modality of sympathetic identification at this level. It must be someone sufficiently like myself so that I can make one of the sympathetic movements noted above, and this means that it is imperative that I also know well enough who "myself" is. Hume makes the point that our capacity for sympathy is intimately linked to our "lively ... conception of our own person",[9] and Butler explicitly pins the capacity for benevolence toward others, facilitated by the

mechanism of sympathy, upon "cool self-love."[10] But "knowing myself" also has its pitfalls, because my "self" can have many layers, which could theoretically all be peeled away to leave what might be termed my "essential" self. Most of us do not get that far, remaining in the realm of the safe and familiar characteristics by which we identify and characterize ourselves to ourselves and to others.[11] Therefore, if I am an English gentleman and remain so during my moral deliberation, I can see the world adequately enough through your eyes, should we get into a conflict, if you are also an English gentleman and I see you as such. If I am an American suburban housewife from Milwaukee, I can do the same if you are also an American suburban housewife from Milwaukee. But if you are alien to me, for example if you are an American suburban housewife and I am an English gentleman, I may not be able adequately to accomplish the movement from my perspective to yours and back again and it is arguable that I would even be motivated to do so, if we were in a conflict.[12] Therefore, on this level, in order to become engaged at all in my sympathetic identification with you, the other, I necessarily must see you as very similar to myself, whether you actually are or not. My movement is toward myself-as-other, or toward the other-as-myself. Smith is increasingly explicit here, although it considerably modulates his previous exchange-of-persons idea, and it will be seen later on that what he says here expresses a flaw in the theory of sympathy that is very difficult to circumvent:

> Every faculty in one man is the measure by which he judges of the like faculty in another. I judge of your sight by my sight, of your ear by my ear, of your reason by my reason, of your resentment by my resentment, of your love by my love. I neither have, nor can have, any other way of judging about them.[13]

With this, he indirectly makes the point that if our sympathy is not engaged, for a variety of reasons, with the point of view of another, we tend instead to impose our own perspective and judge the other as deficient or excessive (or irrelevant) from our own point of view. Rather than considering a point of view so different from ours as legitimate, but temporarily out of our reach pending a sympathetic engagement, we instead judge the other by standards of propriety and impropriety arrived at by our sympathies with our own selves in a like situation:

> In the suitableness or unsuitableness, in the proportion or disproportion which the affection seems to bear to the cause or object which excites it, consists the propriety or impropriety, the decency or ungracefulness of the consequent action. ... When we judge in this manner of any affection as proportioned or disproportioned to the cause which excites it, it is scarce possible that we should make use of any other rule or canon but the correspondent affection in ourselves.[14]

So it appears that the exchange of persons can only proceed if sympathy is already engaged. This may seem an obvious and trivial point, but at this level it constitutes a serious defect in the psychology of impartial moral decision-making by the modality of sympathetic identification since my sympathy can, on this view, only be engaged by someone similar to myself and, moreover, as this quotation suggests, only by someone I already approve of. Fortunately, as we shall see, sympathetic identification need not be limited only to those who are so familiar to us.

It is useful at this time to review some issues that are involved in the psychological process of "identification." If sympathetic identification is indeed the modality that allows me to take in another's perspective, how does it come about? Is the

motivation for trying for such a state of mind natural (or innate or instinctive) or is the desire for it learned by socialization? Is the state of mind itself a natural function that operates independently of my will, similarly to the operation of an instinct? What is the relationship, if any, between identification and imitation? Can I successfully imitate or model one who is alien to me? How successful can I hope to be? The question of how far, in the context of moral decision-making, sympathetic identification can or ought to go, will be explored in some depth as the various patterns of moral theory unfold.

Hume gives some preliminary insight into the mechanisms of the identity problem with which these issues are intimately linked. He tells us that we may receive from others, by "communication," their "inclinations and sentiments, however different from, or even contrary to, our own," and, as noted earlier, links this to our "lively ... conception of our own person." So the first movement must be a clear-enough awareness of who we are, in order for identification to proceed at all. This link between our lively conception of ourselves and the capacity for sympathy is due to nature's having "preserved a great resemblance among all human creatures," since "we never remark any passion or principle in others, of which, in some degree or other, we may not find a parallel in ourselves." This resemblance contributes to "make us enter into the sentiments of others, and embrace them with facility and pleasure." He continues,

> The stronger the relation is betwixt ourselves and any object, the more easily does the imagination make the transition, and convey to the related idea the vivacity of conception, with which we always form the idea of our own person. ...

and stresses the necessity of contiguity, whether of blood or of acquaintance, to make this work effectively:

> ...we must be assisted by the relations of resemblance and contiguity, in order to feel the sympathy in its full perfection.[15]

Sympathy arises from nature: birds of a feather flock together

> ...by a certain sympathy which always arises betwixt similar characters ... [which resemblance] operates after the manner of a relation by producing a connection of ideas.[16]

The smooth elision from the "natural conformity" of one human being to another to the therefore-assumed natural conformity of their psychological attitudes and cultural and intellectual assumptions is the fly in the ointment of many theories of sympathy.

Kohlberg arrives at a similar point, but from an explicitly socially-based etiology for sympathetic identification, which he calls "role-taking" -- taking the perspective of another. He describes role-taking as "... a trend to structure imitative processes in terms of conceptions of structured roles ... [i.e.,] of categories of persons in defined relations to one another with normatively defined functions." These role conceptions themselves depend upon "concrete operations, that is, the logic of classes and relations."[17] He locates the "enduring [human] tendency to model"[18] in an Aristotelian framework: imitation is not instinctual, reinforced, or conditioned, but intrinsically motivated because it is interesting; therefore it is cognitively based but not due to fixed intra-organismic needs.[19] In addition, the relation of mastery between an act (imitation) and its effects (role-taking) leads to "assimilation" -- the establishment of the external event's relation to one's own behavior.[20]

A particular kind of assimilation, called attachment, involves three things: similarity to the other, love or altruism toward the other ("sympathy"), and presupposition of self-love -- "... striving to satisfy another self presupposes the capacity to satisfy one's own self -- which is loved intrinsically and not instrumentally."[21] And the

> primary meaning of the word social is the distinctively human structuring of action and thought by role-taking, by the tendency to react to others as like the self and to react to the self's behavior from the other's point of view.[22]

The emphasis of Butler, Hume, Smith, and Kohlberg on the primacy of self-love for the structuring of sympathy with others is absolutely essential to impartiality and will be examined in greater detail below.

Butler, Hume, Smith, and Kohlberg, arriving from different directions, agree that similarity or resemblance motivates the initial movement toward sympathy. But moral decision-making often requires a taking-in or taking-on of the perspective of someone who is not necessarily similar to ourselves, and "similarity" or "resemblance" sound like they could impose a restraint on sympathetic identification so that one might in principle be unable to break out sympathetically from the closed circle of persons like oneself. Here resides the consternation of the ordinary soul when confronted with the injunction to "love thy neighbor as thyself" when the realization comes home that my neighbor can, by definition, be literally anyone.[23]

Recognition of this restrictiveness, combined with proper motivation, makes it clear that there must be a more generalized other -- not "you" in particular, but anyone who could be you, or me for that matter. However, the problem of familiarity remains.

"Anyone" must still remain at some level of familiarity, perhaps cultural or intellectual, unless a movement of some sort is made to break out of the confinement of the familiar. It is clear that something more is needed than local good-faith sympathy to transcend the constraints of familiarity in order to stimulate the motivation to try to see the world through the eyes of truly *any* other.

At this point, two related ideal concepts of impartiality traditionally brought in to effect this breakout are the "God's-eye view" and the "ideal observer" or "impartial benevolent spectator." Sympathetic identification is still the modality in use, but there is a difference in degree between the sympathetic identification required for these concepts and that required for the Englishman and the housewife. When I try to take on the God's-eye view or become an ideal observer, I need to take on the perspective of *all* others, not just my contextual peers; my sympathetic identification should be with the points of view of everyone. We can consider such an ideal observer as the very exemplar of impartiality, embodying a great many of its concepts. Kurt Baier puts it this way:

> ... we look ... from the moral point of view, that is, that of an independent, unbiased, impartial, objective, dispassionate, disinterested observer. Taking such a God's-eye view...[24]

Acts, 10:34, interprets this idea from a theological perspective, but it is clear what is intended nevertheless:

> Then Peter opened *his* mouth, and said, 'of a truth I perceive that God is *no respecter of persons*: But in every nation he that feareth him, and worketh righteousness, is accepted with him.'[25] [emphasis added]

And Mill writes,

> As between his own happiness and that of others, utilitarianism requires him to be as strictly impartial as a disinterested and benevolent spectator. In the golden rule of Jesus of Nazareth, we read the complete spirit of the ethics of utility.[26]

One who can take on the role of this ideal empath presumably will choose morally, wisely, and unerringly, since he or she knows the world from all viewpoints; this person is not disinterested, but is rather fully and completely interested -- has taken on or taken in all persons' interests together:

> [e]ndowed with ideal powers of sympathy and imagination, the impartial spectator is the perfectly rational individual who identifies with and experiences the desires of others as if these desires were his own.[27]

The common criticism of mere humans being unable in practice and in principle to achieve such status is beside the point and far too literal; it is the *movement toward* incorporating multiple viewpoints, even, by imaginative means, those that are unfamiliar, that counts, and that permits a new conception of the other. It is important to construe the ideal observer as being possible, not actual -- as reflective of a humanly possible state of mind rather than of actual superhuman capacities. In the following exploration of the development of the nature and status of the impartial spectator through sympathetic identification in various moral-philosophical writings, Roderick Firth will provide the most detailed analysis, both moral and moral-psychological, in support of this mode of

moral deliberation. (A somewhat different conceptualization of the operations of sympathy will be found in the discussion of Hare's "Kantian Utilitarianism" in Chapter VII.)

A selective survey of the moral aspects of sympathy and sympathetic identification, beginning with Butler's idea of the relationship between self-love and benevolence, will provide a closer look at the implications and possible outcomes of using this mode of impartiality. One of the more intriguing elements of the theory of sympathetic identification as a tool for moral deliberation is the supposed naturalness of such a process in the human person; but we will see how natural sympathy can operate to close off a moral community, presumably an unintended outcome from a moral point of view. In following the theory of the sympathetic impartial spectator through various writers, we can see that in its most fully developed form, the sympathetic impartial spectator can actually render morality itself almost superfluous.

NOTES

Chapter Two

[1] e.g., Butler, Hume, Smith, et al.

[2] Lawrence Kohlberg, *The Philosophy of Moral Development* (San Francisco: Harper & Row, Publishers, 1981.

[3] The "visual" terminology here is not to be taken concretely. By "seeing" I comprehend not literally having a visual image and interpreting it, but knowing and feeling, all of which enters into taking in or taking on another's perspective in this sense. The language of vision is most evocative in these matters, perhaps because, as Socrates notes, vision is indeed the "keenest mode of perception" (*Phaedrus* 250d). Adam Smith speaks of "picturing out in imagination" to convey this sort of process in *The Theory of Moral Sentiments*.

[4] Rawls, *Theory of Justice*, § 71.

[5] For a deeper discussion of this problem, see the section on Hare in Chapter 7.

[6] Adam Smith, "The Theory of Moral Sentiments," [1759] in *British Moralists*, ed. L.A. Selby-Bigge (Indianapolis, IN: The Library of Liberal Arts, 1897), § 339.

[7] For example, Alasdair MacIntyre argues that all inquiry is steeped in intellectual "traditions of inquiry" and that it is impossible to adopt the perspective of another tradition unless one learns its intellectual language as a second first language. But this principle may hold true for even more local "translations" within the same culture, even within the same neighborhood, if one isn't careful and imaginative. Alasdair MacIntyre, *Whose Justice? Which Rationality?* (Notre Dame, IN: University of Notre Dame Press, 1988) 348 and elsewhere.

[8] Smith, *Theory/Selby-Bigge*, §319.

[9] David Hume, *A Treatise of Human Nature* [1739-40], Book II, Part, Section XI: "No quality of human nature is more remarkable, both in itself and it its consequences, than that propensity we have to sympathise with others, and to receive by communication their inclinations and sentiments, however different from, or even contrary to, our own. ... men of the greatest judgment and understanding ... find it very difficult to follow their own reason or inclination, in opposition to that of their friends and daily companions. ... It is evident that the idea, or rather impression of ourselves is always intimately present with us, and that our consciousness gives us so lively a conception of our own person that is not possible to imagine that any thing can in this particular go beyond it." (Garden City, NY: Dolphin Books, 1961), 287-88.

[10] Butler, *Five Sermons*, Sermon I. Butler begins, as always, from Scripture: "...we, being many, are one body in Christ, and every one members one of another." (Romans XII: 4,5), and proceeds with a meditation and argument concerning human nature. The outline of his argument is that "... men are so much one body that in a peculiar manner they feel for each other..." (§ 10), and therefore the "...natural principle of *benevolence* in man ... is in some degree to society what *self-love* is to the individual." (§ 6) [emphasis in text]. While Butler does not use the express term "sympathy," his meaning in the first Sermon is clear.

[11] For another point of view concerning the "essential" self, see the discussion of Zeno Vendler in Chapter 7.

[12] Butler indirectly supports this point when he says, "There is such a natural principle of attraction in man toward man that having trod the same tract of land, having breathed in the same climate, barely having been born in the same artificial district or division, becomes the occasion of contracting acquaintances and familiarities many years after..." Sermon I, § 10.

[13] Adam Smith, *The Theory of Moral Sentiments* (NY: Augustus M. Kelley, Publishers, [1759] 1966), 18.

[14] Smith, *Theory/Selby-Bigge*, §§265-267.

[15] Hume, *Treatise*, II, 1, xi.

[16] Ibid., II, 2, iv.

[17] Lawrence Kohlberg, *The Psychology of Moral Development* (San Francisco: Harper & Row, Publishers, 1984), 115.

[18] Kohlberg, *Psychology*, 109.

[19] cf. Aristotle, *Poetics*: Although Aristotle declares the imitative tendency to be "natural to man from childhood," he stresses that it is delightful because it is at the same time "learning -- gathering the meaning of things." (1448b5-20). Rawls notes this tendency in what he terms the Aristotelian Principle: "...other things equal, human beings enjoy the exercise of their realized capacities ... and this enjoyment increases the more the capacity is realized, or the greater its complexity." (*Theory of Justice*, §65) And Hume also supports this idea: "When the soul applies itself to the performance of any action, or the conception of any object to which it is not accustomed, there is a certain unpliableness in the faculties, and a difficulty of the spirits moving in their new direction. As this difficulty excites the spirits, it is the source of wonder, surprise, and of all the emotions which arise from novelty, and is in itself very agreeable, like everything which enlivens the mind to a moderate degree." (*Treatise*, II, 3, v.)

[20] Kohlberg, *Psychology*, 123-125.

[21] Ibid., 155.

[22] Ibid., 141.

[23] In Butler's exegesis of these texts, he suggests that the command to love our neighbor was given to compensate for our not being able to love the whole universe; loving our neighbor seemed more manageable. Joseph Butler, *Butler's Fifteen Sermons*, ed. T.A. Roberts (London: S.P.C.K., 1970), Sermon 12 (#5 in Darwall's edition).

[24] Kurt Baier, *The Moral Point of View: A Rational Basis of Ethics* (Ithaca, NY: Cornell University Press, 1958), 210.

[25] *The Holy Bible, Michelangelo Edition: Containing the Old and New Testaments in the Authorized King James Version* (NY: Abradale Press, Publishers, 1969).

[26] John Stuart Mill, *Utilitarianism, Liberty, and Representative Government* [1861] (London, J.M. Dent & Sons, Ltd., 1947 [1861]), 16.

[27] Rawls, *Theory of Justice*, §5.

CHAPTER III

A SURVEY OF SOME SYMPATHY THEORIES

Joseph Butler

Although Butler does not use the idiom of the impartial spectator, it is clear that for him such a function would be carried out by the faculty of "conscience, or reflection," which is the regulatory mechanism for expressions of cool self-love and its offspring, benevolence operating through sympathy. Butler is concerned to argue, in Sermon I, that benevolence toward others is merely the faculty of self-love writ large, since we are all members of one body. As members of the same body, we experience sympathy in a very natural way toward others, which is expressed in acts of benevolence. Again, while not expressly using the idiom of sympathy, he speaks instead of the "correspondence between the inward sensations of one man and those of another". The regulatory mechanism of conscience "tends to restrain men from doing mischief to each other, and leads them to do good, [which] is too manifest to need being insisted upon."[1] In his description of the economy of the organization of human passions, affections, and principles of action, Butler reflects a Platonic scheme of potentially powerful and unruly natural faculties kept under the severe and incontrovertible authority of this superior faculty.

Butler's concern was to counteract what he considered the dangerous and incorrect ideas developed by Hobbes, "a man of learning," in his "grave book upon human nature,"[2]

Leviathan, specifically that the human person is not by nature ethical but instead deeply and only self-regarding; that this self-regard is in fact what generates morals and that it does so in the sole framework of what is advantageous and self-preserving for the individual organism; and that "the good is subjective and ... the right is conventional."[3]

In contrast, Butler argued that self-love, properly understood, is certainly natural, but that *when* properly understood, it is in turn the springboard of a benevolence toward others which both expresses and potentiates fellow-feeling, which is also natural. Further, the mechanism of the expression of both of these natural faculties is the relation between them, and between them and other specific passions and affections, a relation which he terms "reflection or conscience," the superior faculty noted above, which, as a principle of action,

> ...plainly bears upon it marks of authority over all the rest [of the principles of action], and claims the absolute direction of them all, to allow or forbid their gratification...[4]

Butler's method is deliberately naturalistic; he chooses to examine the subject of morals by inquiring into "matters of fact," that is, into the natural constitution of human nature, its "economy," rather than into the "abstract relations of things,"[5] since that (the former) is how Hobbes claims to have arrived at his understanding of human nature. Butler's conception of self-love or "cool" self-love, intended to counteract Hobbes' idea of self-love as merely "the love of power,and delight in the exercise of it,"[6] is not explicitly given; instead, he examines the nature of "nature" and leaves his reader to infer an understanding of self-love as that which motivates actions which most fully express our genuine human nature.

In Sermon II of the *Upon Human Nature* sequence, Butler is concerned to explicate the meaning of "nature" in order to understand what is meant in the epigraph from Romans (II:14):

> For when the Gentiles, which have not the law, do by nature the things contained in the law, these, having not the law, are a law unto themselves.[7]

He argues against the common understanding that, because in humans the principle of reflection or conscience is added to instincts ("that is, appetites and passions"[8]), a person will act

> ...agreeably to his nature ... by following that principle, be it passion or conscience, which for the present happens to be strongest in him.[9]

This common understanding of "nature" can then be misused to excuse wrongdoing:

> ... let not the man of virtue take upon him to blame the ambitious, the covetous, the dissolute, since these equally with him obey and follow their nature. Thus, as in some cases we follow our nature in doing the works contained in the law, so in other cases we follow nature in doing contrary.[10]

It is important, therefore, "in view of all this licentious talk" to really understand what is meant by "nature" and "natural" – to discern

> ...in what sense the word is used, when intended to express and signify that which is the guide of life, that by which men are a law unto themselves.[11]

Butler discerns three senses in which the term "nature" is used. In one sense, a person could equally "follow and contradict his nature," since the term is used only to identify "some principle in man, without regard either to the kind or the degree of it."[12] Since both affection and anger, for example, are such principles, expressing the one necessarily means contradicting the other at that moment. Another sense of "nature" refers us to the strongest and most influential passions in our nature, which, since they are vicious passions, must mean that we are "naturally vicious, or vicious by nature."[13]

These two common understandings of nature are "mentioned only to be excluded"[14] in order to concentrate on the real meaning of "nature" which would explain how it is that men can "do by nature the things contained in the law" and be a "law unto themselves." It is here that Butler introduces the function of the "superior principle of reflection or conscience" which reflects upon actions, passions, and interests and

> pronounces determinately some actions to be in themselves just, right, good; others to be in themselves evil, wrong, unjust, which, without being consulted, without being advised with, magisterially exerts itself, and approves or condemns him the doer of them accordingly; and which, if not forcibly stopped, naturally and always of course goes on to anticipate a higher and more effectual sentence which shall hereafter second and affirm its own.[15]

Conscience or reflection is then the ultimate expression of human nature, the impartial arbiter among passions and interests, which "exerts" itself on its own, "without being consulted," and which nature constitutes as the regulatory agent both of cool self-love and of benevolence. This impartial regulator and arbiter can be seen as the prototype of the external impartial spectator of later writers or, on the other hand, as the

foreshadowing of the internal Kantian lawgiver who "[does] by nature the things contained in the law."[16]

As a "matter of fact," conscience conceives properly functioning self-love impartially as the suitability of actions to the nature of the human person. It is not the action of itself, nor its consequences, that are key, but the relationship of the action to the nature of the agent that is significant. For example, if present gratification of a particular passion or appetite (which Butler is careful to acknowledge are also perfectly "natural," differing from cool self-love not in degree but in kind[17]) is chosen even though it foreseeably leads to personal ruin, this action, although not the passion or the appetite, is conceived as "unnatural" because it is contrary to cool self-love, that is, to actions fostering the development of the person and his character as such. Virtue consists therefore in following, and vice in deviating from, the nature of man, which distinctively includes a capacity for agency, itself uniquely involving acting in conformity with the authoritative principle of reflection.

There is another development, noted in passing in Sermon XII (V in Darwall):

> A man's heart *must be formed* to humanity and benevolence, he must *love mercy,* otherwise he will not act mercifully in any settled course of behavior. ... to get our heart and temper formed to a love and liking of what is good, is absolutely necessary in order to our behaving rightly in the familiar and daily intercourses amongst mankind.[18] [emphasis added]

This passage suggests a problem with the faculty of conscience or reflection, operating as it does through a mechanism of natural sympathy, and expressed in cool self-love (defined as tending to foster the full natural development of the person) and

benevolence. Whether impartiality in any of its forms is a natural faculty or state has obvious implications both for moral education and for the employment of impartiality in moral deliberation, and so it has been important to understand exactly what "natural" refers to. The essential problem can be seen by analogy with other natural faculties of the body, to which Butler makes explicit reference in the beginning of Sermon II:

> ... obligations of virtue shown, and motives to the practice of it enforced, from a review of the nature of man, are to be considered as an appeal to each particular person's heart and natural conscience, *as the external senses are appealed to for the proof of things cognizable by them.* Since then our inward feelings, and the perceptions we receive from our external senses, are equally real; to argue from the former to life and conduct is as little liable to exception as to argue from the latter to absolute speculative truth.[19] [emphasis added]

If impartial conscience or reflection, as the natural regulatory agent of cool self-love, sympathy, and benevolence, is to be regarded analogously to vision, that is, the natural capacity is there but it must be trained, as Butler seems to acknowledge in the "formed to humanity" passage, then the naturalness of conscience or reflection, and its handmaidens sympathy and benevolence, as forms of impartiality, is to be seen only as a potential capacity, for the training required to have these mechanisms *properly* engaged through conscience and reflection is highly dependent upon the traditions and mechanisms in place for training them.

The faculty of vision, for example, is certainly a natural faculty; however, the ability to *see* is evidently not natural, it must be *learned*, and once the time is past for learning it, the skill cannot be gotten. The famous Molyneux problem was played out in real life recently: a man, blind from infancy, received several operations that removed the

obstructions from his eyes; technically, light was afterward entering properly, presumably rods and cones were being stimulated correctly, but in spite of months of intensive coaching, the man was unable to learn how to "see." He could not be made to understand what the various shapes and colors and distances *meant*; for example, he repeatedly failed to recognize his fiancée's face or her bright yellow car, and kept colliding into things, although he was able to navigate perfectly with his eyes closed.[20]

Butler himself indirectly makes the point again when he makes passing mention of superstition as an exception to the natural tendency toward truth and virtue,[21] implying that those trained incorrectly, for example in the ways of superstition, will not behave "naturally." And here lies the rub, for one man's unnatural superstition is another man's natural religion, or perhaps even his science.

The training required to see the world clearly enough to navigate properly and to interpret its colors and shapes correctly is accomplished reliably and quickly enough through early, enforced, and often punishing interaction with a material environment, concerning which there is such a high degree of agreement among persons that one who sees incorrectly is often understood to be either visually or mentally impaired. On the other hand, the training required to "see" *morally*, whether through one's own or another's eyes, is not generally so subtle or reliable, and it is a commonplace that outside one's own social environment, there is a high level of disagreement about what the truths of the moral realm really are. More profoundly, if reflection and conscience are the means to "more or less dispassionate and disinterested"[22] points of view, but, analogously to vision, they must be trained, then this form of impartiality intrinsically depends upon at least the pedagogical adequacy, if not the moral character, of the trainer (who could be a single influential person or a sequence of such persons or an entire society). Butler's

analogy to the workings of the external senses is apt, for it highlights the distinction between what is natural as a capacity and the full functioning of that capacity. This is not an epistemological problem specific to Butler alone, but is a continuing discordant note sounded throughout not only sympathy theories, but many other kinds of moral theories predicated on some kind of naturalness. Impartiality is not natural, in this sense; partiality is. One must be educated into the capability of adopting an impartial perspective. In fact, it is well known that impartial or objective perspectives are often tenuous, easy to fall out of and difficult to return to.

Variations of this problem affect the development of Hume's theory of morals, and are particularly vivid in Smith, as will be seen.[23]

David Hume

The theme of naturalness, with its troublesome implications for sympathetic impartiality, continues in Hume. There is not a fully articulated theory of impartiality in either the *Treatise* or the *Inquiry*, but there is ample opportunity for inferences as to how impartiality could be conceptualized and engaged in moral deliberation, given the premises Hume is working from. Hume's impartial spectator is suggested in some passages from the *Treatise* on the importance of generalization for moral deliberation:

> ...it may ... be objected to the present system, that if virtue and vice be determined by pleasure and pain, these qualities must, in every case, arise from the sensations; and consequently any object, whether animate or inanimate, rational or irrational, might become morally good or evil, provided it can excite a satisfaction or uneasiness. ... it is only when a character is considered *in general*, without reference to our particular interest, that it causes such a feeling or

> sentiment as denominates it morally good or evil. It is true, those sentiments from interest and morals are apt to be confounded, and naturally run into one another. ... But ... *a man of temper and judgment* may preserve himself from these illusions. ... a person ... *who has command of himself*, can separate those feelings and give praise to what deserves it. [emphasis added]

This man must further

> depart from his private and particular situation and must choose a point of view common to him with others: he must move some universal principle of the human frame and touch a string to which all mankind have an accord and symphony.[24]

The "man of temper and judgment," in Hume's theory, who considers character in general terms separated from his particular interest, becomes that way due to his grounding in a natural benevolence and humanity operating through the mechanism of sympathy. Sympathy, according to Hume, is the idea of a passion produced by observation of its effects which is "presently converted into an impression and acquires such a degree of force and vivacity as to become the very passion itself."[25] The capacity for sympathy is innate in all humans and attests to the essential resemblance and similarity of *all* humans to one another, not merely cultural neighbors, as well as to the benevolence and humanity implanted in us all at birth:

> ...there is some benevolence, however small, infused into our bosom; some spark of friendship for humankind; some particle of the dove kneaded into our frame, along with the elements of the wolf and serpent. [26]

> This benevolence and humanity turn our moral distinctions of approval and

disapproval toward the things which are or are not useful to society; whether the utility of an action or a person affects us personally or not, our innate benevolence gives us a cool preference for the useful over the non-useful or pernicious to society:

> Let these generous sentiments be supposed ever so weak, let them be insufficient to move even a hand or finger of our body, they must still direct the determinations of our mind, and, where everything else is equal, produce a cool preference of what is useful and serviceable to mankind above what is pernicious and dangerous. *A moral distinction, therefore, immediately arises;* a general sentiment of blame and approbation; a tendency, however faint, to the objects of the one, and a proportionable aversion to those of the other. [emphasis added][27]

Significantly, here moral distinctions are not *made* in a conscious and intentional, rational, manner; they *arise* of their own accord on account of a perhaps weak benevolence, in response to the perceived utility or non-utility of things. Reason has no place in the final decisions of moral deliberation, which Hume conceives as instinctual responses of a certain kind of approbation or disapprobation. Reason does have its place:

> Reason is the discovery of truth or falsehood ... [which] consists in an agreement or disagreement either to the *real* relations of ideas, or to *real* existence and matter of fact. ... Reason ... can have an influence on our conduct only after two ways: either when it excites a passion, by informing us of the existence of something which is a proper object of it; or when it discovers the connection of causes and effects, so as to afford us means of exerting any passion.[28]

In the *Treatise*, reason is separate from will and functions purely as a sophisticated means-end analyst in moral deliberation, while it is passion that directs, and ought to direct, the choice of ends:

> ...[reason's] proper province is the world of ideas, and as the will always places us in that of realities, demonstration and volition seem upon that account to be totally removed from each other. ... Abstract or demonstrative reasoning, therefore, never influences any of our actions, but only as it directs our judgment concerning causes and effects ...Reason is, and ought only to be, the slave of the passions, and can never pretend to any other office than to serve and obey them.[29]

In short, the impartial, "judicious" spectator,[30] who probably would be modeled upon the "man of temper and judgment," can only be a fully actualized being, fully developed in all natural capacities, with all of these not only intact but developed to their maximum potential, whose moral judgements are, in Hume's words, "infallible" because they are grounded in the original "humanity" and benevolent constitution of human nature:

> The general opinion of mankind has some authority in all cases; but in this of morals it is perfectly infallible. ... The moral obligation is founded on the natural ...[31]

The infallible moral deliberator uses his sentiments of approval or disapproval in the face of certain actions or persons as evidence of the morality or immorality of these actions or persons. But how are these sentiments themselves to be understood? According to Hume, the humanity and benevolence they are grounded in are "instincts":

> The social virtues of humanity and benevolence exert their influence immediately by a *direct tendency or instinct*, which chiefly keeps in view the simple object, moving the affections, and comprehends not any scheme or system, not the

> consequences resulting from the concurrence, imitation, or example of others.[32] [emphasis added]

Alternatively, moral distinctions are also "evidently perceptions":

> Our decisions concerning moral rectitude and depravity are evidently perceptions. ... Morality, therefore, is more properly felt than judged of; though this feeling or sentiment is commonly so soft and gentle that we are apt to confound it with an idea.[33]

Both of these conceptualizations, which are not mutually exclusive but are distinct, illustrate the same problem, but it is a problem that does not arise with them, it is merely highlighted by them. The problem is twofold: on the one hand, an individual wrongdoer could certainly be seen as immoral; but this would merely be the expression given to the sentiment of disapproval that would arise spontaneously in a spectator upon being confronted with the wrongdoer's actions. But he could not be *blamed* for his wrongdoing. If moral distinctions are based solely in instinctual natural capacities, the worst response that could be directed toward the wrongdoer is not blame, but pity. The person is not evil, but *impaired* in his natural instincts of humanity and benevolence, as Hume himself acknowledges:

> ...the notion of injury or injustice implies an immorality or vice committed against some other person: And as *every immorality is derived from some defect or unsoundness of the passions,* and as this defect must be judged of, in a great measure, from the ordinary course of nature in the constititution of the mind, it will be easy to know whether we be guilty of any immorality with regard to others, by considering the natural and usual force of those several affections which

are directed towards them.[34] [emphasis added]

It is not clear how to understand the notion of being "guilty of immorality" in the context of impairment. This is particularly emphasized by Hume's distaste for the idea of freedom of the will; we are all natural creatures who operate predictably and systematically according to laws of human psychology, no less so than stones which "operate" according to laws of gravity when they are dropped from the roof of a building. Morality and immorality then describe the *condition* of a person, in almost a clinical sense, that is, whether the person is impaired or fully functional. A moral person is fully functional, and the fullest and best functioning of something is what makes it most useful, and therefore, in Hume's analogy of moral sensibility with aesthetic sensibility, beautiful:

> Most of the works of art are esteemed beautiful, in proportion to their fitness for the use of man; and even many of the productions of nature derive their beauty from that source. Handsome and beautiful, on most occasions, is not an absolute, but a relative quality, and pleases us by nothing but its tendency to produce an end that is agreeable. The same principle produces, in many instances, our sentiments of morals, as well as those of beauty.[35]

On the other hand, if moral distinctions are perceptions, the same problem of reliably training the moral-perceptual skill arises that was present in Butler. In Hume's theory, the trainer would be experience, based in convention, and even though he grounds convention, for example, justice, in natural human laws and necessities which are the same for everyone (since everyone operates with propensity toward pleasure and aversion toward pain), it is somewhat too optimistic to suggest that because of this, convention is always a reliable and appropriate trainer for moral perception. Here is the shadow of

relativism that first raised its head in Butler and becomes larger in Hume.

In the face of the pervasive, almost physiologically-based naturalism in Hume's theory, with no room for reason except to judge appropriately of the existence or nonexistence of the elements and relations of a moral situation, it is hard to see a function for impartiality, even for sympathetic impartiality. Hume is evidently uneasy about this conception of nature, because certain natural human institutions, such as justice and property, seem to require reason.[36] But if morality is founded in reason and understanding, it would "[suppose] a real right and wrong, that is, a real distinction in morals, independent of these judgments"[37] and Hume argues at some length against this idea.[38] Sympathy is then the mere psychological mechanism that operates to express a common humanity and benevolence, but even when it is consciously adduced, for example in decisions of justice, sympathy is attenuated, abstract, and generalized to the whole of the social environment, rather than operating powerfully (because instinctually) in relation to another person's experiences. In a certain sense, physiologically-based moral naturalism takes impartiality out of the realm of moral deliberation altogether, leaving it only to mediate between means and ends and to assess general utility. A more vivid sympathy may in fact lead to "partiality" which then needs to be "corrected," as perceptions are routinely corrected. In Hume's discussion of necessity in human affairs as contrasted with necessity in natural science, he tells us we must "judge of the actions of men ... upon the same maxims as when we reason concerning external objects" and advises us that experience will correct our perceptions in either case:

> When any phenomena are constantly and invariably conjoined together, they acquire such a connection in the imagination, that it passes from one to the other

> without any doubt or hesitation. But below this there are many inferior degrees of evidence and probability, nor does one single contrariety of experiment entirely destroy all our reasoning. *The mind balances the contrary experiments*, and, deducting the inferior from the superior, proceeds with that degree of assurance or evidence, which remains.[39]
>
> In general, all sentiments of blame or praise are variable, according to our situation of nearness or remoteness with regard to the person blamed or praised, and according to the present disposition of our mind. But these variations we regard not in our general decisions, but still apply the terms expressive of our liking or dislike, in the same manner as if we remained in one point of view. Experience soon teaches us this method of *correcting our sentiments*, or at least of correcting our language, where the sentiments are more stubborn and unalterable.[40] [emphasis added in both quotations]

The command "Thou *shalt* love thy neighbor as thyself," the scriptural injunction to impartial benevolence, loses its grammatical mood in Hume's aggressive naturalism; it changes from an imperative (an "ought"), and must be understood in the indicative mood as "Thou *dost* love thy neighbor as thyself" (an "is" -- but possibly only weakly and without motivating force, and at that only if all is well and the organism's instincts are fully functional and not impaired). "Love thy neighbor" as a command requires agency; but Hume would have us believe that we do not need such a command, nor would it be pertinent to give one, since on the one hand this benevolence is already programmed in "however weakly" and on the other hand, the passions cannot be commanded in any case.

So far, discussion of sympathetic impartiality has disclosed some elements of its anatomy which makes it a far more complex psychological process than merely a generalized fellow-feeling that can be taken for granted. One must take into account the

degree of similarity to oneself in the sympathized-with other, or judgements of immorality may be made having their foundation merely in unfamiliarity. One must also take into account the identities of both sympathizer and sympathized-with, lest moral decisions made by one "person" be different from those made by the other (although, given the degree of familiarity required to get sympathy moving in the first place, this may end up as only a minor difference in outcome). Further, the understanding of morality, sympathy, humanity, and associated concepts as "natural" is also complex, tending not only to enclose and constrict the relevant moral community, but also to mask to what degree "natural" means also "formed" or "educated," not to mention how this training is accomplished and by whom. In Adam Smith, who distills the impartial spectator out of these and other elements, the multi-faceted complexities of sympathetic engagement become clearly visible.

Adam Smith

In Smith, the analysis of sympathy, its mechanisms, and its place in social intercourse, is more detailed and carefully developed than in Hume or Butler, with the impartial or indifferent spectator making his explicit appearance. The ongoing problems with sympathy are highlighted in the careful detail of his analysis: the etiology of sympathy is again assumed to be "nature," with persuasive examples drawn from common experience, but in Smith's "nature" are conflated the *mechanism* (faculty, capacity) of sympathy with the justness or fittingness of the *outcomes* of sympathetic reflection and feeling. Further, various other dimensions of sympathy are conflated into nature here as well, as they have been all along, but because of the detail in Smith, this problem now becomes more obvious. For example, a physiological or kinesthetic

sympathy may indeed be natural, as Smith's examples and ordinary experience make clear, but a different, perhaps more rarefied dimension of sympathy is emotional sympathy, and an even more refined dimension may even be intellectual in character. Since the natural mechanism of sympathy is said to generate the impartial spectator, or conscience, which is the arbiter of the right thing to do,[41] these conflations have implications for the disinterested judgements, those that are "considered without any particular relation either to ourselves, or to the person whose sentiments we judge of," that Smith says are based not on utility but on truth and on what is right.[42] Smith echoes Butler when he says the impartial spectator is

> ... reason, principle, conscience, the inhabitant of the breast, the man within, the great judge and arbiter of our conduct [who is capable of counteracting the strongest impulses of self-love]. It is he who ... calls to us ... that we are but one of the multitude, in no respect better than any other in it; and that when we prefer ourselves ... to others, we become the proper objects of resentment, abhorrence, and execration. It is from him only that we learn the real littleness of ourselves, and of whatever relates to ourselves, and the natural misrepresentations of self-love can be corrected only by the eye of this impartial spectator.[43]

But although reason, rightness, justice, and fittingness play a vital role in the formulation of Smith's impartial spectator, who is our conscience, nevertheless he cannot come to be unless he is capable of sympathizing correctly with the proper objects, thus revealing yet another dimension of sympathetic engagement, one that depends upon degree of emotion:

> There are some situations that bear so hard upon human nature, that the greatest degree of self-government ... is not able to ... reduce the violence of the passions to

> *that pitch of moderation, in which the impartial spectator can entirely enter into them.*[44] [emphasis added]

Correct sympathy relies essentially on a "pitch of moderation," a correct mutual modulation of passion and sympathy between the experiencer of the passion and the spectator. In this passage, which is worth quoting at some length, Smith gives a detailed psychological analysis of the mechanism of sympathy:

> ... [T]hat there may be some correspondence of sentiments between the spectator and the person principally concerned, the spectator must, first of all, endeavour as much as he can to put himself in the situation of the other, and to bring home to himself every little circumstance of distress which can possibly occur to the sufferer. ... After all this, however, the emotions of the spectator will still be very apt to fall short of the violence of what is felt by the sufferer. Mankind, though naturally sympathetic, never conceive, for what has befallen another, that degree of passion which naturally animates the person principally concerned. That imaginary change of situation, upon which their sympathy is founded, is but temporary. ... The person principally concerned is sensible of this, and at the same time passionately desires a more complete sympathy. ... But he can only hope to obtain this by lowering his passion to that pitch, in which the spectators are capable of going along with him. He must flatten ... the sharpness of its natural tone, in order to reduce it to harmony and concord with the emotions of those who are about him. ... In order to produce this concord, as nature teaches the spectators to assume the circumstances of the person principally concerned, so she teaches this last in some measure to assume those of the spectators.[45]

For Smith, sympathy itself is more properly seen as based on reason and judgement than on pure feeling. Sympathy arises from the imagination, not from sense,

and this mechanism works only for certain situations, not all of them; for example, one's being in love with a particular person arouses no sympathy in the bystander in the sense of the bystander's experiencing love as well. It is the situation, not the passion, that engages the bystander. Smith makes it clear that we must consider imagination, not feeling, to be the key element in sympathy. If it were only feeling, then we would merely "mirror" to a weaker degree the feeling that we observe in others, as Hume in fact suggests.[46] Not only does this not always happen, but on occasion the bystander will, upon assessing the situation, actually experience "correct" feelings that are in fact *not* experienced by the person under observation. As an example, Smith offers the madman, who may actually be cheerful, laughing and singing, but

> ...the compassion of the spectator must arise altogether from the consideration of what he himself would feel if he was reduced to the same unhappy situation, and, what perhaps is impossible, was at the same time able to regard it with his present reason and judgement.[47]

To a great extent, our sympathy is engaged when the passion we observe is suited to its object; "propriety" is a key element of Smith's analysis of moral discernment based on sympathy, as is the concept of the object of the passion. Experiencing sympathy means that the spectator approves of the affection or passion of the person under observation as just and proper, as suitable to its object. That suitableness can be judged only by the coincidence of the affection felt by the spectator with the affection experienced by the other; a perception of this coincidence leads to moral approbation:

> When the original passions of the person principally concerned are in perfect

> concord with the sympathetic emotions of the spectator, they necessarily appear to this last just and proper, and suitable to their objects; and, on the contrary, when, upon bringing the case home to himself, he finds that they do not coincide with what he feels, they necessarily appear to him unjust and improper, and unsuitable to the causes which excite them. ... To approve of the passions of another, therefore, as suitable to their objects, is the same thing as to observe that we entirely sympathize with them; and not to approve of them as such, is the same thing as to observe that we do not entirely sympathize with them.[48]

The objects of moral judgements are involved in two distinct perceptions: the rightness or wrongness of conduct, and the merit or demerit of the agent; both depend upon propriety, upon the considered suitableness of the passions to their cause, or of the merit or demerit of the agent based on consideration of their effect. Sympathy, in short, can be not only with affections or passions, but with motives as well;[49] but in the approbation of propriety of a person's sentiments, I must first perceive the correspondence of sentiments between him and myself. On the other hand, when assessing merit, I need not experience either the resentment or the gratitude of the recipient or patient of an action; I necessarily approve or disapprove of the agent when I bring his case home to myself regardless of the recipient's actual emotional responses.[50]

And finally, "this sentiment," sympathy, is one of the "original passions of human nature";[51] "nature ... has stamped [it] upon the human heart."[52] Smith, like Butler, explicitly bases his theory of sympathy on a nature that is "wisely ordered," on matters of fact:

> Let it be considered too, that the present inquiry is not concerning a matter of right, if I may say so, but concerning a matter of fact. We are not at present

> examining upon what principles a perfect being would approve of the punishment of bad actions; but upon what principles so weak and imperfect a creature as man actually and in fact approves of it.[53]

Smith suggests, however, that these "facts" are preordained by the "Author of Nature" who has not "entrusted it to [man's] reason" to find out for himself what is right or wrong, but has programmed in a mechanism of sympathetic reflection

> [w]ith regard to to all those ends which may be regarded as the favourite ends of Nature [operating by means of an endowment of an] appetite for the end which she proposes, [and] likewise with an appetite for the means by which this end can be brought about.... Nature has directed us to the greater part of these by original and immediate instincts."[54]

While Smith is here explicitly talking about propagation and self-preservation, with its mechanisms of sexual appetite, hunger, thirst, and so on, the elision between these natural instincts, which he calls the "mechanism of sympathetic reflection" and his previous discussion of propriety, which is sympathetically based, suggests that our sense of propriety is as naturally based as our biological instincts, a suggestion reinforced by his explicit statement that sympathy is an original passion of human nature and that nature has stamped it upon the human heart. This is the conflation of dimensions of sympathy alluded to earlier: an instinctive, clearly natural, sympathetic wince upon observing someone hit a thumb with a hammer is a far cry from the "instinctive" sympathetic wince upon hearing a friend's complaint that the wrong class of people is moving into the neighborhood.

Taking all the elements of correct sympathy into account, therefore, correct

sympathy relies upon the observer's assessment of the propriety of the passions or motivations observed, which depends upon the congruence of those passions or motivations with the observer's own, were he to place himself into the same situation he observes. If congruent, these imagined passions or motives of the observer are more or less assured by nature to be appropriate to the circumstances observed, or at least the observer believes this to be so. Under these circumstances, the incapacity of the observer to feel sympathy means that he judges that the experiencer's passions or the agent's motives are "not suited," which in its turn means the the observer may legitimately disapprove of them.

It is not a far leap to see how such a conceptualization of sympathy relies entirely upon, on the one hand, the observer's assurance that his sympathies are naturally based – that is, designed by the Author of Nature – and on the other hand, depends upon a closed cultural system to hone these instincts "correctly" and to their optimal pitch; so that the observer is, in the truest sense, "well bred." Ideas, practices, beliefs, and so on that fail to elicit from the observer the correct sympathy can therefore be ostracized, excluded, or punished on the grounds that they are not natural. One might not learn how to engage one's natural sympathies on behalf of someone foreign or alien to one's established way of thinking, and loving one's neighbor as oneself becomes an extremely local matter.

Smith's analysis of the origin of the general rules of morality reinforces this inference by explicitly making the connection between them and divine law in a chapter entitled, "Of the Influence and Authority of the General Rules of Morality, and that they are Justly Regarded as the Laws of the Deity":

> The man who was injured ... could not doubt but that divine being would behold it

> with the same indignation which would animate the meanest of mankind ... [t]he man who did the injury felt himself to be the proper object of the detestation and resentment of mankind; and his natural fears led him to impute the same sentiments to those awful beings, whose presence he could not avoid, and whose power he could not resist. These natural hopes, fears, and suspicions were *propagated by sympathy, and confirmed by education*; and the gods were ... believed to be the rewarders of humanity and mercy, and the avengers of perfidy and injustice. ... [when] philosophical researches came to take place, [they] confirmed those original anticipations of nature. ... It cannot be doubted that [moral faculties, as] a principle of our nature ... were *given* us for the direction of our conduct.[55] [emphasis added]

Furthermore, these divinely given general rules supposedly act as "a corrective to self-love and self-deception"[56] and are to be inferred from observation of the conduct of others.[57]

The foregoing analyses of individual structures of sympathetic identification in Butler, Hume, and Smith have disclosed an intricate web of interdependent elements and dimensions, each of which may influence outcomes of moral deliberation. Smith's analysis is most exhaustive, but reveals in the end that this modality of sympathy, in order to operate properly, requires an almost hermetically sealed, small moral universe, whose members may reasonably be counted upon to be homogeneous enough for sympathy to generate proper and appropriate moral rules and principles.

Issues of restrictiveness or "closedness" that can interfere with proper sympathetic identification with persons outside the system might be addressed with a more open conception of sympathy and a more generalized view of what an ideal sympathetic observer might be like. Sympathy *is* truly experienced as a natural phenomenon, or so it seems to the experiencer, so one wishes not to throw out the baby

with the bathwater when noting the serious difficulties with an individualistic sympathy too narrowly conceived and too readily affirmed by reference to "natural" capacities or divine approval. In a more contemporary refinement, an analysis of the necessary characteristics of such an ideal observer, and particularly of what is meant by attributions of impartiality, disinterestedness, and dispassionateness, are offered by Roderick Firth in his "Ethical Absolutism and the Ideal Observer."[58]

Roderick Firth

Firth is concerned to defend a certain kind of analysis of ethical statements; this analysis is both absolutist and dispositional. By "absolutist" he means that judgements about ethical statements are non-relative; that is, they contain no "egocentric expressions" pertaining to particulars of person, time, place, or tense – in short, anything that can be indicated by a demonstrative pronoun such as "this" or "that."[59] Ethical statements are judged true or false, without reference to these particulars.

By "dispositional" he intends to signify the non-relative ethical reactions to these statements of an ideal observer, construed not as an actual being, who would then have to be God Himself, but as a possible being. If the ideal observer is construed as God, then the truth or falsity of an ethical statement stands or falls with the determination of the existence or nonexistence of God. As a possible being, however, the ideal observer would have the dispositions of all possible beings of a certain kind, whether these beings actually exist or not, returning truth or falsity to ethical statements and rendering truth-judgements objective and independent of the existence of an experiencing subject. All that is needed is the "conceivability" of an ideal observer; ethical judgements are then rendered in the subjunctive conditional mood, both absolutely and empirically, if "empirical" is defined to

include the dispositional concepts of the natural sciences.

Furthermore, such an analysis is relational as well, in that a lawful relation is presumed to exist between "certain reactions of an ideal observer and the acts or other things to which an ethical term may correctly be applied."[60] This aspect of the analysis relies upon the form of ethical statements being the same as that of statements about secondary qualities:

> Not only phenomenalists and subjectivists, but many epistemological dualists, would agree that to say that a daffodil is yellow is to say something about the way the daffodil would appear to a certain kind of observer under certain conditions; and the analysis of ethical statements which we are considering is exactly analogous to this. Thus the sense in which an absolutist dispositional analysis is relational, is the very sense in which a great many philosophers believe that yellow is a relational property of physical objects; and to say that a statement of the form "x is right" is relational, therefore, is not necessarily to deny that the terms "right" and "yellow" designate equally simple properties.[61]

Then the characteristics of the ideal observer are enumerated and analyzed. Here we also see how varying concepts of impartiality are collected together to constitute an ideal observer; it is significant that clear distinctions among the concepts are made here.

To begin with, an ideal observer is *omniscient* about all non-ethical facts, and not merely about the "relevant" non-ethical facts (since judgements concerning relevance and non-relevance in moral matters are necessarily normative, such judgements of relevance by an ideal observer are necessarily circular). An ideal observer is also *omnipercipient*; it is not just the facts that he has knowledge of, but he possesses as well a universal imagination that allows him vividly to imagine ("visualize") all "actual facts, and the

consequences of all possible acts in any given situation, just as vividly as he would if he were actually perceiving them all -- universal perception." This capacity guarantees that "his ethically-significant reactions are forcefully and equitably stimulated."[62]

Furthermore, the ideal observer is *disinterested*, or "completely impartial." What being disinterested actually means, how it is defined, is a matter of some concern as once again, as with morally relevant non-ethical facts, there is a risk of circularity. A too-broad definition is one that might naturally come to mind in evaluating the impartiality of a moral judge by pointing to factors that have perverted or enhanced his or her decisions. But marking these factors presupposes a definition of impartiality that already includes them, thereby rendering the definition circular. On the other hand, the definition can be too narrow. Firth cites "Bentham's maxim" to represent the utilitarian idea of impartiality, that "every man should count for one and none for more than one,"[63] to argue that this evidently excludes duties deriving from special relationships. Disinterestedness then must be carefully defined as "not influenced by particular interests," where "particular" cannot be defined without use of proper names.[64] This definition of disinterestedness allows us to define even a judge who does have particular interests as "disinterested," provided he or she suppresses them during judicial decision-making.[65] One who is disinterested, in short, is one who is entirely lacking in particular interests, if only for the moment.

An ideal observer is also *dispassionate*, which may be defined analogously to being disinterested: incapable of experiencing emotion. Impartiality, says Firth, "cannot be exhaustively analyzed in terms of interest, for an impartial judge, as ordinarily conceived, is a judge whose decisions are unaffected not only by his interests, but also by his emotions."[66] Although Firth allows that "dispassionate" may be defined point by

point analogously to "disinterested" insofar as particular emotions are analogous to particular interests, he says we can go further still and disallow the ideal observer any general emotions as well, whereas we cannot disallow general interests in the same way. This brings the "conception of an ideal observer closer to Kant's conception of a 'purely rational being.'"[67] Curiously, love and compassion are not here characterized as emotions but as virtues, and a window is left open for them to come into the definition of an ideal observer, but only under certain extremely restricted circumstances, namely if (assuming they are really virtues and not emotions) they enter into the factors which influence one or another conception of an ideal observer. This leaves open the possibility of differing culturally- or religiously-specified conceptions of an ideal observer. This point will be discussed in more detail later; for the current conception under consideration, Firth tells us that while attribution of love and compassion as characteristics of an ideal observer would not be a logical mistake, it is nevertheless unnecessary:

> The value of love and compassion to a judge, considered solely as a judge, seems to lie in the qualities of knowledge and disinterestedness which are so closely related to them; and these two qualities... can be independently attributed to an ideal observer.[68]

The last quality of an ideal observer is *consistency*. Consistency is not to be construed merely in the logical sense, however, for in inspecting, say, Singer's generalization principle (what is right [or wrong] for one person must be right [or wrong] for any similar person in similar circumstances),[69] it is immediately obvious that regardless how similar the persons or circumstances may be, the anticipated judgement would

> ...express ethical decisions about two different cases, [therefore] they necessarily refer to different acts or events, and of course *any* two self-consistent statements are logically consistent with one another if they refer to different acts or events. Thus the kind of consistency we have in mind must be 'stronger' than logical consistency."[70]

It is possible to say that a moral judge's decisions are consistent even when, in two separate cases, all things being equal, he or she upholds in the first, then denies in the second, the same principle P. Closer inspection of such a case would probably reveal that there is another principle Q in play which, together with principle P, renders consistent both decisions. Firth concludes then that a decision of consistency in two different cases must necessarily presuppose "a certain amount of ethical knowledge, [thus implying] that our analysis would be circular if we made consistency of this kind one of the defining characteristics of the ideal observer."[71] Therefore, consistency must be construed differently from the other characteristics of the ideal observer -- omniscience, omnipercipience, dispassionateness, and disinterestedness. Consistency does indeed differ from the other characteristics: whereas they are in place specifically to "eliminate some particular source of disagreement in ethical reactions, [consistency] ... is, on the contrary, a *consequence* of eliminating such disagreement."[72]

In other respects, says Firth, the ideal observer is "normal," meaning that he is a *person*. Therefore,

> no analysis in terms solely of such general, and highly ideal, characteristics, could be fully adequate to the meaning of ethical statements [because] our conception of the personality of an ideal observer has not yet undergone the refining processes which have enabled theologians, apparently with clear conscience, to employ the

> term 'person' in exceedingly abstract ways. Most of us, indeed, can be said to have a conception of an ideal observer only in the sense that the characteristics of such a person are implicit in the procedures by which we compare and evaluate moral judges, and it seems doubtful, therefore, that an ideal observer can be said to lack any of the determinable properties of human beings.[73]

However, no matter the difficulties of specifying the determinate characteristics of an ideal observer and mapping them somehow onto the limits of (unspecifiable) human "normality," the thesis stands, for Firth, that "ethical statements are statements about an ideal observer and his ethically-significant reactions."[74]

The concept of the ideal observer or spectator, then, could be a way out of the constraints of the familiar that so restrict individual sympathy. By sympathetic identification, one identifies oneself with this ideally sympathetic person, a person who is in turn able sympathetically to take in the perspectives of all persons concerned, but not be overwhelmed either with their interests, since the ideal observer is disinterested, or with their passions, since the ideal observer is also dispassionate. Moreover, he also knows what those interests and passions are, because he is omniscient, and he has them present vividly to his view, since he is universally imaginative or omnipercipient as well. Finally, he is reliable, since his moral reactions are consistent. At last, one might think, there is a way to understand what loving one's neighbor as oneself, regardless of who the neighbor (or oneself) is, actually amounts to from a moral point of view.

But the problems at the core of the ideal observer concept will not yet permit such understanding. Gilbert Harman touches on the heart of this set of problems when he says that emotivism, of which ideal observer theory is a logical outgrowth, and ideal observer theory itself, are "not specifically moral theories."[75] His particular arguments

focus on both the circularity and the triviality of the theories, and the essential conceptual problem that the ideal observer does not (cannot) appeal to moral principles to resolve difficult problems, like why it is morally permissible to kill fetuses but not infants; he merely *reacts*, dispositionally, in presumed contrast with most ordinary moral decision-makers. In addition, in his analysis the ideal observer as a construct relating to possible, not divine, beings, appears also to lend itself to relativism, since, significantly, a Christian ideal observer may differ radically from a Muslim ideal observer in his ethical reactions, for example in differing conceptions of familial hierarchies and powers.

Another kind of reason for disclaiming the ideal observer as an arbiter of moral rightness or wrongness lies in the very mechanics of the sympathetic identification that gets the whole thing going. The ideal observer must enter sympathetically into the perspectives of everyone, in order to avoid partiality; but this does tend to suggest a God's-eye view. Such a view seems to require an anchoring, screening principle in order to function as a specifically moral theory, for example the principle of utility or some other principle, for otherwise a free-floating God's-eye view places the ideal observer beyond the capacity of making moral distinctions or having moral experiences or reactions at all – effectively verging, in fact, on the depersonalization problem characteristic of a different style of impartiality. This turns us directly back to the "nature" problem of the earlier analysis: true nature must, so to speak, take on such a viewpoint and therefore does not (cannot) discriminate morally; when it rains, it rains on everyone. Laws that operate in nature do so impersonally. In Nagel's terms, this is the "view from nowhere" – it is impossible, in those terms, not to see the persons of the world as caught in a causal nexus from which there cannot be any escape, and therefore no freedom, and therefore no moral responsibility. To know all might very well be to understand and thereby forgive

all, although the epistemological concept of understanding and the moral concept of forgiveness would have to be mapped onto what is actually a purely impersonal mechanism.

Kant seems to add such an anchoring screen to this point of view – the moral quality of goodness – but this serves instead to make a similar point in Kant's own idiom: God (the omniscient-omnipercipient) does not need morality, since he knows all and his will is perfect and therefore already perfectly good:

> A perfectly good will, therefore, would be equally subject to objective laws (of the good), but it could not be conceived as constrained by them to act in accord with them, because, according to its own subjective constitution, it can be determined to act only through the conception of the good. Thus no imperatives hold for the divine will or, more generally, for a holy will. The "ought" is here out of place, for the volition of itself is necessarily in unison with the law. Therefore imperatives are only formulas expressing the relation of objective laws of volition in general to the subjective imperfection of the will of this or that rational being, e.g., the human will.[76]

In Firth's terms, a perfectly good will is perfectly and infallibly disposed to identify correct ethical relations, but then it is not clear what function the term "ethical" serves. And Sidgwick tells us:

> ... 'dictate' or 'imperative' ... describes the relation of Reason to mere inclinations or non-rational impulses by comparing it to the relation between the will of a superior and the wills of his subordinates. This conflict seems also to be implied in the terms 'ought,' 'duty,' 'moral obligation' as used in ordinary moral discourse: and hence these terms cannot be applied to the actions of rational

beings to whom we cannot attribute impulses conflicting with reason.[77]

Morality, the "ought," is not for the un-anchored, omniscient, ideal observer – it is for those with imperfect will and particularly with imperfect knowledge. It is arguable that a holy will could even recognize an "ought" if it saw it, since a holy will is identical with a "would." Omniscience and omnipercipience vitiate morality since it must then be the balance of things in general, and not the morality of things in particular, that is the concern of the ideal observer -- their "justice" in Anaximander's ancient sense and not in the local, human sense.[78]

In the end, this attempt to break out of the closed circle of sympathy leaves us either with a genuinely ideal, perfect, possible observer, who (even if such a being were humanly possible) cannot judge morally because of his own perfection, but who can, indeed must, love his neighbor (if love is permitted into this conceptual scheme), or with a normal human being whose moral processes may incorporate in a natural way an inchoate idea of this being and sympathetically identify with it, but with the knowledge that this idea is vitally dependent upon the cognitive and cultural presuppositions locally in place, returning us in spite of ourselves to a closed and dangerously self-deceptive system, with its suggestion that all of our sympathetic responses are "natural."

Conclusion

Modalities of sympathy offer themselves intuitively for use in moral deliberation in what appears at first glance to be a perfectly natural manner. Based upon the common human experience of sympathizing or empathizing with another, sympathy, with its concomitant benevolence,[79] seems initially to be intuitively unproblematic and even rather

attractive as a means of solving the problem of being motivated to consider the interests of other persons than ourselves -- of loving our neighbor as ourselves. Upon closer examination, however, sympathetic identification takes on a slightly more sinister air as efforts to explicate "sympathy," "identification," and "nature" seem to lead inexorably either to an image of a moral community so narrow and homogeneous that morality is almost superfluous (since it is easy to be impartial among similars), or to an image of an all-knowing dispositionally-correct entity who is so impersonally "natural" that morality once again seems to have little or no function (since nature's laws are as impartial as can be).

We have seen that interpreting the command to love one's neighbor through the modality of sympathy, while intuitively the fitting thing to do, nonetheless can end in poor results for conceptions of who one's "neighbor" necessarily must be in order for one to "love" him – if one "loves" only through sympathy, one's neighborhood can then become somewhat localized and parochial.

Perhaps turning attention away from the loving and toward the nature of the neighbor, and on what it can mean to attend to a neighbor "as oneself," will yield more satisfactory results for inclusion into a moral community. By this time, an idea of what would constitute a satisfactory result begins to take on a rough outline, if only in a negative sense, by eliminating elements that are not wanted: a too-narrow parochialism or a too-wide impersonality. One's neighbor, by tradition, is conceived as "anyone who enters my horizon," either personally or by representation,[40] but we have no way as yet to understand how to be motivated to incorporate our neighbors, in this sense, into a satisfactory moral community. Fortunately, the clause "as oneself" provides a direction into the next modality of impartiality, the modality of detachment, which may resolve

some of the problems that sympathy has led to.

NOTES

Chapter Three

[1] Butler, *Five Sermons*, Sermon 1.

[2] Ibid., 26, n.4.

[3] Ibid., Introduction, 2.

[4] Ibid., Preface, 15.

[5] Ibid.

[6] Ibid., Sermon 1, 27, n. 4.

[7] Ibid., Sermon 2, 34.

[8] Ibid., 35.

[9] Ibid.

[10] Ibid., 36.

[11] Ibid.

[12] Ibid.

[13] Ibid.

[14] Ibid., 37

[15] Ibid.

[16] Darwall makes a different distinction between the internal perspective and the external impartial spectator. In a review article on recent studies of Adam Smith, he distinguishes between the external impartial spectator who observes impersonally, from a perspective associated with aesthetic distance, and the internal perspective which "is not strictly a spectator's standpoint at all." The internal perspective is that of the agent, when judging of the agent's motive, or the patient, when judging of the patient's feeling. Stephen Darwall, "Sympathetic Liberalism: Recent Work on Adam Smith," *Philosophy and Public Affairs* 28, no. 2 (1999): 141-142.

[17] Butler, *Five Sermons*, Sermon 2, 38.

[18] Ibid., Sermon 5 (XII), 59-60.

[19] Ibid., Sermon 2, 34.

[20] Oliver W. Sacks, "To See and Not to See." *The New Yorker*, 10 May 1993, 59-66.

[21] Butler, *Five Sermons*, Sermon 3, 42-43.

[22] Ibid., Introduction, 4.

[23] Mill's take on what is to be considered natural is slightly different for it assimilates what is acquired to what is natural: "... if, as is my own belief, the moral feelings are not innate, but acquired, they are not for that reason the less natural. It is natural to man to speak, to reason, to build cities, to cultivate the ground, though these are acquired faculties." (J. S. Mill, *Utilitarianism, Liberty, and Representative Government* [London: J.M. Dent & Sons, Ltd., 1947 (1863)], 28). In other words, it is natural to acquire such faculties, but this still leaves the problem of determining which of the acquired *moral* faculties are "correct" and which aren't. After all, one may certainly build cities poorly, cultivate the ground incorrectly, and reason in fallacies, none of which vitiates the fact that these acquired faculties are natural.

[24] David Hume, *An Inquiry Concerning the Principles of Morals* (NY: Bobbs-Merrill Company, Inc., 1957 [1752]), IX, 93.

[25] Hume, *Treatise*, II,1, i.

[26] Hume, *Inquiry*, IX.

[27] Ibid.

[28] Hume,*Treatise*, III, 1, i.

[29] Ibid., III, 3, iii.

[30] Ibid., III, 3, i

[31] Ibid., III, 2, ix.

[32] Hume, *Inquiry*, Appendix III.

[33] Hume, *Treatise*, III, 1, ii.

[34] Ibid.

[35] Hume,*Treatise*, III, 3, i.

[36] Hume, *Inquiry*, 124, n. 3.

[37] Ibid.,, *Treatise*, III, 1, i.

[38] Ibid., II, 1, vii and III, 1, i.

[39] Ibid., II, 3, i.

[40] Ibid.

[41] Smith, *Theory/Kelley*

[42] Smith, *Theory/Selby-Bigge*, § 268: " The utility of those qualities [of the man of science and taste], it may be thought, is what first recommends them to us; and, no doubt, the consideration of this, when we come to attend to it, gives them a new value. Originally, however, we approve of another man's judgment, not as something useful, but as right, as accurate, as agreeable to truth and reality ... Taste, in the same manner, is originally approved of, not as useful, but as just, as delicate, and as precisely suited to its object."

[43] Smith, *Theory/Kelley*, III, 3.

[44] Smith, *Theory/Selby-Bigge*, § 279.

[45] Smith, *Theory/Kelley*, I, 1.

[46] Ibid.

[47] Ibid., I, 1, 7-8.

[48] Ibid., I, 1, 3.

[49] Smith, *Theory/Selby-Bigge*, § 299.

[50] Ibid., § 305. This may be compared with Hume: "...morality is determined by sentiment. ... virtue [is] whatever mental action or quality gives to a spectator the pleasing sentiment of approbation; and vice the contrary." *Inquiry*, 107.

[51] Smith, *Theory/Kelley*, I, 1, 1.

[52] Smith, *Theory/Selby-Bigge*, § 293.

[53] Ibid., § 304

[54] Ibid.

[55] Smith, *Theory/Kelley*, III, 5.

[56] Smith, *Theory/Selby-Bigge*, §317.

[57] Ibid., §314.

[58] Roderick Firth, "Ethical Absolutism and the Ideal Observer," *Philosophy and Phenomenological Research* XII, no. 3 (1952): 317–345.

[59] Ibid., 318.

[60] Ibid., 324. A "unique particular" such as Socrates, according to Firth, is merely a disguised universal; Crito would have the same reaction to *any* person who had the qualities he attributes to an accidentally unique particular (Socrates), for example, being the wisest man in all of Athens.

[61] Ibid., 324. But this analogy becomes more difficult to follow the closer it is looked at. While it appears incontrovertible that secondary-quality and ethical-dispositional statements have the same form, the elements of concreteness (or "reality") underpinning the two statements are vastly different in degree. The yellowness of the daffodil typically does not need ideal conditions to be perceived, nor does it need an ideal observer or impartial spectator, as Gilbert Harman notes (Gilbert Harman, *The Nature of Morality* [NY: Oxford University Press, 1977], 44. Any ordinary unimpaired person in ordinary light will perceive the daffodil as yellow (cannot help perceiving it as yellow), because it "is" yellow, regardless of how "is" is construed, e.g., that the daffodil refracts, absorbs, and reflects certain parts of the light spectrum in certain ways, or that certain primary properties of the daffodil have "powers" to "cause" a sensation of yellow in the spectator. But the rightness or wrongness of an act does not and arguably cannot have

correspondingly "real" underpinnings to stimulate the ethical dispositions of an ideal observer (at least, this question is exactly what has been in dispute in most of the history of ethics), and if it is these very dispositions that define rightness and wrongness, there is the potential of circularity. Firth addresses this issue by making a distinction between the ideal observer's experiences of moral data and his moral beliefs about that data. Moral data are "... the moral experiences to which we appeal when *in doubt* about the correct solution of a moral problem, or when attempting to *justify* a moral belief. For the epistemic function of moral data, when defined in this way, will correspond to the function of color sensations in determining or justifying the belief that a certain material object is 'really yellow.' And in that case moral data could play the same role in the analysis of 'right' that color sensations play in the analysis of 'really yellow.' " (327)

[62] Ibid., 335.

[63] Ibid., 336.

[64] Ibid., 338.

[65] But Firth suggests at a later point, without qualification or explanation, that "it seems unlikely that an ideal observer who had no interests at all would ever have any ethically significant reactions." (340) It may turn out that much serious discussion of impartiality will arrange itself around the idea in this throwaway line, particularly in view of Firth's conviction that, although analysis of emotion in impartiality discussions is analogous to analysis of interest, complete lack of emotion would not interfere with impartiality while complete lack of interests would.

[66] Ibid., 340.

[67] Ibid., 340.

[68] Ibid., 341.

[69] Singer, *Generalization*, 13.

[70] Firth, *Absolutism*, 342.

[71] Ibid.

[72] Ibid., 343.

[73] Ibid., 344.

[74] Ibid.

[75] Harman, *Nature of Morality*, 51.

[76] Immanuel Kant, *Foundations of the Metaphysics of Morals.*, trans. Lewis White Beck (Indianapolis, IN: The Library of Liberal Arts, 1959 [1785]), Second Section, p. 30.

[77] Henry Sidgwick, *The Methods of Ethics* , 7th ed. (Indianapolis, IN: Hackett Publishing Company:1981 [1907]), I, iii, 3: 34-35.

[78] Anaximander, via Simplicius, corrected by Heraclitus: "...and the source of coming-to-be for existing things is that into which destruction, too, happens 'according to necessity; for they pay penalty and retribution to each other for their injustice according to the assessment of Time,' as he describes it in these rather poetical terms." G.S. Kirk et al., *The Presocratic Philosophers* (Cambridge, England: Cambridge University Press,1983), 193-94.

[79] However, Bernard Williams makes the point that intelligent cruelty requires sympathy as well. Bernard Williams, *Ethics and the Limits of Philosophy* (Cambridge, MA: Harvard University Press, 1985), 90.

[80] See, for example, Kierkegaard, in "You Shall Love Your *Neighbour*". In M. Pakaluk, ed., *Other Selves: Philosophers on Friendship* (Indianapolis, IN: Hackett Publishing Company, 1991), 233-247.

CHAPTER IV

DETACHMENT AND CONTRACT THEORIES

Background

The previous chapter discussed a mode of impartiality that operates through the mechanisms of sympathy and benevolence. The embodiment of that mode was the ideal benevolent sympathetic spectator, who sympathetically identifies with all parties and has full knowledge and complete disinterestedness, and is ideally rational.

As became quickly evident, no matter how it was looked at, sympathy as a mode of impartiality ended by either closing off the moral community by stressing the naturalness of sympathy, and by extension the naturalness of the judgements that are made under its influence, or by making the idea of moral deliberation and a moral community unattainable by an ideally sympathetic individual.

In this section, I will examine a different mode of impartiality, an impartiality of explicit detachment. A good way to see this kind of impartiality in action is by looking at contract theories, since most contract theories either implicitly or explicitly require that the moral deliberator in the contract situation be detached from his personal interests. The detachment evident here is usually at least theoretically tied to ideas of equality, although it will be seen that the interaction of detached impartiality with ideas of equality can have some difficult and counterintuitive outcomes.

The impartial chooser of contract theories differs from the ideal spectator in

several ways. First, and fundamentally, the impartial chooser must not only observe, reflect, and judge, but must also act (choose). The ideal sympathetic spectator appears to function as a conscience more than anything else (this interpretation is particularly vivid in Butler); and while conscience ought ideally to spur us toward action in accordance with its dictates, it is clear that akrasia, ill-will, self-deception, plain laziness, or any of a number of other factors often prevents us from so acting. However, in contract theories the impartial chooser is brought into existence precisely for the purpose of *acting* in accordance with the outcomes of his deliberations. Similarly, the situations of moral deliberation differ in that regard from those of the ideal spectator as well, in that whether they are concretely delineated or not, they are explicitly the venues for a certain kind of action (choosing) rather than only for judgement, reflection, or evaluation.

Further, the act of choosing is presumed to have consequences that one must then live with, together with one's fellow choosers and, generally, further generations must live with them as well. This aspect of the choosing situation makes the choosing not only morally, but also epistemologically fraught, since the impartial chooser does not possess perfect knowledge or ideal rationality or what Firth terms "omnipercipience," a universal imagination that allows him to visualize all facts and all consequences of any possible acts in any given situation.

It is emotionally risky to use the sympathetic aspect of the chooser, so that aspect is explicitly stripped away. There is a pragmatic acknowledgement that persons may not all be "naturally" sympathetic at all times, in fact, they may be equally naturally antipathetic, and that therefore sympathy is an unstable foundation upon which to make significant moral choices that will impact upon others and upon future generations. Accordingly, the impartial chooser is delineated by the following characteristics: he is

rational, however a particular theorist may construe "rationality," although not ideally so; he is self-interested, but in a general, abstract sort of way, stripped of considerations relating to his personal self, although not perfectly disinterested; he is not expected to be at all interested in the personal interests of others; and finally, not expected to be either benevolent or sympathetic. The impartial chooser, in short, is more like an ordinary human being than like a god-like ideal impartial spectator or even a humanly sympathetic individual (but whose sympathies may be suspect).

This fundamentally altered categorization of the spectator-turned-chooser suggests that a different mechanism of impartiality might be in play. But since not all contract theories have their assumptions about impartiality explicitly spelled out, a way must be found to bring them to the surface. I shall do this following the same informal template as in the preceding chapter, by playing out the implications and meanings the injunction to "Love thy neighbor as thyself" within this modality. While in the examination of sympathy, the focus was on what "love" could mean, and therefore what kinds of "neighbors" one could end up with, here the direction will be reversed. The whole point of a contract theory is to turn already-unloved potential adversaries or enemies into "neighbors," so I shall begin by delineating the kinds of "neighbors" one might have upon implementing the designs of a particular contract theory, and work backward from there to see what kinds of impartiality assumptions could lead to that outcome. Several elements of the chooser and the choosing situation will be examined, including self-interest, rationality, and equality.

Further, the characteristic abstract, general self-interest of this impartial chooser suggests that an examination of what "as thyself" could mean might prove fruitful. Thomas Nagel, in his *Equality and Partiality*[1] and other works as well, has located a

primary ethical tension, not just the well-known tension between personal interests and impartiality, but in the actual psychological processes in which we are more or less trapped as we shuffle back and forth between the two in an effort not only to live ethically but to live also in a way which satisfies us personally. Loving one's neighbor "as oneself" is therefore wide open to interpretation, since "as oneself" can theoretically be located anywhere on a long continuum between full personal self-interestedness and the "view from nowhere" as denoted under a detachment modality – fully detached impersonality.[2]

In contract theories, a rational self-interest is presumed to be the motivating force behind the act of choosing. Contract theories provide a certain kind of answer to the amoralist's question: "Why is there anything that I *should, ought* to do"?[3] Why, in other words, should not a rational being simply become a "parasite on [an existing] moral system"[4] since at first glance it seems obvious that if everyone else is behaving morally, and I am not, I can better my situation substantially provided I am careful to *look as though* I am behaving morally? (As some later contractarian theorists believe, such parasitism is exceptionally easy to pull off in a contracted society if one is not careful in the designing stages.)

But, as the device of the Prisoner's Dilemma (among others) shows, such a narrow interpretation of self-interest often ends by working against the self-interest of all parties.[5] Contract theories, on the other hand, offer a solution to these sorts of situations by demonstrating that it is possible to structure prior agreements that will be in everyone's, or almost everyone's, long-term interest, provided those individual immediate interests that could potentially interfere with this outcome are voluntarily and publicly given up ahead of time (or tacitly acknowledged to have been given up). The motivating

force is the promise of a kind of security and predictability of outcome in life situations structured similarly to the Prisoner's Dilemma (or, as David Gauthier terms them, "Hobbesian" situations).[6]

Again, it will be seen that "rational self-interest" varies widely in interpretation with different contract theories, which also makes a difference in the kinds of neighbors one might find oneself confronting when all is said and done.

Another issue that is essential in the contemplation of contract theories is the concept of moral equality, which makes its appearance here in a different way from the "natural" interpretations of equality to be found in sympathy theories. After some examination, it emerges that equality and impartiality may be, unsurprisingly, deeply in tension with one another, and that one frequently gives way to the other, with mixed results. Again, this depends essentially on how "equality" is interpreted -- whether as a flat numerical or categorical equality, or in a weighted, functional way that tends toward leveling the playing field while not counting everyone the same.[7]

Another issue to consider is that while it is the impartial chooser who is doing the choosing, he is not doing so in a vacuum. It is only recently that the choosing situation has been delineated explicitly in contract theories. This explicit delineation has helped to bring out the conditions of impartiality and to highlight the places where self-deception could take root; however, the characteristics of the choosing situation were always implicit in theories where they were not overtly spelled out.[8] An exploration of the structures of various "original positions" in contract theories will show not only that these starting conditions for impartiality vary widely in their interpretations of equality and lead thereby to different understandings of impartiality, but also that there is a conceptual feedback loop between the concept of what constitutes general human

interests shared by all – the concept of a human "self" – and the structure of the original choosing situation (and therefore of impartiality). The concept of the human self informs the structure of the situation, while the structure of the situation shapes and controls the characteristics of the chooser, who supposedly represents the distilled general interests of the human self. The obvious danger for the theorist is that he will begin where he already is, and then structure the situation and characterize the chooser in such a way as to simply ratify and embalm outcomes arrived at from this starting point as the outcomes of deliberations engaged in by that impartial chooser in this original position.[9] This problem can be seen clearly when one considers that how an original situation is defined in order to assure the impartiality of the chooser, and the type of impartiality that is thereby engaged, depend in an important way on whether the presuppositions of the human self in place are those of a rational interest-maximizer, an embattled and thereby warlike human who needs a sovereign to both control and protect him, or a compassionate or self-interested (or both) recognizer of the arbitrariness of life's initial lottery of endowments. The original choosing position is then defined accordingly, and sets the criteria and standards for the choices to be made. The outcomes thereafter are, unsurprisingly, those that were predicted.

I have differentiated the psychological mechanisms of contractarian impartiality from those of sympathy theories by collecting them under the general concept of "detachment." The main distinction is between the "taking-in" of another's viewpoint as opposed to a "detaching-from" one's own; although these mechanisms are certainly not mutually exclusive (they can and do co-exist), it makes a difference in theorizing and in outcome whether one is emphasized or the other. In contract theories, where sympathetic mechanisms are not an essential part of theorizing, and in fact can usefully be ignored

altogether, detachment mechanisms take the foreground. These will be seen to be more stable and reliable over time than sympathy mechanisms, but I will show that they also have their own problems, mostly in relation to how the ideas of equality, and their influence on the way impartiality is used, are played out.

A brief overview of some of the psychological elements of detachment, separately from their use in ethical decision-making, will reveal some areas where inherent weaknesses may be imported into the ethical area.

Psychological Issues

It has long been known that ordinary consciousness can "split" away from itself in pieces, that the human mind has the capacity to mediate complex mental activity in channels split off from or independent of conscious awareness. An ordinary example of this phenomenon is a relatively common occurrence: while driving long distances alone, a person is thinking about something, gets deeply involved in his thoughts, and suddenly comes back to himself ten miles later, without the slightest recollection of the actual act of driving or of the passing landscape. Yet he did drive, performing all the constant intricate temporo-spatial adjustments driving entails – maintaining speed, awareness of road conditions and traffic, stopping at lights, keeping a safe distance from the car ahead, checking the rear-view mirror regularly, and so on. But while "he" was so deeply involved in what he was thinking about, "who" was driving the car? These kinds of mild dissociative states occur regularly in most people's lives and are described by Milton Erickson, a pioneer of techniques of naturalistic medical hypnosis in the areas of chronic pain and certain kinds of mental disorder, as

> ...a period during which the limitations of one's usual frames of reference and beliefs are temporarily altered so one can be receptive to other patterns of association and modes of mental functioning that are conducive to problem-solving.[10]

Although Erickson was speaking of therapeutically induced hypnotic trance states, the modalities of both deliberate and unconscious alteration of perspective, or frames of reference, are for practical purposes the same. This is precisely what is asked of a moral decision-maker when the adoption of one or another stance of impartiality is called for. In the generality of this particular formulation it is clear that this temporary altering of frames of reference is the preliminary movement both for sympathetic identification and for detachment from personal interests. In the present case, the continued movement toward detachment (rather than toward sympathetic identification) aims for a location on what may be conceptualized as a continuum of dissociative states including mild everyday trance states, deliberate hypnotic alteration of perspective, automatic dissociative states such as those experienced in extreme emergencies or in dissociative and somatoform pathology, and conscious deliberate dissociation such as that aimed for (in varying degrees and for different purposes) in moral deliberation and in some forms of religious contemplation. One might also include the mental state associated with absorption in an intellectual problem or in a novel, play, or film.

In this mode of impartiality, one makes use of conscious and deliberate dissociation. One alters one's frames of reference, beliefs, and modes of mental function so as to separate oneself as far as possible from one's personal interests and from others' as well, in the service not of broadening perspective by identification with the interests of another, but of narrowing and paradoxically thereby generalizing perspective to the

interests that are presumed to be shared among persons as individuated rational beings. In most contract theories, it is potential individuated personality with its basic human desires, preferences, or utilities, that is being consulted to serve as the impartial chooser, separated from the actual individual with his or her particular interests. The distinction between the general interests of a potential individuality, which can take indefinitely many kinds of forms in actual individuals, and the particular interests of actual individuals, is precisely that particular interests are not shared by all actual individuals, whereas general interests presumably are. To be able to distinguish one from the other in one's own mind is part of the movement toward impartiality of this sort of detachment.

Detachment in Contract Theories

Contract theories depend on an understanding that individuals cooperating to undertake a collective task (typically to form a society) are, and morally can remain, mutually disinterested. Disinterestedness refers to naturally self-interested persons not having or taking a personal interest in others' personal interests; the movement toward this style of impartiality seeks a perspective of similar temporary disinterestedness in one's own personal interests as well. In so doing, one detaches from one's particularized self and tries to adopt the perspective of a pure potential individuality, the success of whose ultimate actual individuation depends upon being safeguarded from unwarranted intrusion and constraint by the unrestrained personal interests of others. Having successfully adopted this perspective, one can then proceed to choose (or evaluate existing) principles upon which a just society ought to based.

But this is easier said than done. The danger here is once again Nagel's "view from nowhere." Stephen Darwall, in *Impartial Reason*,[11] remarks that impartiality and self-

interestedness (in a fully personalized sense) are in principle incompatible;[12] following this idea, might not one unwittingly "detach" so far from the personal that one may forget what one is supposed to be doing, or perhaps not take it seriously? Hume is clear that a lively sense of self is necessary for sympathetic identification; what sort of sense of self might be needed for appropriate detachment if one is not to end up "nowhere" again? And how is one to arrive at this "appropriate" self that is neither too personal to be impartial nor too impersonal to be human (and therefore, moral)? As noted above, it is presumed that there is a core of general interests that all persons share by virtue of being persons, and it is the conception of this core that constitutes both the abstracted human "self" (self as subject, as Gauthier terms it)[13] and sets the mandates of the original choosing position, with a particular presupposition of this self informing the construction of each original choosing situation.

Impartiality is not a given, not immediately present by fiat in the choosing or contracting situation; it is brought into being because of other factors, usually either a recognition of the essential human equality of powers, or a specifically tailored understanding of human rationality, depending on the characteristics of the chooser. In the contract theories of Hobbes, Rousseau, and Rawls, the qualities of the chooser are less explicit and more hidden, typically submerged in the perceived or imposed equality of persons, which then becomes the prime characteristic of the chooser, that is, being one among equals. I shall call these "equality" theories to distinguish them from a different class of contract theories to be discussed later. In Hobbes, for instance, the chooser is already perceived as equal in power to everyone else *prior* to the choosing or contracting situation; the contracting situation in fact comes about as a result of this equality. Recognition of this equality forces a type of impartiality that is based in all persons'

physical and intellectual powers being equal-enough to create danger for everyone unless something is done. In Rousseau, the chooser is not characterized explicitly at all but is submerged in the "general will" of the corporate body of which he is an equal member with other members; the general will expresses the decisions of the essential self and thereby urges choices that are impartially benevolent, directed toward the health of the entire body. For Rawls, on the other hand, the impartial chooser is forced by the characteristics of the original position into an equality of unindividuated personhood by being stripped of certain key elements of his personal knowledge, specifically, his knowledge of the form that his possible *in*equality may take due to the initial morally-arbitrary lottery of endowments. Since he will not know ahead of time where he will start out, he must strip himself to his essence in a very conscious way in order to maximize his situation should he be required to begin life in a worse-off position.[14] Equality, in these theories, drives the transition from the state of nature to civil society, or from unjust or non-just to just society, and subsequently requires a specific implementation of a detached impartiality in order to effect the transition.

In the other class of contract theories, such as those of Locke, Nozick, and Gauthier, the chooser is visible in clearer outline, having left less of his personal self out of the picture. In these theories, which I shall term "Lockean," the delineation of the chooser is sharper and more explicit. In Lockean theories it is equality which is initially submerged, to resurface later, but in an attenuated form based on an initial freedom of association. As in the "equality" theories, rationality is the soul and motivator of the impartial chooser. However, how "rationality" is to be construed is here also dependent upon the presuppositions about the self in place. It stands to reason that a self believing he is equal in principle to *all* others, and vice-versa, will construe rational decision-making

differently from a self who knows that others who are not equal in certain key respects will be excluded from the outcomes of the original situation. For example, the Lockean theories are explicit in their desire to arrive at a place where personal self-interest, usually expressed in property rights, can be secured and made safe from predation; original decision-making is oriented toward these kinds of goals. Left out of consideration is concern with the arbitrariness of initial endowment and how that might impact upon the ability to acquire property in the first place; the resulting society, therefore, will necessarily leave out of account concern with those persons who do not fit the initial presuppositions of self, the endowments required to have the ability to acquire property. Gauthier and Nozick are quite explicit about this; Locke is more indirect, but it is his formulation of who is and is not included in the society that other theorists build upon. This issue will be discussed more fully in Chapter VI.

These two types of contract theory have decidedly different characters and tend to suggest different interpretations of what it may mean to love one's neighbor as oneself, and thereby, different understandings of what it means to be impartial as well. The spirit of the theories built on equality suggests that because everyone is equal, and is to be publicly recognized as such, either initially or eventually, therefore, necessarily and dangerously, *everyone* is my neighbor, and loving my neighbor can easily be expressed in negative or positive versions of the Golden Rule (this is explicit in Hobbes): loving my neighbor means doing (or not doing) unto him as I would (or would not) be done by.

In Lockean theories, this is not the case, at least not at the initial stages of a society built upon principles of rationality and self-interest. Rationality not only trumps equality in these theories, even under differing conceptions of rationality, but it eliminates equality altogether from the choosing situation, only to reinstate it in somewhat rarefied

form at a later stage. Impartiality in these theories is motivated by and characterized by individuals' rational preferences or interests, but significantly, only those of economically productive or self-sufficient individuals, those whom Locke terms the "Industrious and Rational."[15] Those who cannot or will not produce (the "Quarrelsome and Contentious") are simply not invited into the construction of the moral community to begin with, and impartial rationality (and thus morality) proceeds over their heads, as it were. Who one's neighbor actually is, therefore, in these theories is a safe matter of prior selection based in the assumed right of freedom of association among those who share certain interests. It is obvious that here we may not end up much farther away from where we finished in considering sympathetic engagement as a modality of impartiality; we may find ourselves in yet another closed, exclusive moral community. It is worth noting, however, that in the case of sympathy, this restriction of the moral community is arrived at more or less innocently, as an artifact of sympathy and benevolence, whereas in Lockean theories it is an explicit structural and exclusionary factor from the beginning.

How the shape of the moral community is influenced by prior conceptions of the person, and therefore by the types of impartial rationality in play, will emerge in the following two chapters, which give a brief survey of highlights of some leading contractarian theories in each category.

NOTES

Chapter Four

[1] Thomas Nagel, *Equality and Partiality* (NY: Oxford University Press, 1991).

[2] It will be recalled that the "view from nowhere" within a sympathy modality involves a total sympathetic absorption of all points of view whatever, a *hyper*-personality, so to speak. Within the modality of detachment this same "view from nowhere" would, on the contrary, be completely *impersonal*.

[3] Bernard Williams, *Morality* (Cambridge, England: Cambridge University Press, 1993 [1972]), 3-4.

[4] Ibid., 5.

[5] R. D. Luce and H. Raiffa, *Games and Decisions* (New York: John Wiley & Sons, Inc., 1958), pp. 95-102 ff.

[6] David Gauthier, *Morals by Agreement* (Oxford: Clarendon Press, 1986), Chapter V and elsewhere.

[7] Nagel: "...egalitarian not merely in the sense that it counts them all the same as inputs to some combinatorial function, but in the sense that the function itself gives preferential weight to improvements in the lives of those who are worse off as against adding to the advantages of those better off." *Equality & Partiality*, 12.

[8] Interestingly, in some of those theories, Hobbes' for instance, the processes of impartial contractual choice are summed up in the injunction to "Love thy neighbor as thyself" as well as in positive and negative versions of the Golden Rule. (Thomas Hobbes, *Leviathan* [Middlesex, England: Penguin Books, Ltd., 1968 (1651)]), Chapter 30 and elsewhere.

[9] Rawls' concept of reflective equilibrium, another feedback loop wherein intuitive ideas are balanced against derived ideas, both subsequently mutually adjusted till they match, yields a more benevolent outcome, largely because the circular path between the sets of ideas is not only explicitly acknowledged but is indeed an integral part of the process. For example, the "intuitive" ideas are not random or instinctive, but carefully birthed as "considered judgements" in an environment that will give them their most honest formulation. This process will be discussed below in the section on Rawls.

[10] Milton H. Erickson, "Hypnotic Alteration of Sensory, Perceptual, and Psychophysiological Processes," in *The Collected Papers of Milton H. Erickson, Vol. II*, ed. Ernest L. Rossi (NY: Irvington Publishers, Inc., 1980), 3.

[11] Stephen Darwall, *Impartial Reason* (Ithaca, NY: Cornell University Press, 1983).

[12] Ibid., 17. Darwall imposes a crisp dichotomy between impartial reason and what he terms "self-centered[ness]." In Chapter 2 he explicitly seeks to "place a wedge between reasons and desires."

[13] Gauthier, *Morals by Agreement*, 7.

[14] Stephen Darwall (among others) has argued that it is "a widespread misconception that the parties to the original position are self-interested in the sense that their exclusive motivation is to secure the greatest utility for themselves once the veil is lifted." (Stephen Darwall, "Is There a Kantian Foundation for Rawlsian Justice?" in *John Rawls' Theory of Social Justice*, ed. H. Gene Blocker and Elizabeth Smith [Athens, OH: Ohio University Press, 1980], 232-235.) But leaving exclusivity out, self-interest certainly is an integral and logical part of the motivation (otherwise why would there be a need for a veil of ignorance in the first place?), and Rawls himself encourages this idea. See, for example, *Theory of Justice*, 19.

[15] John Locke, *Two Treatises of Government* (NY: New American Library, 1963 [1690]), §34.

CHAPTER V

CONTRACT THEORIES BASED IN EQUALITY

Hobbes

Hobbes' system calls upon what seems at first to be the simplest and most accessible form of detachment -- detachment from the illusion that human persons differ from one another in their fundamental physical or psychological nature. He constructs the initial situation out of a plain recognition of the essential equality of appetites, aversions, hopes, and powers among humans which leads, by the simplicity and strength of his argument, inexorably to "warre ... of every man, against every man" and renders individual life in the state of nature "... solitary, poore, nasty, brutish, and short."

> Nature hath made men so equall, in the faculties of body, and mind; as that though there bee found one man sometimes manifestly stronger in body, or of quicker mind then [sic] another; yet when all is reckoned together, the difference between man, and man, is not so considerable, as that one man can thereupon claim to himselfe any benefit, to which another may not pretend, as well as he. ... And as to the faculties of the mind, ... I find yet a greater equality amongst men, than that of strength....From this equality of ability, ariseth equality of hope in the attaining of our Ends. And therefore if any two men desire the same thing, which neverthelesse they cannot both enjoy, they become enemies; and in the way to their End, ... endeavour to destroy, or subdue one an other. ... Hereby it is manifest, that during the time men live without a common Power to keep them all

> in awe, they are in that condition which is called Warre; and such a warre, as is of every man, against every man.[1]

The seeds of impartiality are already present in this universal equality, specifically in the desire and mandate for self-preservation, as well as in the Right of Nature, the

> ... Liberty each man hath, to use his own power, as he will himselfe, for the preservation of his own Nature; ... and consequently, of doing any thing, which in his own Judgement, and Reason, hee shall conceive to be the aptest means thereunto.[2]

The detachment required to achieve an explicitly and consciously impartial stance in order to move from a state of war to a condition of peace is not so much detachment from one's own particular interests; it is detachment from what drives the *fulfillment* of those particular interests, the Right of Nature. If anyone can and may do anything according to one's own Reason and Judgement, then impartiality of a certain sort is already built into the prior system; it is a force of nature, as impartial as the weather and as unconscious to the potential contractor. It is only when the chronic state of war has become intolerable that it is time to consider giving up, or detaching from, the Right of Nature; the Reason of Nature (as opposed to one's own reason) comes to light and becomes explicit in the form of the "Fundamentall and Second Laws of Nature":

> *That every man, ought to endeavour Peace, as farre as he has hope of obtaining it;* ...the first, and Fundamentall Law of Nature; which is, *to seek Peace, and follow it.* ... From this Fundamentall Law of Nature, by which men are commanded to endeavour Peace, is derived this second Law; *That a man be willing, when others are so too, as farre-forth, as for Peace, and defence of himselfe he shall think it*

> *necessary, to lay down right to all things; and be contented with so much liberty against other men, as he would allow other men against himselfe.* .. This is that Law of the Gospell; *Whatsoever you require that others should do to you,that do ye to them,* And that Law of all men, *Quod tibi fieri non vis, alteri ne feceris.*[3]

"Every man," then, must become a Reasoner of Nature and agree that setting aside his Right of Nature, no matter how inimical to his immediate personal interests it might appear, is the only true rationality, because it springs essentially from his equality with everyone else, and it is the only thing which can safeguard his current and future interests. This mutual setting-aside of the Right of Nature generates the only condition under which the contract can become possible. With the birth of the contract comes the birth of conscious, explicit impartiality, in contrast to the unconscious force-of-nature "impartiality" reigning, by virtue of essential human equality, in the state of nature.

It is instructive to look now at who therefore must be included in the resulting moral community: *everyone,* without exception. In the "Nature hath made men so equall" passage, Hobbes has set the stage for the transition from the "warre ... of every man, against every man" inexorably to the *contract* of "every man [with] every man" – there is no room after this passage to carve certain moral kinds or classes of persons out of the contract. Indeed, it is not easy to see consistently upon what grounds these kinds or classes of persons may even be identified and distinguished from one another.

If one is minded to do so, one may evaluate the strength of a moral community by looking to see who is included in it and who is left out. A moral community that matter-of-factly makes room for the sick, the disabled, the poor, the insane, the very old, the very young, even the criminal and the lazy, is arguably stronger than one which

systematically excludes any of the above. Excluding all of these would leave people obviously more similar to one another, and we have seen in the chapter on sympathy that it is very easy to be sympathetic to those similar to oneself; if almost *everyone* is like oneself, morality must also be correspondingly easier and the institutions in place to maintain morality need not be very strong – a "minimal" government such as Nozick envisions[4] may be all that is necessary. On the other hand, if all are formally included, but are actually unwelcome, laws might need to multiply and proliferate to keep them "in their place" – for example, through expanded institutions of incarceration, involuntary civil commitment, juvenile institutions, nursing homes, and various other forms of human warehousing.[5]

"Similarity" here does not refer to obvious superficial, personal interests or values; valuing art, say, more highly than business, or golf more than baseball. The point is that it can, and in contract theories presumably is intended to, refer to the understanding of the *essential* self, stripped or detached from personal desires, values, and interests. However, it is evident that theoretical selves such as these can be stripped or detached in specific ways from specific things; the core of what constitutes humanity can therefore be understood in vastly different ways. Even self-consciously seeking to characterize a Kantian, "noumenal" self, can import biases into its construction.[6]

But morality seems to require more of a society which includes everyone without qualification and genuinely regards all its members as moral equals and potential participants. A place must be found for all the types of persons listed above, and more besides. The institutionalization of equality, whether recognized beforehand or imposed later, means that *everyone* must be taken account of, and the appropriate implementation of these institutions means that one must from time to time remind oneself that this or

that undesirable or inconvenient person merits recognition as an equal, at least in essential respects, the same as anyone else does. A convenient name that is later attached to this recognition is "respect," an interpretation of the love in "love thy neighbor" that does *not* rely on sympathy and can be called forth even when one feels decidedly unsympathetic. So when one's neighbor is recognized one's moral equal, one may respect him as such, as one respects oneself, no matter how apparently dissimilar the neighbor might be to oneself. The appropriate implementation of institutions of equality might then include not only respectful care of the elderly, poor, and infirm, but also respectful punishment of wrongdoers and respectful correction of the lazy (perhaps by public disapproval). A moral community built on institutions of equality that embody respect, in short, has room for compassion, as opposed to mere sympathy, where compassion can operate between dissimilars and in the face even of disapproval or dislike.[7]

Hobbes is explicit about some of these issues, for both prudential and compassionate reasons:

> ...needy men, and hardy, not contented with their present condition ... are enclined to continue the causes of warre; and to stirre up trouble and sedition.[8]

The fifth Law of Nature, Compleasance, suggests that one who will not aid the needy, by hoarding riches that others require as necessities, is to be excluded from the commonwealth:

> ... a man that by asperity of Nature, will strive to retain those things which to himselfe are superfluous, and to others necessary; and for the stubbornness of his

> Passions, cannot be corrected, is to be left, or cast out of Society, as combersome thereunto. For seeing every man, not onely by Right, but also by necessity of Nature, is supposed to endeavour all he can, to obtain that which is necessary for his conservation; He that shall oppose himselfe against it, for things superfluous, is guilty of the warre that thereupon is to follow; and therefore doth that, which is contrary to the fundamentall Law of Nature, which commandeth to seek Peace.[9]

Finally, Hobbes makes the duty of the Sovereign in this regard explicit:

> ...whereas many men, by accident unevitable, become unable to maintain themselves by their labour; they ought not to be left to the Charity of private persons; but to be provided for, (as far-forth as the necessities of Nature require,) by the Lawes of the Common-wealth. For as it is Uncharitablenesse in any man, to neglect the impotent; so it is in the Soveraign of a Commonwealth, to expose them to the hazard of such uncertain Charity.[10]

The inclusion of all into the moral community is thereby made explicit, as a necessary outcome of the contract, with specific duties of inclusion falling upon both the private citizen and the sovereign. Throughout the construction of his argument, Hobbes reminds us that although persons may appear different from one another, they are nevertheless "equall"; and the outcome is necessarily that therefore they are to be cared for if they are needy or disabled. In fact, it is the miser hoarding his superfluity in the face of another man's need who is to be cast out of the society, as not able to comprehend the meaning of equality.

In Hobbes these exhortations are the inevitable result of his argument but are essentially unnecessary (since the argument speaks for itself in its common sense and

matter-of-factness); but in Rousseau the same issues are handled in a way that takes a subtly more disturbing turn.

Rousseau

In *The Social Contract* Rousseau seeks to uncover the root of the legitimacy of the rule of administration and considers first whether force is what legitimates civil administration. If it turns out not to be force, then it must be an agreement of some sort. He begins by looking at "[t]he oldest of all societies, and the only natural one," – the family. Immediately Rousseau presents the presuppositions of freedom and equality that resurface later in this work both as the explicit goals of law and as evidence for the existence and operation of conscience:

> ... the family is the first model of political societies: The father corresponds to the ruler, the children to the people; and all, *having been born free and equal,* give up their freedom only for their own advantage.[11] [emphasis added]

It is a radical idea to consider that parents and children are equals, and moreover that they were all born free, only giving up their freedom for the sake of their own advantage. Rousseau contrasts this view of the original nature of humankind with that of Grotius (who "leans" toward the view that "the human race belongs to a hundred men"), Caligula ("...either kings [are] gods or peoples [are] animals"), and Aristotle ("... men [are] not naturally equal, but some [are] born to be slaves and others to be masters").[12] But examining arguments in support of the supposed legitimacy of slavery shows the

weakness at their core, namely that the proposed essential inequality of human beings upon which this institution rests is spurious. On the one hand, rule by force is self-contradictory and cannot be maintained because it presupposes that might makes right; and

> ... [a]ny might greater than the first will take over its right... and since the strongest is always in the right, one has only to act in such a way as to be the strongest. But what kind of right is it that ceases to exist when strength perishes? If a man is forced to obey, he no longer has any obligation to do so. It is clear that the word "right" adds nothing to force ...[13]

On the other hand, one cannot voluntarily agree to be enslaved; one cannot sell one's own freedom, because in so doing one sells one's own humanity as well, and thus vitiates morality altogether:

> A man who renounces his freedom renounces his humanity, along with the rights of humanity, and even its duties. ... Such a renunciation is incompatible with the nature of man, and to remove all freedom from his will is to remove all morality from his acts.[14]

Not even soldiers making an agreement with their captors to allow themselves to be enslaved in exchange for their lives is legitimate, because war is a relation of things and not men. The presumed right to kill an enemy soldier is terminated at the instant the soldier lays down his arms; at that point, he ceases to be an enemy, so the right to kill, and therefore also enslave, him vanishes.[15]

Therefore, the legitimacy Rousseau seeks is that which obligates by duty rather

than by force -- that which persuades rather than compels a minority to obey a majority, that which is presupposed in the giving of a people to a king. "[A] people is a people before giving itself to a king," but what is the "agreement by which a people is a people? Since the latter necessarily precedes the former, it is the true foundation of society."[16] Force as a legitimator is predicated upon the presumed natural inequality of human beings; but given that humans are presumed to be equal, what could motivate them to be ruled by an administrative body? There must be a prior agreement, based in the recognition of not only the equality but also of the ultimately inadequate strength of men to survive on their own:

> I suppose men to have reached the point where the obstacles to their survival in the state of nature have a resistance that cannot be overcome by the forces each individual has at his disposal for preserving himself in that state. ... Since men cannot engender new forces, but can only unite and direct existing ones, they now have only one means of preserving themselves: to form by aggregation a sum of forces capable of overcoming the resistance, then direct them toward a single goal and make them act together.[17]

This leads to the paradox of the social compact – how to both give up and yet simultaneously retain the strength and freedom of each individual, his "primary means of self-preservation":

> The problem ... can be stated as follows: 'To devise a form of association which will defend and protect the person and possessions of each associate with all the collective strength, and in which each is united with all, yet obeys only himself and remains as free as before.'[18] [quotation marks in text]

The terms of this contract are "everywhere the same" and require the "complete surrender of each associate, with all his rights, to the whole community."[19] This surrender solves the problem of the social compact, for it maintains each individual's equality and thereby his motivation ("since each man gives himself entirely, the condition is equal for all, and [therefore] it is to no one's interest to make it burdensome for others.").[20] Further, "since the surrender is made without reserve"[21] the union is perfectly balanced and thereby fortified against tyranny. Last, the paradox is resolved, because

> ...in giving himself to all, each man gives himself to no one, and since he acquires the same right over all the other associates as they acquire over him, he gains the equivalent of everything he loses, plus greater power to preserve what he has.[22]

This is summarized in the following formulation:

> 'Each of us puts his person and all his power in common under the supreme control of the general will, and we collectively receive each member as an indivisible part of the whole.'[23] [quotation marks in text]

and thereby

> ... the fundamental pact substitutes a moral and legitimate equality for whatever physical inequality nature has produced among men, so that while they may be unequal in strength or intelligence, they all become equal by agreement and rights.[24]

The general will embodies and represents the impartiality of Rousseau's social contract: "... an individual will ... tends toward partiality, and the general will tends toward

equality."[25] Here, by countering "partiality" with "equality," Rousseau makes explicit the impartiality born of equality that was implicit in Hobbes. Equality in its turn assures that the "general will is always enlightened and the people never mistaken."[26]

But how can that be, and what is the general will? It is the will of the "public person formed by the union of all other persons ... [the] body politic."[27] It is not necessarily unanimous, but all the votes must be counted, because "any formal exclusion destroys generality"[28] and throws the members back into partiality. Furthermore, it stems from "the preference that each man gives to himself, and therefore from the nature of man"[29] – this is why the general will is always well-meaning and why everyone "constantly will[s] the happiness of each individual."[30] This may be one interpretation of the relationship between self-interest and impartiality that Hume hinted at when he said that one must have a "lively sense of self" in order to experience sympathy. The sympathy here has Butlerian echoes in its organicity – it is in its public form, benevolence, and exists only because each person has first detached himself from his partial interests and surrendered his freedom completely to every other person, thereby reconstituting himself as a member, with everyone else, of a single corporate body, the "public person." Butler predicated both self-love and benevolence on persons being all members of "one body in Christ," which generates an individual conscience and allows each individual to be "a moral agent, ... a law to himself."[31] The general will, in its turn, is the expression of the newly formed moral individual, a member, together with all the others, of the public person, who has "moral freedom, which alone makes man truly his own master."[32]

Loving one's neighbor "as oneself," therefore, is explicit in Rousseau. "As oneself" quite literally refers to all individuals being as members of one body, which

generates perfectly logically the desire to do well by that body by wishing happiness to all other members as one does to oneself. It is as illogical for the general will to wish harm to an individual member as it would be for a human person to wish harm to his own left foot. The benevolent impartiality that one feels toward all one's members – left foot, right arm – is exactly what Rousseau has in mind here.

But what does the general will say about who one's "neighbor" is? Here is the real paradox that threatens the integrity of the Rousseauean contract. Rousseau made the interesting points that "any formal exclusion destroys generality,"[33] but, in another context, that "anyone who refuses to obey the general will shall be compelled to do so by the whole body. ... forced to be free."[34] This suggests not only that everyone *ought* to be considered one's neighbor, but in fact that everyone *must* be considered one's neighbor, for otherwise the public body is at least corrupted, if not destroyed outright. Furthermore, since exclusion destroys the general will, it appears that any attempt to exclude must be countered with force, and the inclusion of all is compulsory. In sum, therefore, it appears that in voluntarily surrendering oneself to everyone else, one has not only agreed to the creation of an institutionalized compulsion to obey law, but also to the compulsory inclusion of everyone as one's neighbor. This is obviously very different from Hobbes' education to justice.

Rousseau's "forced to be free" locution interprets in a troubling way the powers vested in Hobbes' sovereign – troubling because in Rousseau there may not be an effort made to make those powers understood and accepted by all the members from within themselves; they are instead imposed upon some or even many members by fiat and their acceptance insinuated into the public consciousness by outright mythology. While at first glance it is reassuring to know that *no one may be excluded* from the moral community,

upon reflection there is a hidden flavor of brutality in the complementary requirement that *everyone must be included* on pain of compulsion; somehow, the two formulations seem not to be equivalent or even compatible. This uneasiness is intensified by Rousseau's observation that the "larger a state becomes, the more freedom is diminished ... [it] moves farther away from equality ... [and] repressive force must be increased."[35] The general will is to the public body what our own individual will is for us: a will toward what is good for us as humans. But Rousseau, in making the point that neither we as individuals, nor therefore as members of the public body, always *know* what is best for us,[36] neglects to concern himself with the case in which we do know what is good for us and choose consciously in opposition to that end. One may then still say, consistently, that we *desire* the best for ourselves, or *wish* the best for ourselves, but in the face of actually choosing otherwise, it is hard to see what it might mean to *will* the best for ourselves. In that case, and also in the case of not actually knowing, combining compulsion and repression with either ignorance or willful wrong choice seems a dangerous thing to permit or even insist upon.

Thus, in the case of who one's neighbor is or must be, the idea of a general will to mandatory inclusion for the good of the public body, extended out from an individual will for the well-being of one's own members, seems wishful, vague, and un-anchored to any hard principle to guide action, much as the free-floating, all-knowing ideal observer was unable by his nature to appeal to any anchoring principle. The general will has, moreover, in this case as well as in the case of theories built upon sympathy, a formidable opponent in ordinary human nature. In relation to who one's neighbor is, it must take account of the existence, in many cases, of deep-seated *antipathies* to certain persons or kinds of persons that are arguably as "natural" and "instinctive" as one's sympathies are alleged to be. For

example, it may be as natural and instinctive for some to shrink from misfortune, poverty, disfigurement, and insanity as it may be for others to feel sympathy for those who suffer from these afflictions. In Hobbes, antipathy can be countered by compassion; but in Rousseau's own terms, as the population increases, and with it the incidence of such misfortune and illness, repressive measures may have to be instituted to enforce an equality that people don't really feel or experience or even see the need for. This seems to signify more than the ordinary impartial enforcement of law that is expected in a civil society. If the theory of the general will is explicitly available to the rulers, but not necessarily to the members, it can and probably will serve as the justification both for the exclusion of dissenters, even though this would be incompatible with the original intention to include all, and for compulsion of the rest.

Hobbes, for his part, has taken care to educate, rather than trick, the people to the ways of justice, with the intention of having people eventually behave justly "sincerely from the heart."[37] Having begun from a place of equal-enough mental and physical power, equal hope, and therefore equal danger, Hobbes believes that the teachings of justice (including the commandment "*Thou shalt love thy neighbor as thy selfe*"), are "so consonant to Reason, that any unprejudicated man, needs no more to learn it, than to hear it."[38] Equality is internally recognized by Reason as an initial condition which, if unregulated, can lead to danger, and is transformed subsequently to moral equality by the teaching of justice, thus obviating the need for overt compulsion and repression at this level.

But in Rousseau, because equality is assumed without argument and not further motivated, the theory of the general will leaves the door open for repressive enforcing measures that he suggests might be necessary for other reasons with an increase in

population. Further, Rousseau himself seems to believe that persons are not really so equal after all and should not really be permitted to be so free. To begin with, the permission for repression is combined both with an avowed imperfection in knowledge of the general will and with Platonic-Machiavellian protocols concerning how the general will is to be made known to the public (by "appeal to divine intervention ... [so that] their peoples ... will ... bear the yoke of public well-being with docility."[39]). The willingness to use these protocols thereby gives the lie to Rousseau's avowed conception of equality, since it gives permission to treat the persons of the civil body with the same kinds of attitudes that he himself decried in Caligula, Grotius, and Aristotle, that is, as if they were children or even inferior humans.

This leads to a peculiarity in this conception of impartiality: the general will is brought into being by individual persons in an avowed state of moral equality who detach themselves temporarily from their particular interests and come voluntarily together into a collective; this same general will sees itself as then being required to turn around and enforce an equality at the collective civil level that it now somehow presumes is foreign and unknown to the very individuals who would ordinarily be contracting in the above situation.

In Rousseau, the end result of such systemic coercion and manipulation can only be extreme but shallow conformity and ultimately, a compulsory, politically-correct, but brittle homogeneity of population. The love of one's neighbor that Hobbes expected to flow sincerely from the heart following education to justice is instead imposed by force, under the auspices of a spurious divine mythology. Clearly, if persons are to be actually motivated to recognize all others as their moral equals, this motivation must be educated so that it flows from within. Rawls' contemporary contract theory tries to re-establish

the "education to justice" that Hobbes relied upon.

Rawls

In his *A Theory of Justice*, Rawls makes explicit some of the reasons for the human tendency to exclude dissimilars and builds his theory of justice as fairness around the recognition of one of the origins of such exclusion, life's initial lottery of endowments:

> ... each person finds himself placed at birth in some particular position in some particular society, and the nature of this position materially affects his life prospects.[40]

Specifically, Rawls is concerned to argue for a conception of justice that recognizes and adjusts for the inequality, not of the moral status of persons, but of arbitrarily assigned starting points which, in their random distribution of advantages and disadvantages, may work against the ability of individuals to assume their own, or to recognize others', full equality of moral status. To aid his search for "a conception of justice that nullifies the accidents of natural endowment and the contingencies of social circumstance as counters in quest for political and economic advantage,"[41] he marshals the psychological mechanisms of reflective equilibrium, the original position, and the veil of ignorance to justify two principles of justice which he argues would together be chosen by impartial choosers behind a veil of ignorance in an original position. These principles embody the ideals of liberty, equality, and a missing element in other theories, which he seeks to reinstate, fraternity. The two principles are (in their final formulation):

First Principle:

> Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all.

Second Principle:

> Social and economic inequalities are to be arranged so that they are both:
>
> (a) to the greatest benefit of the least advantaged, consistent with the just savings principle, and
>
> (b) attached to offices and positions open to all under conditions of fair equality of opportunity.[42]

Partly because of the unusual and radical nature of the second principle, the so-called "difference principle," and its implicit overturning of the Lockean absolute property in the self, the psychological mechanisms that are used to delineate the original position, eventuating in the choice of these two principles, are of particular importance in an examination of styles of impartiality. The question is whether the particular kind of impartiality embodied in these mechanisms supports or weakens the legitimacy, as something reasonable persons *would* choose, of the derivation of these principles. His theory must draw upon these psychological mechanisms in order to construct the original position and delineate the characteristics of the impartial chooser. In his close attention to how these mechanisms operate and why they are to be used, he reveals a new wrinkle concerning the analysis of impartiality in general.

In previous discussions of both contract theories and sympathy theories, impartiality was implicitly taken for granted by theorists, presumably as the appropriate mindset to be adopted in moral deliberation, and indeed, this is probably the attitude adopted toward the state of being impartial by most moderately reflective and conscientious persons in ordinary life. The various analyses revealed the uncomfortable, if

not novel, fact that although "being impartial" sounds like a relatively straightforward state of mind to aim for, in fact preconceptions, assumptions, and various types of epistemological obstacles and psychological baggage ("Idols") tended to interfere with the attainment of a "pure" impartiality that might aid in nudging the outcomes of moral deliberation in the direction of something approaching objective moral "truth." What was evident in the analyses of how impartiality could be interfered with was how this interference seemed almost to guarantee exclusion, with Hobbes as an exception.

In Rawls, however, the concept of impartiality is made explicit, carefully analyzed and meticulously reconstructed to derive in the first instance as solid a foundation for subsequent deliberation as possible, the original position. The original position stipulates a formal moral and epistemological equality among deliberators by stripping them of the particular inequalities of their natural and social endowments. Two mechanisms or devices serve to assure the essential impartiality of this original position.

The first, reflective equilibrium, is a touchstone validating our description of the initial contractual situation whereby our considered intuitive convictions of justice "in which we have the greatest confidence"[43] are balanced against the principles of justice derived from that situation to see if they can be made to match. Initially, they probably will not; equilibrium between the two is eventually arrived at by alternately modifying and adjusting either our convictions or our derived principles until we have arrived at a balanced description of the initial contractual situation that "both expresses reasonable conditions and yields principles which match our considered judgements duly pruned and adjusted."[44] "Considered judgements," in their turn, are

> those rendered under conditions favorable to the exercise of the sense of justice,

> and therefore in circumstances where the more common excuses and explanations for making a mistake do not obtain.[45]

Judgements made hesitantly in which we have little confidence, or those made under conditions of high emotionality, or in circumstances where we have personal interests, are not "considered judgements." Considered judgements are "those judgements in which our moral capacities are most likely to be displayed without distortion"[46] -- in short, judgements derived under conditions approximating those optimal for the exercise of impartial deliberation.

The other device invoked to support the favored interpretation of the initial situation is the veil of ignorance – a listing of restrictions upon the arguments for the principles of justice in the initial situation, and "therefore on these principles themselves."[47] To assure impartiality in the choice of principles, the contracting parties must exclude from their knowledge those facts of life which individuate and particularize them: their place in society, their natural assets or abilities, their own conceptions of the good, the particulars of their life plans, special features of their own psychology, particular circumstances of their own society, and even what generation they belong to. They know only the general facts of life: they know that their society is subject to the circumstances of justice, and they know general facts about human society -- political affairs, principles of economic theory, the basis of social organization and fundamental facts of human psychology.[48] Rawls also makes the "special assumption [of rationality]," that none of the parties is afflicted with envy.[49]

These restrictions assure not only the impartiality of the parties, but also their status as equal moral persons. These persons are similar to one another in respect of their

having a conception of their good and being capable of a sense of justice;

> ...[t]ogether with the veil of ignorance, these conditions define the principles of justice as those which rational persons concerned to advance their interests would consent to as equals when none are known to be advantaged or disadvantaged by social and natural contingencies.[50]

The original position thus characterized will then presumably lead, by virtue of the rationality and forced neutrality of the parties, to selection of the two principles of justice as described rather than to other principles from the available alternatives.[51]

As we have seen so far, a state of impartiality can come about (at least theoretically) either by sympathizing with all parties or by detaching from one's personal interests without necessarily being sympathetic.[52] But here we ought to remember what impartiality is *for*. Briefly, if we invoke states of impartiality in order to "do what's right" without narrowing the horizon of options only to what will benefit us in particular, then there seems to be a need to look at what we mean when we say that a certain decision is "right" -- whether what we mean is something we intend to be objectively valid for all persons, or whether we will garner what is *considered* right from our peers and fellow citizens. Rawls makes his concept of right completely coherent with criteria that would have been imported into the original position as tests or standards for "right" things:

> ... the concept of something's being right is the same as, or better, may be replaced by, the concept of its being in accordance with the principles that in the original position would be acknowledged to apply to things of its kind.[53]

> ... a conception of right is a set of principles, general in form and universal in application, that is to be *publicly recognized* as a final court of appeal for ordering the conflicting claims of moral persons. [emphasis added][54]

This is not surprising, for the concept of reflective equilibrium, wherein our considered judgements and our derived principles of justice undergo mutual adjustment until they match, is also explicitly coherentist in this way, resting as it does on already-existing standards of rightness. (The only test, in other words, is a test *within* rightness so understood, and not *of* rightness.)

Further, we are meant to understand that the mechanism of the veil of ignorance strips us of particularized knowledge about ourselves, and that what is left is the more abstract and generalized self that knows only certain general things. In this way, decisions arrived at are as close to being universal ("objective") as imperfect humans can arrive at. But it is obvious that the list of items of general knowledge is far from being generally valid and homogeneous; it is clearly already so strongly biased that all ethical concepts subsequent to the choices made in the original position will reflect these biases, and "right" will come to mean no more than something that these already-primed individuals will all agree to as being right. This is not far from Rousseau's objection to might as right; all it takes is someone stronger – or a different, but coherent, set of values – to change the meaning of "right." The veil of ignorance, in other words, is translucent at least, if not transparent altogether. For example, Rawls tells us that, among other general knowledge, we know the laws of human psychology. But the "laws" of human psychology not only vary strongly among cultures, but also within the same culture at different times. A

culture that values personal humility and a readiness to sacrifice oneself for the good of the whole is going to seek and find "laws" of psychology that are radically different from those of a culture that values individual self-sufficiency and competitiveness, and the psychological concept of self-realization (or the moral concept of flourishing) is going to take correspondingly different forms in the two cultures, as will the concept of a noumenal self and also theories of personality and personhood. So an original position arrived at by means of Rawls' mechanisms must necessarily reflect the state and particulars of general knowledge of an already-existing society, no matter what efforts are made to generate an abstract, universal, or "noumenal" self to participate in the contract decisions.

However, in "On the Original Contract," Hume observes

> that though an appeal to general opinion may justly, in the speculative sciences of metaphysics, natural philosophy, or astronomy, be deemed unfair and inconclusive, yet in all questions with regard to morals ... there is really no other standard by which any controversy can be decided.[55]

This sentiment is echoed in Rawls when he tells us that there is a "standard" interpretation of the original position, which

> ...best expresses the conditions that are *widely thought reasonable* to impose on the choice of principles yet which, at the same time, leads to a conception that characterizes our considered judgements in reflective equilibrium.[56] [emphasis added]

In another context, he says,

> A conception of justice cannot be deduced from self-evident premises or conditions on principles; instead, its justification is a matter of the mutual support of many considerations, of everything fitting together into one coherent view.[57]

It appears, therefore, that for those who might decry this very coherentism in Rawls' theory and seek something more objective in their moral decision-making (or at least believe or hope that such greater objectivity may in principle be possible), far from the original position embodying a firm standpoint wherein impartiality is assured, in fact impartiality of a qualitatively different, presumably non-coherentist, sort must be invoked in order to assess the original position, and at a much earlier stage of the proceedings, and the concept of reflective equilibrium either modified or done away with altogether – unless, of course, impartiality in general is in fact to be understood, and *can* only be understood, as reflecting the coherent equilibrium of considered public judgements in a slice of time at a particular place.

In this regard, Ronald Dworkin identifies two models of what he calls "moral coherence".[58] In one model, which he calls "natural," there is presumed to exist an objective moral reality, and therefore it is possible that the moral universe can "contain" reconciling principles when moral intuitions conflict. In the second model, which he terms "constructive," there is a requirement of publicity and because of this something must *be constructed* to reconcile conflicting intuitions. These constructions must be consistent with what has gone before and provide a public standard for debate, thereby correcting for biases that may be present in unique intuitions. In the former epistemological model, moral intuitions are analogous to perceptions; in the latter, they are not. Dworkin

identifies reflective equilibrium as an example of the second model but concludes that "... principles of justice selected in this spirit are compromises with infirmity."[59]

Being impartial even in Rawls' careful and subtle sense has not yet allowed us to escape from the confines of potentially closed moral communities, even in the case of a moral community specifically constructed to be broadly if not universally inclusive, and to embody the values of liberty, equality, and (according to Rawls, the usually-missing) fraternity. This is because Rawls himself is clear, in a different context, that the basic structures of a society mold and shape its members' beliefs and their very personalities; whereas elsewhere he observes that one may enter the original position at any time, simply by performing the requisite mental acts to draw the veil of ignorance around oneself. It is an open question, therefore, how the veil of ignorance, as defined by Rawls, will allow the parties to the original position to put aside the beliefs already deeply ingrained by their presence in a certain place at a certain time in order to attend impartially to principles of justice that may not cohere with those beliefs. Even if this view of impartiality is not the point, and Rawls explicitly says it is not, the question then is to define impartiality in such a way that it "coheres" not only with a coherence argument, such as the one that is formalized in the procedures of reflective equilibrium, but also with our common understanding of impartiality in ordinary life as being "outside" what is being evaluated, in this case the deliberative framework itself.

Rawls addresses this kind of objection in his "Kantian Constructivism in Moral Theory"[60] by explicitly contrasting his style of Kantian constructivism with rational intuitionism. The key contrast is that in the latter there is presumed to be an independently existing moral order that it is for the rational individual to "find" and thereby be motivated to act upon. Rawls argues that this requires only a very "sparse"

conception of the person, "founded on the self as knower. ... since the content of first principles is already given, a more complex conception of the person ... is simply unnecessary."[61]

But in Kantian constructivism, the "framework for deliberation" relies upon our "powers of reflection and judgement," powers that need not be much developed for rational intuitionism but that for constructivism require ongoing development in a "shared public culture" and which are also "shaped by that culture."[62]

> [T]he principles adopted by the parties in the original position are designed by them to achieve a public and workable agreement on matters of social justice which suffices for effective and fair social cooperation. From the standpoint of the parties as agents of construction, the first principles of justice are not thought to represent, or to be true of, an already given moral order, as rational intuitionism supposes. The essential point is that a conception of justice fulfills its social role provided that citizens equally conscientious and sharing roughly the same beliefs find that, by affirming the framework of deliberation set up by it, they are normally led to a sufficient convergence of opinion.[63]

Impartiality, in this sense, clearly means impartiality *within* a framework of deliberation that is established by parties who cannot, and probably for practicality's sake should not, detach themselves from their already-existing socially-based understandings of general social, economic, and psychological laws. These understandings obviously lend themselves to the danger that, once again, principles chosen impartially (in this sense) from within that framework may operate to select out, however indirectly, certain kinds of persons, even if the framework is set up explicitly to not do so.

Hobbes, Rousseau, and Rawls' "equality" theories have been examined to see

whether a style of impartiality that takes as a primary principle a given equality among persons would succeed in breaking out of a closed moral community. So far, only Hobbes' argument has given the hope that this may be accomplished, with the proper education to justice of the citizens. Rousseau's theory seems to require mechanisms of mythmaking, combined with a compulsion far beyond ordinary law enforcement, leading to a society where equality is formally mandated but where large numbers of people potentially could actually be excluded from the opportunity to participate genuinely in the formation and operation of the society. Rawls' coherentist theory explicitly is directed toward an egalitarianism that takes differences of all sorts into account, offering fair equality of opportunity to all, and offering hope that "fraternity" may be reinstated next to liberty and equality; yet it may well founder on the quality of the materials available as considered judgements to begin with, since those judgements will be imported directly into the choosing situation as part of its construction and contrasted only with decisions arrived at within the choosing situation itself.

In the next chapter, a different category of contract theory will be examined, one not necessarily based in the explicit equality of persons. It will be seen that without the assumption of equality, the resulting moral communities are increasingly narrowly conceived.

NOTES

Chapter Five

[1] Hobbes, *Leviathan*, I, §13.

[2] Hobbes, *Leviathan*. I, §14.

[3] Hobbes, *Leviathan*, I, §14.

[4] Robert Nozick, *Anarchy, State, and Utopia* (NY: Basic Books, 1974).

[5] Interestingly, a similar pattern comes to light very readily in an examination of Rousseau, one of the "equality" theorists. See below on Rousseau.

[6] As Rawls has been accused of doing: see, for example, Thomas Nagel, "Rawls on Justice," in *Reading Rawls*, ed. Norman Daniels (NY: Basic Books, 1975), 10. Nagel reviews some criticisms of Rawls that note that his original position is biased toward liberal individualism and therefore that life conceptions and plans that are not liberal and individualistic are unfairly pushed aside.

[7] Distinctions between sympathy and compassion are difficult to pin down, but I believe they may boil down to an essential difference in "person," in Adam Smith's sense. Experiencing sympathy, I may feel "for" you; you are in a particular circumstance, and I consciously or unconsciously put myself in your place and feel what you are feeling, see the world from your point of view, and so on, either "as" myself or "as" you. For these reasons, similarity of outlook to begin with seems an important element in order to accomplish this temporary exchange of persons or perspectives successfully. Compassion, however, seems to involve something very different, perhaps an essential recognition that it isn't just "you" in that circumstance, but that it might just as well be "me" – not "me-as-you" but *actually* me, and it is just sheer luck or grace that I am where I am and you are where you are. Our situations might just as easily have been reversed. Therefore, compassion seems to be a feeling "with" rather than "for," embodying a recognition of an essential equality of possible circumstance that is only by chance what it is, and could just as well have been different.

[8] Hobbes, *Leviathan*, I, Ch. 11.

[9] Ibid, I, Ch. 15.

[10] Ibid., II, Ch. 30.

[11] Jean-Jacques Rousseau, "The Social Contract," in *The Essential Rousseau*, trans. Lowell Bair (NY: The New American Library, 1974 [1762]), I, ii, 9.

[12] Ibid., I, ii,10.

[13] Ibid., I, iii, 11.

[14] Rousseau, *Social Contract*, I, iv, 13. Kant subsequently makes the point as follows: "In submissiveness there is ... a certain ugliness. ... that man himself should stand in need of no soul and have no will of his own, and that another soul should move his limbs, ... [he] is no longer a man, he has lost his standing, he is nothing but the possession of another man." (Immanuel Kant, notes to the "Observations on the Feeling of the Beautiful and the Sublime," quoted in Ernst Cassirer, *Rousseau, Kant and Goethe* [NY: Harper Torchbooks, 1945], pp. 17-18.)

[15] Ibid., I, iv,13.

[16] Ibid., I, vi,16.

[17] Ibid.

[18] Ibid., I, vi, 17.

[19] Ibid.

[20] Ibid.

[21] Ibid.

[22] Ibid.

[23] Ibid.

[24] Ibid., I, ix, 23.

[25] Ibid., II, i, 24.

[26] "When partial societies do exist, they must be made numerous and prevented from being unequal, as was done by Solon, Numa, and Servius. These are the only effective precautions that can be taken to ensure that the general will is always enlightened and never mistaken." Ibid., II, iii, 27.

[27] Ibid., I, vi, 18.

[28] Ibid., Notes, n. 6, 116.

[29] Ibid., II, iv, 28.

[30] Ibid.

[31] Butler, *Five Sermons*, Sermon II, 37.

[32] Rousseau, *Social Contract*, I, viii, 21.

[33] Ibid., Notes, n. 6, 116.

[34] Ibid., I, vii, 20.

[35] Ibid., III, i, 50.

[36] Ibid., II, vi, 35.

[37] Hobbes, *Leviathan*, II, §30, 383.

[38] Ibid., 379.

[39] Rousseau, *Social Contract*, II, vii, 38.

[40] Rawls, *Theory of Justice*, 13.

[41] Ibid., 15.

[42] Ibid., 302.

[43] Ibid., 19.

[44] Ibid., 20.

[45] Ibid., 47-8.

[46] Ibid.

[47] Ibid., 18.

[48] Ibid., 137-8.

[49] Ibid., 143, also §§ 80-81.

[50] Ibid., 19.

[51] Rawls lists the available alternatives on page 124 of *Theory of Justice.*

[52] Rawls makes these two forms of impartiality explicit in his discussion of classical utilitarianism: "In the one case perfect knowledge and sympathetic identification result in a correct estimate of the net sum of satisfaction; in the other, mutual disinterestedness subject to a veil of ignorance leads to the two principles of justice." *Theory of Justice*, §30, 187.

[53] Ibid., 111.

[54] Ibid., 136.

[55] Hume, "On the Original Contract." [1752], in C. W. Hendel, ed . , *David Hume's Political Essays* (NY: The Liberal Arts Press, 1953).

[56] Rawls, *Theory of Justice*, §20, 121.

[57] Ibid., 21.

[58] Ronald Dworkin, "The Original Position," in Norman Daniels, ed., *Reading Rawls* (Stanford, CA: Stanford University Press, 1989 [1975]).

[59] Ibid., 34.

[60] John Rawls, "Kantian Constructivism in Moral Theory," in Stephen Darwall et al., eds., *Moral Discourse and Practice* (NY: Oxford University Press, 1997)

[61] Ibid., 256.

[62] Ibid., 256-7.

[63] Ibid., 257.

CHAPTER VI

LOCKEAN CONTRACT THEORIES

Detached impartiality is used in a different way and for different purposes in another kind of contract theory, which I have termed "Lockean." These theories do not have equality as their primary emphasis, although Locke's own theory takes equality as a starting point; they are concerned explicitly with persons who are motivated to production and achievement wishing to form a society with others whom such persons would naturally associate with. In these cases, the desirable association is with persons who are also high achievers and have some wealth, whether in actual economic terms or in intellectual or even artistic terms. The emphasis is on protection of the property in the self that is either assumed, as in Locke, or derived, as in Gauthier, and thereby of property in an external sense. In the following sections I will examine this kind of theory, which does not rely upon a prior understanding of equality among persons (even if, as in Locke, there is an explicit initial commitment to that equality), to see whether the style of impartiality engaged there might enable us to break free of the closedness in moral community that has resulted thus far from applications of various kinds of impartiality.

I will look at the contract theories of Locke, Robert Nozick, and David Gauthier. It is not my purpose to survey all contemporary variations on contract theory; I have chosen Nozick and Gauthier because they best exemplify the consequences of adopting without question a particular and significant Lockean distinction, that between persons

who are "Industrious and Rational" and those who are "Quarrelsome and Contentious."[1] Locke's theory will show how even an argument built upon an explicit acknowledgement of equality, as in the preceding three theories, can nevertheless take a concept (in Locke's case, rationality) and use it to subvert and eliminate actual equality. Impartiality, in that case, while apparently active in a broad venue to start with, is later narrowed and focused, and used as a tool once again to close a moral community.

Locke

Despite his initial humane commitment to equality, it is actually in Locke that there appears a much more radical constriction and diminution of moral equality, which comes about because of two of the elements of Locke's system: the assumed property in the self, and the explicit restriction of participation in the state of nature and, by implication, later on in civil society, to the "Industrious and Rational."[2]

Locke begins by describing the natural state of humans: it is

> a *State of perfect Freedom* to order their Actions, and dispose of their Possessions, and Persons as they think fit, within the bounds of the Law of Nature, without asking leave, or depending upon the Will of any other Man. A *State* also of *Equality*, wherein all the Power and Jurisdiction is reciprocal, no one having more than another: there being nothing more evident, that that Creatures of the same species and rank promiscuously born to all the same advantages of Nature, and the use of the same faculties, should also be equal one amongst another without Subordination or Subjection ...[3] [emphasis in text]

He quotes Hooker, who notes that this self-evident equality underlies men's " ... Duty, to Love others [as] themselves."[4]

In the next passage, Locke bounds the liberty to dispose of one's own person in the state of nature by the prohibition against destroying oneself; this restriction, because of humankind's natural equality, also prohibits one from destroying another. The complementary affirmative duty to preserve oneself is transferred for the same reason to the affirmative duty

> *to preserve the rest of Mankind,* and may not unless it be to do Justice to an Offender, take away, or impair the life, or what tends to the Preservation of the Life, Liberty, Health, Limb or Goods of another.[5] [emphasis in text]

So far, Locke's understanding of the natural state of humankind is deeply egalitarian, and it continues in this vein through discussion of the state of war and of slavery. Locke's egalitarianism differs from Hobbes' up to this point only in his more optimistic view of human nature as being at least capable of trust within the state of nature –

> For Truth and Keeping of Faith belongs to Men, as Men, and not as Members of Society.[6]

– and in his conviction that humankind are "naturally in that State [induced to seek Communion and Fellowship with others]"[7] and can live together "according to Reason."[8] But, as C.B. MacPherson suggests in a different context, what Locke gives with one hand, he takes away with the other.[9] The elements and implications of Locke's arguments in his discussion of property lead to a severe restriction of his moral community, taking it ultimately far from the equality that he had urged in the earlier passages.

Impartiality, presumably already built into the explicit natural equality of persons in community with one another and responsible for one another's well-being, is called

upon chiefly to regulate the acquisition of, and enduring title to, property and goods in the state of nature. This proceeds by the use of reason, which is the Law of Nature, teaching the preservation of self and others that is a result of persons equally being the property and servants of God, and prohibiting the impairment of others' well-being for the same reason.

Reason, the Law of Nature, was given to humankind for three purposes: to preserve oneself and prevent oneself from self-destruction, to preserve others and refrain from destroying others (save for punishment of transgressions), and to make use of the world given to humankind by God by properly appropriating its goods.

How does one appropriate land and other goods without impairing "what tends to the preservation of another"? God has given the Earth, and "all inferior Creatures ... to all Men, yet every Man has a Property in his own Person."[10] His exclusive right to his own person leads to exclusive possession of his own work and labor, and therefore

> ...whatsoever then he removes out of the State that Nature hath provided, and left it in, he hath mixed his *Labour* with, and joyned to it something that is his own, and thereby makes it his *Property.* ... For this *Labour* being the unquestionable Property of the Labourer, no Man but he can have a right to what that is once joyned to, at least where there is enough, and as good left in common for others.[11]

Appropriation proceeds by carving out from the Common enough goods to sustain oneself and one's family, the logical crux of "Common" being that it exists for appropriation.

"Enough and as good left" is Locke's "proviso" and the venue of the impartiality required by natural equality and the Law of Nature in order properly to assess what is

meant by "as good" and "enough" and "others." Since in the state of nature no one's consent is required for this appropriation save one's own, the granting of such self-consent places the onus of fairness on the appropriator. There is a strict guideline: these things were "richly" given to *enjoy;* this means "as much as any one can make use of to any advantage of life before it spoils;"[12] waste is not permitted by the Voice of Reason. Moreover, land was also richly given, but one may "inclose ... from the Common"[13] only as much as one "...Tills, Plants, Improves, and Cultivates" without "prejudice" to anyone else, if there is "still enough, and as good left."[14] Goods are given for *use,* not merely for accumulation; personal "use" bounds appropriation in the state of nature (at least until the appearance of money, which then gives freedom to accumulate a surplus). Land is the chief good of the Earth; by cultivating land one increases, rather than decreases, the common stock of humankind[15] and therefore follows the Law of Reason, to preserve oneself and others.

However, in this framework of impartiality, the embedded conception of the person comes to light, since it now appears that God gave the world not to "all Men," exactly, but specifically to those who deserve it,

> to the use of the Industrious and Rational, (and *Labour* was to be *his title* to it;) not to the Fancy or Covetousness of the Quarrelsom and Contentious. He that had as good left for his Improvement, as was already taken up, needed not complain, ought not to meddle with what was already improved by another's Labour: If he did, 'tis plain he desired the benefit of another's Pains, which he had no right to, and not the Ground which God had given him in common with others to labour on, and whereof there was as good left, as that already possessed, and more than he knew what to do with, or his Industry could reach to.[16]

This particular set of ideas, that God gave the world to the industrious and rational rather than to the quarrelsome and contentious, initially seems unexceptionable in this formulation, until one sees what Locke has to say about waste and puts it together with the preceding ideas of property deriving from a mixing of one's labor with the goods of the earth, these goods being reserved for the Industrious and the Rational, and what is meant by being "Rational." If one appropriates more than he can use, and the part of what he appropriates therefore goes to waste,

> he offended against the common Law of Nature and was liable to be punished; he invaded his Neighbour's share, for he had no Right, farther than his Use called for any of them, and they might serve to afford him Conveniencies of Life.[17]

Further, labor augments the value of land[18] and makes up

> ...*the far greatest part of the value* of things, we enjoy in this World: And the ground which produces the materials, is scarce to be reckon'd in, as any, or at most, but a very small, part of it; so little, that even amongst us, Land that is left wholly to Nature, that hath no improvement of Pasturage, Tillage, or Planting, is called,as indeed it is,*wast*...[19] [emphasis in text]

The Voice of Reason naturally recommends that the industrious have title to the land, for their labor will add value to the goods of the earth, especially land, whereas "land that is left wholly to Nature ... is *wast*."[20]

Civil government comes into being by "compact and agreement" after the Common has mostly been appropriated or when the use of money has made land scarce; when the proviso can no longer be applied, the burden of impartiality then transfers from

individual interpretation of the proviso to government in the form of impartial legislation and just adjudication to settle disputes. The chief end of civil government is the preservation of property,[21] where by "property" is understood "Life, Liberty, and Estate."[22]

At this point, what Locke means by characterizing persons as either "Industrious and Rational" or "Quarrelsome and Contentious" becomes important, because it is increasingly clear that only the former have title to the goods of the earth in the state of Nature, and by extension, they will be the ones "by compact and agreement" to come together into civil society for the express purpose of protecting their property, including that which they have appropriated.

Persons who are industrious and rational are the ones who are going to appropriate and make good use of the goods of the earth. They are industrious insofar as they mix their labor with these goods in the act of appropriation; they are rational insofar as they use these goods well, without wasting them, and also insofar as they properly interpret the proviso, in this way preserving others' rights to appropriate goods and their ultimate title to those goods (again, until the introduction of money).

But who are the quarrelsome and contentious? They could be thieves, or the idle rich; this would certainly be an intuitive interpretation. But Locke is not as clear about these types of persons as he is about the former. He characterizes them as being prone to "Fancy or Covetousness"[23] and describes them as tending either to complain and meddle concerning what an industrious and rational person has already appropriated, or as sinking into covetousness, desiring "the benefit of another's pains, which he has no right to"[24] rather than appropriating and working some part of the Common themselves – thus, lazy as well.

Because Locke leaves it to our imagination to fill in the class of the Quarrelsome and the Contentious, what he sets up here is an ultimately toxic dichotomy of kinds of persons, one that makes it too easy to account for the many other kinds of persons there may be – the weak, disabled, elderly, or ill, to name several – as being merely Quarrelsome and Contentious. The problem comes about in assessing what a person has done in his life; if it cannot be construed that the person has been "industrious" in the approved sense, i.e. appropriating the goods of the earth and thereby increasing the common stock, it is easy to assume that thereby he has not been fully rational either, since industry and rationality appear to go hand in hand by the Law of Nature, which has given the world for our use. If persons are neither industrious nor thereby fully rational, then they must fall into the only other category of person there is, in this particular universe – the quarrelsome and contentious. In later contract theorists, this dichotomy becomes truly dangerous: for one writer, this portion of humanity is indiscriminately lumped in with "parasites."[25]

Those who have been industrious and rational, who have well used the goods of the earth to increase the common stock, and who have, either by their industry or by the introduction of money, used up all the Common, now come together by compact and agreement to create a civil government which is charged with the protection of property. But "property," in Locke's terms, sometimes means more than just external goods; it includes Life and Liberty as well,[26] and here is where Locke's initial egalitarianism returns, but only in one aspect of government: the Laws, designed for the good of the people are "to ... have one Rule for Rich and Poor, for the Favourite at Court, and the Country Man at Plough."[27] On the other hand, the express or tacit consent of the people is required for making and enforcing law, but for this purpose "the people" are, or ought to be,

represented only by that "part of the People" that is "in proportion to the assistance, which it affords to the publick."[28] So it seems clear that while the laws ought to be designed to protect and dominate everyone equally, including presumably the quarrelsome and contentious, the legislative organ making the laws and the judiciary adjudicating them ought to be comprised only of representatives of the industrious and rational, who have afforded assistance to the public by increasing its common stock. "And hence subduing or cultivating the Earth, and having Dominion, we see are joyned together."[29] Those who have not done so, those whose "property" includes Life and Liberty, but not Estates, are excluded from full participation in crafting and enforcing the laws that will affect them equally with everyone else. This puts the lie to Locke's contention that the introduction of money has not injured anyone:

> ...it is plain, that Men have agreed to disproportionate and unequal Possessions of the Earth, they having by a tacit and voluntary consent found out a way, how a man may *fairly* possess more land than he himself can use the product of, by receiving in exchange for the overplus, Gold and Silver, which may be hoarded up *without injury to any one* ...[30] [emphasis added]

The injury does not comes directly in the nonpossession of land, for one may "tenant" someone else's land, but indirectly in the ultimate disenfranchisement that results from not possessing land or other property. Civil government that comes into being precisely for the purpose of protecting, among other things, estates (including land), is not likely to compact and agree with non-holders of estate in setting up its structures. Impartiality, then, must lead an uneasy and divided life, striving as it were to apply Law equally and impartially to everyone, but within a system that has *prima facie* denied full participation

to those it will affect.

If it were true that the Industrious and Rational, and the Quarrelsome and Contentious, were the only categories of persons that there were, then Locke's egalitarianism could possibly be salvaged, since one could argue that initially, in the state of nature, *everyone* has the right to appropriate and make use of the goods of the earth given by God and everyone is presumed to have the capacity to do so; therefore, it is merely a matter of choice by weak-charactered persons that renders them propertyless in the end. (MacPherson believes that this was a sincere belief of the times.[31]) But if not everyone who is not industrious is thereby not rational, or quarrelsome and contentious, but may in fact be disabled, weak, elderly, or disadvantaged by some other factor outside of the person's control, then Locke's egalitarianism disappears and "preserving others" falls by the wayside to a great extent.

Moreover, there is an even more sinister consequence of Locke's argument, in that it can be made to justify the simple tossing-away of whole groups of people. If the goods of the earth also include human reason, given to everyone in order to allow them to preserve themselves and others, but reason is not made use of (that is, the individual is not industrious and does not appropriate), then, like other unused goods, it lies "wast," which is contrary to that reason itself (it offends the Law of Nature). If those who are not industrious are therefore seen as *intentionally* not rational, it is but a short leap from there to consider such persons as being, literally, *wastrels*, or perhaps even wasted persons, and thereby not worthy of full participation in, or the full protection of, the community.

The system then becomes one in which, regardless of the good and just intentions of the constructors of the system, inevitably the propertied (landed, moneyed, or merchant) class, the class most motivated to bring the system into being, will *de facto*

tend to hold sway over the unpropertied masses, in everything from equality of opportunity to the just adjudication of law. The impartiality that was initially predicated upon the recognition of everyone's equal moral status under God, has given way once again to another type of closed society -- one that is formally open, but in fact severely restricted.[32]

Two more-contemporary contract theorists show the very long life of this division of humans into worthy and unworthy classes (for that is what it amounts to), based upon their ability to, as Locke puts it, "shift for themselves." Robert Nozick and David Gauthier, among others, in responding to Rawls' challenge to Locke's assumption of absolute property in one's own person, carry forward Locke's argument to an extreme that relinquishes even the pretense of considering persons as moral equals (and in fact, argues *against* such consideration as being immoral in its own right). The question that needs to be considered, then, is about the status of the assumed moral link between impartiality and equality. Perhaps that intuitive linkage will be, and ought to be severed; this will certainly give a new conception of impartiality to be mulled over.

Nozick

So far, discussion has centered on how various forms of impartiality – those used in sympathy theories and in contract theories – can end by being exclusionary in their implementation. My "neighbor" turns out to be either someone very similar to myself, or else tyrannical and repressive measures must be instituted in order to make sure that everyone indeed *is* my neighbor.[33] The only possible exception so far has been Hobbes. In Hobbes' system, persons are educated to justice and justice is inclusive. Impartiality in Hobbes is inextricably linked to his premise of equality; if everyone is equal, there is no

venue for institutionalized partiality, and the education to justice will eventuate in personal impartiality as well. With the other theorists, however, the failure of equality as an outcome of various impartiality mechanisms seems to point to some deficiency in those mechanisms.

But in *Anarchy, State, and Utopia*, Robert Nozick asks the counterintuitive question whether equality of this sort is actually desirable in the first place, whether it is even moral. Nozick offers an explicitly Lockean understanding of contract theories, especially Rawls', but without beginning where Hobbes, Locke himself, Rousseau, and Rawls have begun – with the unquestioned premise of equality. He begins instead with another Lockean premise, that of absolute property in the self, and shows how this leads to a specific outcome: a society built on an entitlement principle that is explicitly neither based in equality, nor concerned to establish equality as an end state. One's entitlement to one's holdings is based upon their just acquisition and subsequently their just transfer, and, if necessary, the rectification of injustice in holdings.[34] The question is whether a certain kind of equality is assumed here, and of course it is; but it is the impersonal "equality" of nature, as interpreted by the unargued assumption that all are equally capable of voluntary actions, not the moral equality of rational and self-directed human beings.

Nozick is concerned to show that only a minimal property- and life-protective state (the "night-watchman" state) is morally permissible and that this is the only kind of state that legitimately can arise from a Lockean state of nature, a "non-state situation in which people generally satisfy moral constraints and generally act as they ought."[35] Any more extensive state, one that, for example, affirmatively seeks to improve the position of a disadvantaged class or group, cannot be morally justified (provided the disadvantage is

not a result of prior injustice), since it will inevitably involve or even require violations of other persons' entitlement rights somewhere along the line in order to effect the remedy.

A significant portion of Nozick's arguments are presented as a critique of Rawls' theory, which Nozick uses as a foil to sharpen his own opposing position. The fact that Nozick has worked out an *opposing position* to Rawls' and not merely a critique of Rawls' techniques, arguments, and so on, deserves some emphasis. Nozick takes the uneasy results of Locke's contract theory but does not see them as disadvantaging it; on the contrary he develops those results into what he claims is not only a fully moral outcome, but in fact is the *only* moral outcome possible. He provides thereby a serious challenge to the intuitive idea that impartiality somehow depends upon or requires or embodies equality, and in fact to the idea that a state of affairs in which a recognition of equality is affirmatively implemented is itself desirable or even moral.

In order to make his case that absolute property in the self and all that legitimately flows from that trumps equality, Nozick develops three ideas: the distinction between historical and structural principles of distributive justice; the idea of voluntary choice as the bearer of all rights concerning justly acquired holdings, including rights of transfer; and so-called "invisible-hand" mechanisms of explanations of outcomes.

"A distribution," says Nozick, "is just if everyone is entitled to the holdings they possess under the distribution."[36] In determining the justice of holdings, "subjunctives" do not apply;

> ...that from a just situation a situation *could* have arisen via justice-preserving means does not suffice to show its justice ... the fact that a thief's victims *could* have presented him with gifts does not entitle the thief to his ... gains. Justice in holdings is historical; it depends on what actually has happened.[37] [emphasis in

text]

The entitlement theory of justice is historical in this way; Nozick contrasts this with unhistorical structural end-state (or a subcategory, "patterned") principles of distribution. These hold that "...the justice of a distribution is determined by how things are distributed (who has what) as judged by some structural principle(s) of just distribution."[38]

> Two distributions are structurally identical if they present the same profile, but perhaps have different persons occupying the particular slots. My having ten and your having five, and my having five and your having ten are structurally identical distributions.[39]

> ... *historical principles* of justice hold that past circumstances or actions of people can create differential entitlements or differential deserts to things. An injustice can be worked by moving from one distribution to another structurally identical one, for the second, in profile the same, may violate people's entitlements or deserts; it may not fit the actual history.[40] [emphasis in text]

A distribution based on entitlement operates entirely according to the voluntary choices persons make with regard to their (justly acquired) holdings and thus preserves justice. These include persons receiving

> ...their marginal products, others win[ning] at gambling, others receiv[ing] a share of their mate's income, others receiv[ing] gifts from foundations, others receiv[ing] interest on loans, others receiv[ing] gifts from admirers, others receiv[ing] returns on investment ...[41]

Although "heavy strands of patterns"[42] may permeate the outcomes, over large numbers of persons in a society, of these voluntary choices, the "principle of entitlement ... is *not* patterned."[43]

This second idea, the idea of voluntary choice as an outgrowth of absolute property in the self, including one's own actions,[44] is made to bear the burden of the rights persons have in their holdings. All of the transfers above are based on the voluntary choices persons make in disposing of their justly-held property. Even foolish choices are morally legitimate because of the person's just entitlement voluntarily to do what he pleases with his holdings.

Nozick realizes that in the area of distributive justice there may be controversy concerning the notion of "voluntary choice." The specific interpretation of voluntary choice that he responds to asks about the true idea of voluntary choice in a situation where there are limited options and each is worse than the (undesirable) choice that is actually made. Nozick's response is Hobbesian:

> Whether a person's actions are voluntary depends on what it is that limits his alternatives. If facts of nature do so, the actions are voluntary. ... Other people's actions place limits on one's available opportunities. Whether this makes one's resulting action non-voluntary depends upon whether these others had the right to act as they did.[45]
>
> A person's choice among differing degrees of unpalatable alternatives is not rendered nonvoluntary by the fact that others voluntarily chose and acted within their rights in a way that did not provide him with a more palatable alternative.[46]

The third idea Nozick develops is the "invisible-hand" mechanism as explanation of some outcomes that, on the surface, may appear to have emerged by some different explanation. "Invisible-hand explanations [have] a certain lovely quality," says Nozick;

> ...[t]hey show how some overall pattern or design, which one would have thought had to be produced by an individual's or group's successful attempt to realize the pattern, instead was produced and maintained by a process that in no way had the overall pattern or design 'in mind.'[47]

The "specially satisfying quality" of these explanations is connected with the "notion of fundamental explanation":

> Fundamental explanations of a realm are explanations of the realm in other terms; they make no use of any of the notions of the realm ... Invisible-hand explanations minimize the use of notions constituting the phenomena to be explained; in contrast to the straightforward explanations, they don't explain complicated patterns by including the full-blown pattern-notions as objects of people's desires or beliefs. Invisible-hand explanations of phenomena thus yield greater understanding than do explanations of them as brought about by design as the object of people's intentions. It therefore is no surprise that they are more satisfying.[48]

(These are contrasted with their mirror image, "hidden-hand" explanations, the soul of conspiracy theories, which connect disconnected facts according to the pattern that the connector generally wishes to realize.)

As a foundation for his later arguments opposing end-state distributions, Rawls' in particular, Nozick develops the explanation of the emergence of the minimal state from

the state of nature as an invisible-hand explanation, to show that no rights that are held by the state were not held previously by individuals in the state of nature. The two indispensable conditions for being a state are the monopoly over the use of force (this single condition alone denoting an "ultra-minimal" state) and the fact that the state protects everyone, even "non-clients" (both conditions together denoting the "minimal" state). He uses the device of persons in the state of nature ultimately developing so-called "protective associations" to deal with those persons in the state of nature who choose *not* to "satisfy moral constraints" and do *not* "generally behave as they ought." Ultimately, a single protective association achieves dominance over a territory; this is the precursor to the minimal state, the ultraminimal state which achieves monopoly over the use of force. At that point, the transformation from the ultraminimal to the minimal state is a matter of moral obligation; the dominant association must offer compensation to those non-clients of the association whom it prohibits from acting with force on their own behalf (providing their own protection) due to their tendency to judge situations or persons incorrectly or engage in risky behaviors endangering their own clients; this compensation is the agreement to protect them against their own clients regardless of whether the non-clients have paid for such protection. Although this looks like "redistribution," in fact it is an invisible-hand explanation justified by the moral principle of compensation, rather than by the end-state principle of everyone's needing to be protected.[49]

It is at the point of voluntary choice that an immovable object (the assumption of a person's absolute property in the self) meets an irresistible force (the incontrovertible truth, brought out by Rawls, that differences in one's starting place in life can generate vastly different possibilities for that life, that any starting place at all is completely arbitrary from a moral point of view, and last, that it can have nothing at all to do with

one's voluntary choice). This point of extreme dynamic tension brings into question all the concepts upon which theorists rely so heavily: entitlement, equality, property in the self (what is the true nature of that self and how is it begotten?), voluntary choice, just processes leading necessarily to just outcomes, justice itself.

Having established his foundations -- the legitimate emergence of a minimal state and *only* a minimal state from a Lockean state of nature and the illegitimacy of end-state distribution principles – Nozick then turns to Rawls' *A Theory of Justice*. Nozick asks the question that he wants Rawls to answer: why does social cooperation necessarily introduce "complications" to raw entitlement, that is, why must it change the *modus vivendi* of cooperating persons from the historical principle of entitlement to the structural principle of patterning (exemplified in this case by Rawls' difference principle)? "Why," he asks, "does social cooperation *create* the problem of distributive justice?"[50] In relation specifically to Rawls, he asks how the difference principle may at all be justified.

After several unsuccessful attempts to develop examples of typical kinds of social cooperation and thereby to find some way to justify, within those models, the necessity for moving away from entitlement as a template for justice, he zeroes in on what he considers to be the main culprit in Rawls' theory – the construction of the original position. Given the task at hand, Nozick says, the structure of the original position simply guarantees that end-state, rather than historical, principles *must* be selected.

It is the "social pie" that must be somehow divided; the question is how? Suppose the pie to have materialized out of nowhere; an equal distribution might be agreed upon as a "focal point solution,"[51] but then someone might realize that "the size of the pie [isn't] fixed and ... that pursuing an equal distribution ... would lead to a smaller total pie."[52] In

that circumstance people might agree to "an unequal distribution which raised the least share."[53] But who would increase the size of the pie to make that possible and thus deserve incentives for so doing? The answer to that question, says Nozick, should "reveal something about differential claims on parts of the pie."[54]

But the pie did *not* fall "from heaven like manna,"[55] so Nozick offers another analogy. A group of students is asked to decide a distribution of their grades, and they are to assign a particular grade to each identifiable person. After a couple of attempts at actual assignment – assigning the same grade to everyone, rejecting the instructor's actual distribution – they understand that they are to agree to a *principle* to govern the distribution of the grades. Not surprisingly, self-interest dictates the rejection of any principle resembling entitlement, since those who are likely to do poorly will be unhappy with this solution. In this situation, they might well agree to an end-state principle which maximizes the lowest grades. A significant observation is that they could not choose an entitlement distribution on the grounds of *fairness*, because moral concepts are not permitted to influence choice in the original position, so self-interest behind the veil of ignorance must be the ruling justification; this not only precludes the choice of a historical principle but actually *requires* the choice of an end-state principle. This is Nozick's point:

> The nature of the decision problem facing persons deciding upon principles in an original position behind a veil of ignorance limits them to end-state principles of distribution.[56] ... The whole procedure of persons choosing principles in Rawls' original position presupposes that no historical-entitlement conception of justice is correct.[57]

Last, Nozick searches for some features of Rawls'construction in virtue of which it can not (ever) yield a historical principle. Surprisingly, Nozick fails to find such features and acknowledges his failure, asserting that his critique up to this moment is not yet deep enough; self-interest and the prohibition against moral terms is not sufficient. If he were to rest here, he would be saying nothing more than "that the construction is incapable in principle of yielding any conception other than the one it actually yields. ... it seems clear," he continues, that the

> criticism goes deeper than this ... but it is difficult to formulate the requisite criterion of depth. ... [The] root idea underlying the veil of ignorance ... is to prevent someone from tailoring principles to his own advantage. ... But not only does the veil of ignorance do this; it ensures that no shadow of entitlement considerations will enter the rational calculations of ignorant, nonmoral individuals constrained to decide in a situation reflecting some formal conditions of morality.[58]

It is irresistible to wonder, at this point, whether maybe that is precisely the point; that as the original position is in the way of being a metaphor for a pre-birth condition, where morality may not have much of a function but where self-interest surely must (given the arbitrary lottery of original endowments), perhaps the original position not only cannot but *ought not* to yield the entitlement principle, which will inevitably favor the already-favored. But Nozick is prepared for objections of this sort; in an astonishing footnote he writes,

> Someone might think entitlement principles count as specially tailored in a morally objectionable way, and so he might reject my claim that the veil of ignorance accomplishes more than its stated purpose. Since to specially tailor

> principles is to tailor them unfairly for one's own advantage, and since the question of the fairness of the entitlement principle is precisely the issue, it is difficult to decide which begs the question: my criticism of the strength of the veil of ignorance, or the defense against this criticism which I imagine in this note.[59]

Nozick creates a select, narrow, and explicitly Lockean society, based not in the equality of all persons, Locke's opening premise, but in the absolute property in one's own person, a subsequent Lockean premise which unavoidably vitiates original equality. The Natural Law, Reason, tells us we must obtain – "appropriate" – the goods of the earth given to us by God, by mixing our labor with those goods, our labor being one of the aspects of our absolute property in our own person. Nozick takes "just holdings" as his starting point and derives a society, from the Lockean state of nature, that is built solely on the entitlement to those goods we have justly come by. In Locke, before the introduction of money, impartiality was needed to understand the proviso correctly: how does one appropriate correctly, to leave enough and as good for others? In Nozick, impartiality seems almost not to be needed on a conscious level at all for that purpose; it inhabits, instead, the "invisible hand" that impersonally transforms members' individual actions of a certain sort into a state of affairs of a certain sort, a state that may not have been overtly intended as such but is in place nevertheless. If those individual actions are just, says Nozick, the end result must also be just: "[w]hatever arises from a just situation by just steps is itself just," by analogy with truth-preserving inferential transformations.[60] So at some point, individuals in Nozick's society may relinquish the burden of impartiality altogether, leaving it to the formal processes required for adjudicating between entitlements. But its severance from equality is complete – the invisible hand, by moving justly, has transformed the just, voluntary actions of

individuals into a Nozickean entitlement-based state, one in which it may be regretted that there are poor, dispossessed, disabled, and otherwise disadvantaged persons, but in which it is not at all morally required for anyone or for the state to do anything about it. The invisible hand has taken care of it already, in its godlike way, the same way that sometimes causes persons to scratch their heads in perplexity at some state of affairs and then shrug with the acknowledgement that God's ways are mysterious indeed.

The movements of the invisible hand may, however, be observed, and perhaps seeing it in action might give an answer to the issue of whether its severance from equality is morally acceptable. Nozick offers a counterexample to a point that Bernard Williams develops. Williams makes the case for a poor individual's need for medical care being a sufficient warrant for providing that care:

> ...the proper ground of distribution of medical care is ill health ... [when] we apply the notions of equality and inequality ... in connection with the inequality between the rich ill and the poor ill, since we have straightforwardly the situation of those whose needs are the same not receiving the same treatment, though the needs are the ground of the treatment. ... [This is] a situation insufficiently controlled by ... reason itself.[61]

Nozick here confronts directly what he obviously considers the sacred cow of theories of justice: "It cannot merely be *assumed* that equality must be built into any theory of justice."[62] He takes back a great deal of this apparent boldness with his next statement, however: "There is a surprising dearth of arguments for equality capable of coming to grips with the considerations that underlie a nonglobal and nonpatterned conception of justice in holdings"[63] – the point being that the morality of a "nonglobal and

nonpatterned conception of justice in holdings" is precisely what is at issue, and precisely because such conceptions tend to favor and foster inequality.

But his argument against equality is weak here, relying too heavily on sarcasm, a dubious analogy, and an ultimately question-begging insistence on the overriding morality of entitlement theory. He responds to Williams' point about need being the sufficient (as well as necessary) criterion for receiving medical treatment by making receiving medical care analogous to receiving a haircut. In both cases, Nozick says, Williams (and by extension everyone else who holds social-justice views tied to equality) ignores the fact that these services are provided by the *actions* of *persons* – medical care by someone's doing the activity of doctoring and haircuts by someone's doing the activity of barbering. Because these actions come "already tied to people who have entitlements over them"[64] it is immoral to require that any portion of either one's activities be simply allocated to others on the grounds of others' needs. And if society is to provide the medical treatment and pay for it, ought it do the same for haircuts as well? Needless to say, the analogy of medical treatment with barbering is simply coarse; presumably no one ever died from not receiving a haircut in a timely manner due to lack of funds, and inequality of moral status in barbering results only in aesthetic offense and not in debilitation, illness, and death. Moreover, the assumption of full entitlement to all one's voluntary actions is exactly what is at issue.

The key problem for Nozick (and for the rest of us who were genuinely hoping for a powerful argument, as powerful as Rawls' in favor of equality, *against* the assumption of equality, so we could finally have these two ideas confront each other in detail), is that he does not ever confront effectively the issue of the lottery of initial endowments – the grounds upon which Rawls urges us to understand that we ought to

agree to share each other's fate (because we do, in fact, already share each other's fate, pre-birth; it is just that some of us get lucky and some don't). Nozick quotes carefully, selectively, and legalistically from Rawls; he seems particularly exercised by Rawls' (perhaps unfortunate) use of the phrase "nullifying the accidents of natural endowments and the contingencies of social circumstance,"[65] interpreting it out of context to mean that Rawls has a "negative reflective evaluation of allowing shares in holdings to be affected by natural assets"[66] – overlooking the point of the difference principle, which is that in Rawls' system shares *will* be affected by natural assets in any case, and that for Rawls, this is morally permissible and even desirable, under the provisions of the principle.

Similarly, in another passage Nozick asks,

> Why shouldn't holdings partially depend upon natural endowments? ... Rawls' reply is that these natural endowments and assets, being undeserved, are 'arbitrary from a moral point of view.'[67]

Again, Rawls' noting that these assets are arbitrary from a moral point of view does not mean that he is urging that such persons are not entitled to their differential rewards and that these holdings should *not* "partially depend upon natural endowments"; only that in the interest of making a distribution "just and not merely ... efficient"[68] these rewards ought not to be absolute but subject to the conditions of the difference principle. This is not to argue that the difference principle is or is not a good thing, but to show that Nozick did not respond to it effectively enough to be persuasive. The charge of question-begging that he levels against his own self is a valid one.

It is irresistible to close with a quip from Brian Barry concerning Nozick's ideas.

In his discussion of coordination in *Justice as Impartiality*, Barry places Nozick at one extreme of a coordination continuum on the grounds that very little coordination would be necessary in a minimal state. He lists three disadvantages of Nozick's system which might lead to its *not* being chosen from some differently characterized original position.[69] These are: the increased probability of destitution for any member of the society ("Nobody can be sure of continuing to be physically and mentally capable..."); the tenuousness and inevitably poor quality of public services ("...cooperation on collective projects of mutual advantage ... cannot be mandated"[70]); and probably rampant discrimination, or in Barry's terms, a failure of first-order impartiality ("[i]t would minimize the role of public institutions with their function of applying without fear or favor a uniform set of rules to everybody; ... no firm...or other body that was privately owned could be held accountable for its decisions, however discriminatory."[71]). He closes by praising Rawls for

> recognizing explicitly that societies have patterns of inequality that persist over time and systematic ways of allocating people to positions within their hierarchies of power, status, and money. It is depressing evidence of the social-scientific illiteracy of so many philosophers that someone like Nozick, who is in these terms the equivalent of a pre-Copernican astronomer, should ever have been taken seriously.[72]

Nozick has confronted the question of equality openly, acknowledging that there is a presumption favorable to equality in theories of justice but also that there are few arguments to support such a presumption. He concludes that justice does not readily support any concept of equality but that of humans being capable of voluntary actions; these in turn naturally spawn inequalities in distributions of all sorts tied to these

voluntary actions, which by virtue of their historicalness (or organicity) are thereby justifiable. This is his "entitlement theory," summarized as "From each as they choose, to each as they are chosen."[73] Although he mentions rectification of injustice in holdings in passing, he asserts that he "do[es] not know of a thorough or theoretically sophisticated treatment of such issues"[74] and moves on, although it could be argued that injustice in holdings in this sense is at least part of what is at issue, in relation to the initial, arbitrary lottery of endowments.

David Gauthier's theory of morality goes even farther toward inequality than Nozick's. His presumption of inequality is justified by a premise and a process: an idiosyncratic interpretation of Locke's proviso facilitated by institutionalizing the effects of the process of survival of the fittest.

Gauthier

In his *Morals by Agreement*, David Gauthier quotes, twice, from a "Locke MS" that is itself quoted elsewhere: "an Hobbist ... will not easily admit a great many plain duties of morality."[75] Leaving aside a tendency to distrust that the ellipsis in fact preserves the sense of the original quotation (and that the quotation itself was not taken out of a qualifying context, as quotations from Rawls by his opponents tend to be), the statement is ironic. Hobbes' system, although based drearily in a mechanistic, atomistic view of human beings, nevertheless ends by structuring, through its education to justice, a society that is humane and that can genuinely afford respect to *all* persons; while the selective use of Locke's premises have led, in their successive incarnations, to increasingly cold, narrow, Darwinian, exclusive societies in which entire classes of persons are excluded *ab initio* by those very premises – "gated communities" writ large.

In Gauthier's market society, a recent Lockean incarnation, contract theory becomes the last refuge of a scoundrel. "Impartiality," which in Gauthier might be better characterized as "impersonality," is made, via the offices of Locke's proviso (as revised by Gauthier), to be his market society's bouncer, screening out "animals, the unborn, the congenitally handicapped and defective"[76] from the conviviality and life support of a club structured on mutual advantage, and relegating them to the street. Since they are not made worse off by their exclusion than they would be had they not applied for admission – they were simply not allowed in, not expelled once they were in – such exclusion is, by Gauthier's interpretation of Locke's proviso, moral, and no further notice need be taken. In Gauthier's theory,[77] the ideal of fraternity is not only not realized (except among the narrowly select band permitted membership in the society), it is seen as positively immoral, unless it is an association undertaken voluntarily for mutual advantage. One's fellow human beings are not only *not* presumptively one's brothers, but outside of the society they are by definition potential "parasites" and "free riders." The smallest structural diminution of one's entitlement in order to provide aid or assistance to the excluded is seen as those persons "taking advantage."

It is odd that Locke's proviso is made to carry so much of the moral weight of this particular theory, since Locke's own words in the statement containing the proviso do not support only the narrow reading that Gauthier has given them. Further, it is to be remembered that Locke himself began with a premise of equality and derived from that premise a duty to preserve others by virtue of the Law of Nature, Reason:

> For Men being all the Workmanship of one Omnipotent, and infinitely wise Maker. ...[a]nd being furnished with like Faculties, sharing all in one Community

> of Nature, there cannot be supposed any such *Subordination* among us, that may Authorize us to destroy one another, as if we were made for one anothers uses, as the inferior ranks of Creatures are for ours. Every one as he is *bound to preserve himself*, and not to quit his Station Wilfully; so by the like reason when his own Preservation comes not in competition, ought he, as much as he can, *to preserve the rest of Mankind*, and may not unless it be to do Justice on an Offender, take away, or impair the life, or what tends to the Preservation of the Life, Liberty, Health, Limb or Goods of another.[78] [emphasis in text]

While it is clear that exclusionary premises may be extracted from this passage (for example, we are not "made for one anothers uses," one may not "take away ... or impair ... what tends to the Preservation of Life" including the "Goods of another"), premises which are fully exploited in Gauthier to define the idea of affirmative duties to the disadvantaged to be so impairing,[79] nonetheless it must be clear to the most casual reader that both the letter and the spirit of Locke's passage stress that *all* share "in one Community of Nature" and that we are bound "to preserve *the rest of Mankind*," not just the persons we approve of. "Preservation," in Locke's qualification "when his own Preservation comes not in competition," can only in the bandit Procrustes' iron bed be interpreted as being "in competition" when affirmative duties to the disadvantaged are at issue; yet this is exactly as Gauthier sees it. The above-mentioned "animals, the unborn, the congenitally handicapped and defective *fall beyond the pale of a morality tied to mutuality*"[80] [emphasis added] and one's "preservation being in competition" gets significantly downsized and added to the proviso as *"worsening one's own position."*[81] Gauthier ties it up neatly (and begs the important question of "plain duties") when he says

> The crucial distinction that we must establish is between worsening someone's situation and failing to better it, since the proviso prohibits only the former, not the latter.[82]

The "crucial distinction," however, is not in fact this one; it is the distinction between what Locke actually said and intended in his proviso, its letter and spirit, and how Gauthier reinterprets Nozick's interpretation of it:

> For this Labour being the unquestionable Property of the Labourer, no Man but he can have a right to what that is once joyned to, at least where there is enough, and as good left in common for others. (Locke[83])

> Locke's proviso that there be 'enough and as good left in common for others' [sect. 27] is meant to ensure that the situation of others is not worsened. (Nozick[84])

> Following Locke who allows one's own preservation to take justifiable precedence over that of others in one's deliberations, we modify Nozick's interpretation of the proviso, so that it prohibits worsening the situation of others except where this is necessary to avoid worsening one's own position. (Gauthier[85])

The important thing to notice in Locke's passage is two words that add emphasis to the way he intended the proviso, and which are utterly ignored by Gauthier:

> ...no Man but he can have a right to what that is once joyned to, *at least* where there is enough, and as good left in common for others. [emphasis added]

By his careful structuring of this sentence, with its deliberate use of the proviso clause as a *qualification* of the first part of the sentence, and by emphasizing this placement with "at least," Locke has left the door open for others indeed to have some sort of right to what someone's labor has been joined to, in the event that there is *not* enough and as good left in common for others. While Gauthier (and Nozick before him) has interpreted the entirety of this passage to support Locke's idea of absolute property in the person, including his labor and therefore by extension his property, neither has seen that in fact this passage also provides some warrant for what is anathema to both of them – the annexing of some of the property in someone's person to assist others for whom there is *not* enough and as good left – in fact, an avenue for the duty to preserve others that emerges from the Law of Nature. From this open door, it is not inconceivable that some sort of Rawlsian argument could also possibly be constructed, supporting, say, the idea of a person's natural talents and skills being considered as a common asset, in the benefits of the distribution of which the community may rightfully share.[85] The point here is not whether this would be a strong argument, or whether such an attitude toward natural talents is or isn't a good idea or a moral idea. The point is that under a more straightforward and honest reading of Locke, such an idea would not be inconsistent with a morality built on the proviso, as Gauthier would have us believe.

Gauthier's addition "except when this is necessary to avoid worsening one's own position" is not a reinterpretation of the proviso passage as much as it is an importation of his Procrustean "preservation" interpretation into a passage where it doesn't belong, thus assuring himself of a foundation for his mutual-advantage society that appears to have Locke's posthumous imprimatur.

Gauthier's theory depends in its entirety upon this reading of the proviso. He

begins by declaring his interest to be the development of a theory of morals that shows morality as a "subset of rational principles for choice."[86] His primary concern is the problem of subsequent individual compliance with principles of justice chosen to structure a society from an initial bargaining situation; in this he distinguishes his theory from both Rawls' and Harsanyi's, who he claims do not focus on individual rational choice but on the general structural choices made for the society. But Gauthier wishes to claim that "there are situations in which an individual must choose morally in order to choose rationally."[87] He is concerned to show that his theory must

> generate, strictly as rational principles for choice, and so without introducing prior moral assumptions, constraints on the pursuit of individual interest or advantage that, being impartial, satisfy the traditional understanding of morality.[88]

Here Gauthier makes a significant choice between two possible conceptions of rationality. The "maximizing" conception, his choice, assumes that a rational person will continue to seek the satisfaction of his own interests whether or not other persons' interests are also involved. This is contrasted with the "universalistic" conception of rationality, which says that "what makes it rational to satisfy an interest does not depend on whose interest it is."[89]

The two conceptions have differing requirements of justification for the implementation of "impartial constraints on the pursuit of individual interest" – the heart of morality, according to Gauthier.[90] This justification is relatively unproblematic for proponents of the universalistic conception, since

> the rational requirement that all interests be satisfied to the fullest extent

possible directly constrains each person in the pursuit of her own interests.[91]

It is more complicated for proponents of the maximizing conception, and it is this justification that Gauthier develops in his theory.

He begins from an initial presumption against morality and develops a "contractarian rationale for distinguishing what one may and may not do."[92] Although the contractarian agreement is hypothetical, the parties to the agreement are "real, determinate individuals" who "recognize a place for mutual constraint, and so for a moral dimension in their affairs." Morality "emerges quite simply from the application of the maximizing conception of rationality to certain structures of interaction." The problem is, of course, the "step from hypothetical agreement to actual moral constraint" in individual interactions.[93]

Gauthier acquiesces in the contractarian perspective offered by Rawls: a "'cooperative venture for mutual advantage' among persons 'conceived as not taking an interest in one another's interests.'"[94] Mutual advantage is a necessary, but not yet sufficient condition for contractarian agreement; Gauthier intends to develop a set of conditions for sufficiency as well. Sufficiency will ensure compliance. The theory aims to generate a society of "constrained maximizers" who are "disposed to comply with mutually advantageous moral constraints, provided [they] expect similar compliance from others."[95]

The role of the proviso enters the theory in constraining the initial bargaining position by "prohibit[ing the] bettering [of] one's position through interaction worsening the position of another"[96] – Nozick's interpretation; the proviso is a condition of rational agreement.

However, the controversial primary assumption that drives Gauthier's (and, he implies, any contractarian) theory is that "morality requires a context of mutual benefit;" therefore,

> only beings whose physical and mental capacities are either roughly equal or mutually complementary can expect to find cooperation beneficial to all. Humans benefit from their interaction with horses, but they do not co-operate with horses and may not benefit them.[97]

Continuing the cautionary note he sounds here, he acknowledges that contemporary Western society has succeeded in the cooperative venture beyond previous imaginings, having "discovered how to harness the efforts of the individual, working for his own good, in the cause of ever-increasing mutual benefit."[98] But partly because of this, a deep flaw in the contract has now been exposed:

> From a technology that made it possible for an ever-increasing proportion of persons to increase the average level of well-being, our society is passing to a technology, best exemplified by developments in medicine, that make possible an ever-increasing transfer of benefits to persons who decrease that average. Such persons are not party to the moral relationships grounded by contractarian theory.[99]

The question irresistibly arises here of what we are actually to *do* with "such persons." The assumption must be made that Gauthier does not intend for us simply to have them killed outright, especially since this would be bettering one's situation by worsening that of another's, prohibited by both Nozick's and Gauthier's interpretations

of the proviso. And these persons are already members of the society; they slipped past the bouncer, so to speak, so expelling them would also worsen their situation while bettering ours. Yet his position is clear – these persons disadvantage everyone else by decreasing average well-being through a transfer of benefits and are therefore outside the scope of morality as defined by his theory. How can such a counterintuitive, shocking, and cold attitude develop from Locke's proviso, based as it is in equality before God and a duty to preserve Mankind? The answer is that it requires some strenuous reinterpretation of Locke, which Gauthier takes care of in the course of developing conditions for the initial bargaining situation.

The proviso constrains the initial bargaining situation by setting out conditions on the legitimacy of the endowments that are brought to the table by each party.

> The initial bargaining position must be non-coercive ... each individual's endowment ... must be considered to have been initially acquired by him without taking advantage of any other person – or, more precisely, of any other cooperator.[100]

Otherwise, if endowments have been arrived at by previous coercion, there is no incentive to comply after the agreement. Legitimizing these initial endowments requires an intricate process of justification that begins with revising the proviso; the revised proviso is then used step-wise to convert a pure, Hobbesian state of nature to an appropriate initial bargaining position. The proviso also defines the nature of the person, at least a "suitable [non-literal] reading"[101] of it does. Gauthier revises the proviso by adding the clause "except where this is necessary to avoid worsening one's own position."[102] An analysis of what is meant by "worsening" and "bettering" someone's situation leads to the derivation

of the Lockean property in the self, which Gauthier says Locke simply assumes. The property in the self is derived directly from the nature of the human person, which emerges through one's progressive understanding of how one's actions can better or worsen the situation of others and how others' actions can similarly affect one's own situation.[103]

This derivation of rights to one's own body, and therefore one's own labor, is the first step in the conversion of the Hobbesian state of nature to the initial bargaining position. Three more steps are required. The second step derives a right in one's products from the right to one's body and its labor;[104] from these the third step derives the right of full compensation for displaced costs generated by someone else's activities;[105] and from these the fourth step introduces exclusive rights to land and other goods.[106] This fourth step secures the morally-free market because, thanks to the proviso, property is now possessed exclusively and this is mutually beneficial, since everyone's needs can now be met through market exchange. This provides a basis for the emergence from the state of nature into a cooperative society for mutual advantage by securing rights to the endowments brought to the bargaining table.

However, "different persons will of course benefit differentially ... advantage is not taken, but equality is not assured.[107] The next task, then, is to show that the "inequality allowed by the proviso is no indication of partiality." The key question is whether the proviso fails to be impartial "because it is based on considerations about bettering and worsening."[108] Impartiality does not necessarily translate into "equalizing" or "meeting needs."

Here Gauthier confronts the most controversial element of the controversial difference principle – the idea that because natural talents and endowments and the social

circumstances one finds oneself in are morally arbitrary in their distribution, they not only ought to be compensated for in the structure of a just society by the difference principle, but in fact the core idea is that because of the moral arbitrariness of their distribution (the initial "lottery" of endowments[109]), they ought to be considered a common asset of the community in order to level or equalize the playing field. Gauthier confronts this argument with his revised proviso, and it is now clear that because of the neat and definitive way this version of the proviso disposes of the equalizing problem, by deriving unconditional and exclusive entitlement to natural endowments, community rights in individual talents must have been his target from the beginning.

Gauthier's argument can be pulled and simplified from his contrived and lengthy example of sixteen "Robinson Crusoes," each of whom is "endowed" with a certain mix of personal skills, attitudes, and talents, and relegated to his own desert island, each of which in its turn has certain fixed characteristics of abundance or scarcity. Some are lucky either in their natural endowments or in those of their island, or in both. Some are unlucky in either or both. After tracking all the possible circumstances under which each Robinson Crusoe could come to know of others' good or ill fortune, and the various interactions they might voluntarily or coercively engage in, Gauthier comes out with the following conclusions:

1. It is true, as Rawls says, that no one "deserves" either their natural talents or the circumstances one finds oneself in. But nothing follows from this.

2. The reason nothing follows from this is that "[o]ne's natural capacities determine what one gets, given one's circumstances, in a condition of solitude. ... Why

should they not determine, or contribute to determining, what one gets in society?"[110]

3. The reason one's natural capacities ought definitively to determine what one gets in society is because the (revised) proviso forbids "worsening the situation of others except where this is necessary to avoid worsening one's own position."[111]

4. From 1 – 3 several conclusions emerge for individuals in a society:

(a) I am not worsening your situation by not sharing my endowments with you, since in solitude you would be in exactly the same situation as you are now. Therefore I am not violating the proviso.

(b) However, if you insist upon my bettering your situation by sharing my endowments with you, then *you* are violating the proviso by making my situation worse. You are taking advantage of me, coercing me.

(c) Therefore I am morally entitled to ignore your needs, no matter how dire they are, even if they are life-threatening, if my helping you would worsen my situation even slightly and I don't wish to have my situation worsened, even slightly. "The rich man may feast upon caviar and champagne while the poor woman starves at his gate. And she may not even take the crumbs from his table, if that would deprive him of his pleasure in feeding them to his birds."[112]

However, the revised proviso contains its own contradiction. By the terms of Gauthier's revision, it appears that you may indeed worsen my situation if it is to avoid worsening yours. Borrowing from Nozick, if I have ten and you have five, and for some reason you stand to lose one unless I give you two, then it appears that by the terms of the revision I ought, morally, to give you the two. This is what comes of attenuating

"one's preservation being in competition" to merely "worsening one's situation" – your losing one will not put your preservation in competition, but it will certainly worsen your situation, and therefore you seem to have acquired a right against me to worsen mine (because clearly my preservation would not be in competition either should you do so).

In Gauthier's theory, the impartiality of detachment appears in its worst possible light, licensing the complete avoidance of "plain duties" – affirmative duties to those less fortunate than oneself. His theory relies upon an idiosyncratic reinterpretation of only one of Locke's premises, the proviso, and also upon the process of survival of the fittest, an evolutionary process. In trying to make these two disparate elements work in tandem in his argument, Gauthier highlights a major problem for those who must select a conception of rationality that argues either for or against equality: the uneasy positioning of humankind in relation to the rest of the world's creatures.

Humankind is unquestionably a part of nature; the forces of nature, such as laws of physics, chemistry, and biology, operate as impersonally upon human beings as they do upon other animals. Evolutionary processes operate impersonally as well on all animals, including human beings. But humankind is presumed to be distinct from other animals in the possession of rationality. Human rationality enables scientific investigation, literary and artistic pursuits, philosophical speculation, and morality. Animals cannot engage in these pursuits. This suggests that the possession of rationality creates a categorical distinction between humans and other animals.

Nozick has complained that there was a presumption of equality in moral theories without very many arguments to support them. Most of the arguments for equality, for example Hobbes', rely upon this very categorical distinction between humans and other animals, and conclude that this is what renders humans equal to one another *as humans*,

and therefore, since morality is a uniquely human pursuit, morally equal as well. This is the presumption of equality, hardly needing any argument to support it because it seems so obvious that it is often almost intuitively felt.

But Gauthier and Nozick go farther; they imply that there are categorical distinctions based on rationality *among* human beings, a more controversial claim because it immediately calls into question the presumption of human, and therefore moral, equality. Gauthier uses Nozick's Robinson Crusoe example to demonstrate that there are those whose rationality is such that, combined with the requisite physical capacities, and given specific, generally favorable circumstances, it would permit their survival in solitude on a desert island. There are correspondingly those whose physical and mental capacities would not permit such survival. We are then urged to look upon persons as we find them in society, and imagine their capacity for survival on a solitary desert island. From the conclusions we come to, we are to categorize them as capable or not capable of taking care of themselves, and thereby, via the invisible hand, of the rest of us, in a society built upon conceptions of mutual benefit. We are asked, in short, to categorize them according to differential rationality. Those whom we deem not capable of such survival are thereby thrust outside the bounds of the moral system; they are "beyond the pale," "not party to the moral relationships grounded by contractarian theory."

The assumption that we should operate in the first instance according to nature, by principles of natural selection, and not by a presumption of moral equality, is itself unargued for. Gauthier attempts to mask this procedural assumption by using Kantian concepts as a "beard" – the concepts of autonomy and dignity, for example, in his discussion of the liberal individual[113] and elsewhere. The description of the autonomy of the liberal individual in Gauthier's society is plausible and even inspiring; but it was

purchased at the great cost of the wholesale stripping of large portions of humanity of *their* dignity and the reduction or elimination of their access to opportunities for exercising and increasing their autonomy, without having argued for the initial assumption.

The discussion of autonomy in solitude is plausible as well – our abilities and capacities in fact *are* all we have to make some headway under our particular circumstances. But to import this perspective, eminently reasonable in solitude, into a society is to use survival of the fittest to beg the question of what constitutes human worth. Kant explicitly distinguishes between "price" and "dignity":

> In the realm of ends everything has either a *price* or a *dignity*. Whatever has a price can be replaced by something else as its equivalent; on the other hand, whatever is above all price, and therefore admits of no equivalent, has a dignity. That which is related to general human inclinations and needs has a *market price* ... but that which constitutes the condition under which alone something can be an end in itself does not have mere relative worth, i.e., a price, but an intrinsic worth, i.e., *dignity*.[114]

Leaving aside the question of whether Kant or Gauthier is right about what constitutes human worth, there is a further problem concerning Gauthier's standard of human valuation. If Gauthier's Robinson Crusoes were Miltons or Hawkings, they clearly could not survive in solitude on a desert island and would therefore be eliminated from the initial bargaining position. However, were they to be simply found in society, undoubtedly in real life Gauthier would permit some sort of state subsidy for them, on the grounds that they would benefit the society – they would be valuable, or in other words, they would have a price. But under the rigorous theoretical screening process that Gauthier institutes, neither of these would in fact have a chance to, as he puts it, realize

their autonomy.[115]

In Gauthier's theory, the impartiality characteristic of moral deliberation has been made to resemble the impersonality of nature, which is generally understood to be a non-deliberative, non-intentional set of processes. I suggest that this qualitative difference puts his theory itself "beyond the pale" of morality and reduces it instead to a force of brute nature in its own right.

NOTES

Chapter Six

[1] Locke, *Second Treatise*, V, §34.

[2] Ibid.

[3] Ibid., II, §4.

[4] Ibid., II, §5.

[5] Ibid., II, §6.

[6] Ibid., II, §14.

[7] Ibid., II, §15.

[8] Ibid., III, §19

[9] "Locke's astonishing achievement was to base the property right on natural right and natural law, and then to remove all the natural law limits from the property right." C.B. MacPherson,*The Political Theory of Possessive Individualism* (Oxford, England.: Oxford University Press, 1962), 199.

[10] Ibid., V, §27. Gauthier notes that this property right in one's own person is assumed by Locke, rather than argued for; Gauthier derives this right directly from Locke's proviso. *Morals*,VII, 4.1.

[11] Ibid.

[12] Locke, *Second Treatise*, V, §31, 332.

[13] Ibid., V, §32, 332.

[14] Ibid., V, §33, 333.

[15] Ibid., V, §37, 336.

[16] Ibid., V, §34, 333.

[17] Ibid., V, §37, 337.

[18] Ibid., V, §40, 338.

[19] Ibid., V, §42, 339.

[20] Ibid.

[21] Ibid., IX, §124, 395.

[22] Ibid., IX, §123, 395.

[23] Ibid., V, §34, 333.

[24] Ibid.

[25] For example, in Gauthier.

[26] Locke, *Second Treatise*, IX, §123, 395.

[27] Ibid., XI, §142, 409.

[28] Ibid., XIII, §158, 419.

[29] Ibid., V, §35, 334.

[30] Ibid., V, §50. 344.

[31] "[W]hile the labouring class is a necessary part of the nation its members are not in fact full members of the body politic and have no claim to be so; and ... the members of the labouring class do not and cannot live a fully rational life. 'Labouring class' is used here to include both the 'labouring poor' and the 'idle poor', that is, all who were dependent on employment or charity or the workhouse because they had no property of their own by which, or on which, they might work. These ideas were so generally prevalent in Locke's day that it would be surprising if he had not shared them." MacPherson, *Possessive Individualism*, 222.

[32] MacPherson's interpretation of the cause of these outcomes is somewhat different (MacPherson, *Possessive Individualism).* His argument is that Locke reads back into his theory some unchallenged assumptions concerning the natural differential status of the propertied and unpropertied classes in his own society. MacPherson differentiates two stages of Locke's state of nature, pre- and post-monetary. While all are assumed to be equally rational and therefore equally capable of fending for themselves in the pre-monetary stage, appropriating goods according to the limits of the proviso, after the introduction of money the proviso has no immediately obvious function; it is reserved and reinterpreted at a later time as justification of the unlimited appropriation of goods made possible by the tacit consent to use non-perishable gold and silver to obviate the earlier-stage objection concerning waste. Therefore, "[t]hose who were left without property after the land was all appropriated could not be accounted fully rational." (p. 238) Then, with the development of civil society, MacPherson notes the ambiguity in Locke's use of "property" -- "...everyone ... is included, as having an interest in preserving his life and liberty [but] only those with 'estate' can be full members" since only they are fully rational. (p. 248)

In MacPherson's analysis, it is the introduction of money that definitively and irrevocably divides previously-equal persons into two classes -- those who are fully rational (and thereby deserving of full participation in the society to be compacted and agreed to) and those who are not fully rational because they have no estates and are forced to sell their labor, thus alienating even the property in their self that was initially given, and necessarily diminishing their own personhood thereby and therefore voluntarily sacrificing their opportunity for full participation.

[33] This is similar to what Brian Barry calls the "coordination problem of first-order impartiality," the imposition by law of strict impartiality in all ordinary encounters. Brian Barry, *Justice as Impartiality* (Oxford, U.K.: Oxford University Press, 1995), 204.

[34] Nozick, *Anarchy*, 151-152.

[35] Ibid., 5.

[36] Ibid., 151.

[37] Ibid.

[38] Ibid., 153.

[39] Ibid., 154.

[40] Ibid., 155.

[41] Ibid., 157.

[42] Ibid.

[43] Ibid.

[44] "... actions ... come already tied to people who have entitlements over them." Ibid., 135.

[45] Ibid., 262.

[46] Ibid., 263-4.

[47] Ibid., 18.

[48] Ibid., 19.

[49] Concerning the appearance of redistribution in this instance, Nozick says, "...the term 'redistributive' applies to types of *reasons* for an arrangement, rather than to an arrangement itself. ... Finding compelling nonredistributive reasons would cause us to drop this label." (*Anarchy*, 27)

[50] Ibid., 185.

[51] Ibid., 198.

[52] Ibid.

[53] Ibid.

[54] Ibid.

[55] Ibid.

[56] Ibid., 201.

[57] Ibid., 202.

[58] Ibid., 203.

[59] Ibid., 203n.

[60] Ibid., 168.

[61] Bernard Williams, "The Idea of Equality." In *Philosophy, Politics, and Society*, 2nd ser., eds. Peter Laslett and W.G. Runciman. (Oxford, England.: Blackwell, 1962). Reprinted in Joel Feinberg, ed., *Moral Concepts* (New York: Oxford University Press. 1969). Quoted in Nozick, *Anarchy*, 233.

[62] Ibid.

[63] Ibid.

[64] Ibid., 235.

[65] Ibid., 215. Elsewhere in *A Theory of Justice*, Rawls uses the softer term "mitigates" (e.g., p. 73). The sense of the whole theory supports interpretation of the softer term rather than "nullifies," since it is clear that Rawls does expect and support the differential rewards earned by the better-endowed and more socially advantaged, so long as these rewards follow a distribution structured by the difference principle.

[66] Ibid.

[67] Ibid., 216.

[68] Rawls, *Theory of Justice*, 72.

[69] "The construction proposed by T.M. Scanlon departs from that of Rawls in two fundamental ways. The first is that the parties are aware of their identities and hence of their own interests. The second is that they are not motivated simply by the wish to advance their interests. Rather, we are to conceive of them as motivated by 'the desire for reasonable agreement'." (Barry, *Justice*, 67.)

[70] Ibid., 202.

[71] Ibid., 204.

[72] Ibid., 214.

[73] Nozick, *Anarchy*, 160.

[74] Ibid., 152.

[75] Gauthier, *Morals*, 268, 17.

[76] Ibid., 268.

[77] Barry terms *Morals by Agreement* "Gauthier's morally pathological universe." Barry, *Justice*, 42.

[78] Locke, *Second Treatise*, II, §6, 311.

[79] For example, see his n. 30 on page 18: "Speaking euphemistically of enabling [the handicapped] to live productive lives, when the services required exceed any possible products, conceals an issue which, understandably, no one wants to face."

[80] Gauthier, *Morals*, 268.

[81] Ibid., 203.

[82] Ibid., 204.

[83] Locke, *Second Treatise*, V, §27, 329.

[84] Gauthier, *Morals*, 203.

[86] Rawls, *Theory of Justice*, 101, 179, and elsewhere.

[85] Gauthier, *Morals*, 4.

[87] Ibid., 5.

[88] Ibid., 6.

[89] Ibid., 7.

[90] Ibid.

[91] Ibid.

[92] Ibid., 9.

[93] Ibid.

[94] Ibid., 10.

[95] Ibid., 15.

[96] Ibid., 16.

[97] Ibid., 17.

[98] Ibid., 18.

[99] Ibid.

[100] Ibid., 200-201.

[101] Ibid., 202.

[102] Ibid., 203.

[103] Ibid., 209.

[104] Ibid., 211.

[105] Ibid., 214.

[106] Ibid., 214-217.

[107] Ibid., 217.

[108] Ibid.

[109] Here again an author who opposes the difference principle tendentiously misreads what Rawls has actually said and meant. Gauthier says, "Justice ... is the disposition not to take advantage of one's fellows, whether as a free-rider or as a parasite. It appears that the lexical difference principle licenses those with lesser natural talents to take advantage of those naturally more fortunate, requiring the latter to use their abilities, not primarily for their own well-being, but to maximize the minimum level of well-being."

(252) However,Rawls says that the interpretation of the two principles "seeks ... to *mitigate* the influence of social contingencies and natural fortune on distributive shares" (*Theory of Justice*, 73; emphasis added), *not* to "require" the more fortunate to work "primarily" to "maximize the minimum level of well-being." The difference principle is in fact intended to "license" those with *greater* natural talents to indeed benefit differentially, under the constraints of the principle.

Gauthier's argument against Rawls' idea of a "natural lottery" verges on the ridiculous, without having passed first through a stage of being sublime. He says that there is no "pool fixed to guarantee winners and losers ... and if there is a distribution, there is no distributor -- unless we assume a theistic base ..." (220). But of course, how the distribution occurs, and whether there is a literal lottery or not, is entirely beside the point. Endowments, deficiencies, and social circumstances *are* "distributed," whether by a theistic distributor or by genetic and environmental chance, and to deny that one's draw in this "lottery" can significantly affect one's chances in life by transforming "affecting one's chances" to "guaranteeing winners and losers" is simply deliberate misreading. Objecting to Rawls' metaphor does not refute Rawls' argument. Moreover, to address a further objection of Gauthier's, one can consistently view persons both as "creatures of a distributor" *and* as "rational individual actors."

[110] Ibid., 220.

[111] Ibid., 203.

[112] Ibid., 218. Gauthier finds this scenario "distressing" but urges us not to be "misled" by it. Surely it has come about by either a violation of the proviso somewhere along the line or by some other violation, which ought, morally, to be rectified. The implication is that certainly this scenario would not come about by straightforward application of his moral theory as it stands and therefore be "beyond the pale of a morality tied to mutuality."

[113] Ibid., 346-7.

[114] Kant, *Foundations*, Second Section, 53.

[115] Gauthier, *Morals*, 339.

CHAPTER VII

IMPARTIALITY IN KANTIAN THEORIES

So far, impartiality has not shown itself to be free of problems in its application, either in the mode of sympathy or in the mode of detachment. The difficulties in both are those that are typically associated with internal coherence and both can lead to potentially exclusionary outcomes.

In the mode of sympathy impartiality has shown itself to be arriving on the scene of moral deliberation in a sense too late, well after bounded affiliations have been established and biases developed – the same affiliations that enable a capacity for sympathy to begin with but that subsequently interfere with its operation. One is able to set aside one's particularities temporarily in order to incorporate those of another, but usually only some other who is in key respects similar to oneself. The outcomes are perhaps predictable – one can more easily mobilize sympathetic impartiality for those whose affiliations resemble one's own, but has more difficulty with those whose affiliations are alien or incomprehensible, which is exactly what leads to these kinds of exclusionary potentials.

When one acknowledges this natural tendency in oneself and desires to be as fair-minded as possible in moral deliberation, one may try instead for a modality of detached impartiality, seeking a broader inclusiveness from a farther perspective. Detached impartiality can end up, however, running aground on the same rocks as sympathetic

impartiality. The specifications of the "farther perspective," the location on the spectrum of moral engagement at which one is presumably far from one's own particularities and yet close to the recognition of human commonalities, are themselves shaped by the same affiliations as before. The influences are more deeply masked from oneself by one's intention to conduct a sincere search for an impartial stance, but detached impartiality may very well end up with similar outcomes of exclusion from the moral community.

In examining both of these modalities of impartiality, scrutiny has focused on these and other (sociological or cultural) influences upon the deliberating agent. Alasdair MacIntyre,[1] among many others, has famously recognized at least one of these influences: the standards of universality of liberal individualism, he says, are but one "partisan" approach to theorizing among others;[2] and are part of an intellectual "tradition of inquiry," where "tradition" is defined as

> an argument extended through time, in which certain fundamental agreements are defined and redefined in terms of two kinds of conflict: those with critics and enemies external to the tradition who reject all or at least key parts of those fundamental agreements, and those internal, interpretative debates through which the meaning and rationale of the fundamental agreements come to be expressed and by whose progress a tradition is constituted.[3]

MacIntyre's challenge is that impartiality, which plays such a vital theoretical role in moral deliberation and is presumed to be a state of mind that can be fruitfully accessed, may turn out to be another "Idol" of the liberal-individual tradition. Concerning impartiality, MacIntyre tells us that the "relationship between theory and fact cannot be gotten neutrally" since any description of examples, or "facts ... already proceeds from a

theoretical framework. [Therefore] each theory of practical reasoning is ... a theory as to how examples are to be described."[4] There can be, in short, no non-tradition-constituted moral (or, presumably, other) inquiries, because to imagine a standpoint such as that would be to have no intellectual resources whatsoever; "... to be outside all traditions is to be a stranger to inquiry" altogether[5] – in short, it is to occupy once again the standpoint of the View from Nowhere. But if there can be no impartial (non-tradition-constituted) standpoint within theory, and outside tradition we are Nowhere, how can impartiality then be fruitfully conceptualized? Or must it be discarded altogether?

MacIntyre himself must acknowledge the operations of an impartiality that is in fact outside any particular tradition of inquiry when he says that traditions are to be conceived as characterized by change and conflict; in the movement from the first stages of a tradition, characterized by "unquestioning obedience," to the later stages, characterized by "rival interpretations incompatible with the original," even MacIntyre must concede the existence of "inventiveness."[6] The question, in his own terms, must be where the inventiveness comes from, if not from a standpoint somehow outside the tradition and yet not so far removed that it is outside the possibility of inquiry altogether.

But attempts to theorize about impartiality are indeed vulnerable to being skewered by these types of criticisms, as we have seen in the analyses of sympathy and detachment, and in fact the basic concept of impartiality has itself come under fire as a product of certain assumptions that prevail about the relationship between rationality and morality. A general focus of criticisms of impartiality has come in recent times from a recognition that partiality, while it certainly seems to be the antithesis of impartiality, is nevertheless the foundation of the relationships and connections that make our lives meaningful. The "ethic of care" which has contrasted itself to the "ethic of justice"[7]

attempts to supplement the function of impartiality with the function of care, conceptualizing these two ethics as differing frameworks within which questions of morality and ethical dilemmas will be differently described and thereby differently answered.[8] The "attachment" characteristic of the ethic of care is further contrasted with "detachment," and the concept of "need" is positioned in contrast to "equality":

> [T]hese perspectives denote different ways of organizing the basic elements of moral judgment: self, others, and the relationship between them. With the shift in perspective from justice to care, the organizing dimension of relationship changes from inequality/equality to attachment/detachment. ... Within the context of relationship, the self as a moral agent perceives and responds to the perception of need. ... [C]are is grounded in the assumption that self and other are interdependent ... Within this framework, detachment, whether from self or from others, is morally problematic, since it breeds moral blindness or indifference – a failure to discern or respond to need.[9]

The two styles of impartiality discussed so far seem particularly vulnerable to criticisms of this sort. The question is, given that this is a reasonable characterization of two purportedly different frameworks of moral deliberation, whether the frameworks of justice and care are really distinct, or whether they exist in some relationship to one another, for example, of parallelism or subsumption. Since the present exploration concerns the functioning of impartiality, a classic "justice" concept, the question becomes, for our purposes, whether justice can subsume, include, or otherwise be integrated with care. If it can, this would offer the possibility of a theoretical framework within which impartiality need not necessarily function to exclude persons from the moral community, as it seems prone to doing in the two modalities discussed so far. It is in this

context that I want to examine the operations of impartiality in two Kantian theoretical frameworks to see whether either of them can deliver a style of impartiality that will accomplish this goal. I will use Rawls for his Kantian interpretation and Kantian constructivism and Barbara Herman's conception of "deliberative fields".[10] Finally, I will look at Hare's utilitarian theory, which conceptualizes sympathy in such a way that it not only avoids the problems of sympathy as duscussed earlier, but also has an outcome that is distinctly Kantian. To begin with, a summary of the main elements of Kant's ethics will provide a touchstone to understanding the contrasts of these various interpretations.

Summary of Kant's Ethics

Kant conceptualized impartiality in a specific way which offers a clue as to why impartiality, both in theory and in practice, has been so difficult both to grasp and to implement. In Kantian ethics, the piece missing from previous modalities of impartiality emerges as a way to ground a genuine impartiality. His conceptualization gives a way to revitalize impartiality as an essential component of moral decision-making and fortify it against criticism, by showing, first, how a true impartiality actually may be possible, and second, why it need not be conceived separately from relationships of partiality. Kant focused his scrutiny not in the first instance upon the moral individual, the agent, but upon the moral act itself – what is actually done, chosen, acted upon. Only after the structure of the action is thoroughly understood and its principle revealed does he work backwards to discover therein the necessary specifications of the moral individual, and subsequently his conception of impartiality can be revealed. These specifications are universal; they are both available to and apply to every rational being without exception,

and therefore give reason to imagine the possibility of a genuine, practical impartiality.

Kant does not speak about impartiality as such. A concept of impartiality must be constructed or gleaned from what he uncovers concerning the structure of rationality itself, specifically of its employment as practical reason. Kant asks the question, "Can reason be practical?" This question is far from being as simple as it sounds and there are two aspects to the answer. The question must be answered to account for how it is that as a part of law-ruled nature we can also be free; it must also be answered to show whether pure practical reason can provide its own motivation to act morally. At the end of these explorations, the impartial stance emerges from conceptualizing the intelligible world as a standpoint from which the individual is moved to reconcile his own and others' actions in the world of appearances against the constraint of the categorical imperative. Specifically, it is found in the concept of the rational being forced to conceptualize an intelligible world as a standpoint from which to resolve the apparent contradiction between himself as part of nature and himself as free causality. This is possible because Kant recognizes that a genuinely rational being may, or in fact, must see himself under two entirely different aspects.

Kant begins with the program to "construct a pure moral philosophy which is completely freed from everything empirical"[11] and observes that such a philosophy must be possible, as is "self-evident from the common idea of duty and moral laws"[12] which apply to all rational beings, including but not limited to human beings. Therefore the particularities of being human cannot apply to pure moral philosophy; the study of these belongs to "anthropology."[13] What Kant seeks is a metaphysics of morals, a system of pure *a priori* moral principles knowable to reason; to undertake this task he will

> proceed analytically from common knowledge to the determination of its supreme principle, and then synthetically from the examination of this principle and its sources back to common knowledge where it finds its application.[14]

The root of morality is what Kant calls the "good will" – the only thing either in or out of the world "which could be called good without qualification."[15] The good will is embodied in the concept of duty and is expressed in actions that are done for the sake of duty, rather than merely in accordance with duty.

The first principle of morality, then, is that "to have moral worth an action must be done from duty."[16] The second principle is that its moral worth resides not in the outcome to be accomplished but "in the maxim by which it is determined."[17] From these comes the third principle: "Duty is the necessity of an action executed from respect for law" irrespective of inclinations or other motivations.[18] It is only under this condition that "the will can be called absolutely good without qualification," i.e., the condition of "universal conformity of its action to law as such."[19] Kant observes that this is precisely how persons act when they employ their common reason in practical judgements: they subject their will to the form of universal law, asking (in effect, if not in this abstract form) whether they could will that the (subjective) maxim of their actions become universal law.

Because it is difficult, if not impossible, to determine whether someone (including oneself) acts from pure duty, there is a temptation to infer that the concept of duty resides in empirical grounds, such as self-love or the love of humanity; however, the point is not to describe what *is* done but what *ought* to be done, "even if there never were actions springing from such pure sources."[20] But the source of unlimited and unconditional respect for law, a respect inherent in all rational beings, not just human ones, cannot in

principle be derived from contingent human experiences or characteristics. Nor can this respect be logically derived from examples of morality, since examples themselves, even of the "Holy One of the Gospel,"[21] must be judged ahead of time in order to qualify as examples, necessitating a prior standard. Even the judgement of God as the highest good must come "solely from the idea of moral perfection which reason formulates a priori and which it inseparably connects with the concept of a free will."[22]

Accordingly, in order to move from a popular philosophy of morals to a metaphysics of morals, a system of a priori principles which "is not held back by anything empirical," it is necessary to examine closely the "practical faculty of reason from its universal rules of determination to the point where the concept of duty arises from it."[23]

Will is the capacity to act according to a conception of law, and this capacity is reserved to rational beings alone. This capacity is in contrast to all other elements of nature, which are necessarily bound by its laws. A holy will is infallibly determined by reason; but a will that is also subject to the pull of inclination will experience duty as a constraint. This constraint is "a command of reason, and the formula[s] of this command [are] called ... imperative[s]," all of which are expressed by an "ought."[24] The "ought" of an imperfect will is a degraded form of the "is" of a holy will, for which imperatives are out of place and beside the point because "according to its own subjective constitution, it can be determined to act only through the conception of the good."[25]

The moral imperative for the imperfect will must be categorical, since the imperative is independent of empirical considerations. It determines an action that is good in itself, and not good only as a means to something else (which would make it "hypothetical"): the moral imperative reflects the good will in the intention of the action.[26]

The question now is, how is this possible? How can this "constraint of the will ... be conceived?"[27] This question must be investigated a priori, since, as before noted, we have no means of ascertaining by experience, either others' or our own, whether this imperative is real. It is evident that the imperative itself is contained within the mere concept of a categorical imperative; since there are no restrictions of the law in the form of conditions and the only other element is that of the necessary concordance of one's subjective maxim with this law, "there is nothing remaining in it except the universality of law as such."[28]

> There is, therefore, only one categorical imperative. It is: Act only according to that maxim by which you can at the same time will that it should become a universal law.[29]

This answers the first part of the question, whether practical reason is possible. The second part of the question, how it is possible for reason to determine conduct, is the next step in the deduction, and concerns incentives.

Incentives concern subjective material ends which a rational being proposes to himself and which are therefore merely relative.[30] They have worth only insofar as the desire of the subject gives them worth. But suppose, Kant asks,

> that there were something the existence of which in itself had absolute worth, something which, as an end in itself, could be a ground of definite laws. In it and only in it could lie the ground of a possible categorical imperative, i.e., of a practical law.[31]

This is the case for every rational being, who is an end in himself and bears the designation, because he is self-legislative, of "person."

> Thus if there is to be a supreme practical principle and a categorical imperative for the human will, it must be one that forms an objective principle of the will from the conception of that which is necessarily an end for everyone because it is an end in itself. ... The ground of this principle is: rational nature exists as an end in itself.[32]

The categorical imperative thus transforms itself into a formula bearing within it the universal motive for human conduct:

> Act so that you treat humanity, whether in your own person or in that of another, always as an end and never as a means only.[33]

Again, this imperative is not and cannot be derived from experience – since it applies to all rational beings, of which we have experience only of human rational beings, and we cannot ascertain the quality of human action in relation to inner understandings of motive in any case. This idea therefore must arise from pure reason and it is recognized as "the idea of the will of every rational being as making universal law,"[34] as self-legislative: "subject to the law (of which it can regard itself as the author).[35]

This recognition of the self-legislating nature of the rational being, "the principle of every human will as a will giving universal laws in all its maxims,"[36] is what Kant terms "autonomy," which leads to the concept of a realm or kingdom of ends: the "systematic union of different rational beings through common laws."[37]

The answer to the second part of the question, therefore, is that autonomy is the form of volition in general, which would be the sole content of the absolutely good will (the holy will);

> ...the capability of the maxims of every good will to make themselves universal laws is itself the sole law which the will of every rational being imposes on itself, and it does not need to support this on any incentive or interest.[38]

Kant has shown from analysis of the development of the "universally received concept of morals" that its foundation is autonomy and that "will is a kind of causality of living beings so far as they are rational." Freedom is the

> property of this causality by which it can be effective independently of foreign causes determining it, just as natural necessity is the property of the causality of all irrational beings by which they are determined in their activity by the influence of foreign causes.[39]

But how do we know that we can be free? A free will, in its positive concept, is identical with a will under moral laws, and must be presupposed in order to have morality at all since morality follows by "mere analysis" of the concept of freedom.[40] But the "negative" concept of freedom, that will can have its own causality independent of nature, is insufficient to motivate moral action; for this, the concept of autonomy must be understood and placed in its proper context.

Morality and its supreme principle, the absolutely good will, may be analytically broken out of the concept of freedom of the will. But the "absolutely good will is one whose maxim can always include itself as a moral law." This is a "synthetical" proposition because "by analysis of the concept of an absolutely good will the property of the maxim cannot be found." It requires a unifying third cognition, containing the previous two, in order to make the proposition possible. This third cognition is

"furnished" by the positive concept of freedom, and is the capacity of the will to recognize its own lawful causality under freedom, and hence to have the property of being a law unto itself. Kant terms this capacity "autonomy," which

> only expresses the principle that we should act according to no other maxim than that which can also have itself as a universal law for its object. And this is just the formula of the categorical imperative and the principle of morality."[41]

This does not yet explain what it is that motivates moral action, or respect for the moral law which must lead to moral action; nor does it explain how this motivation is possible.

As rational beings we must necessarily regard ourselves as free; the defining characteristic of such a being is that it "cannot act otherwise than under the idea of freedom, [and] is thereby really free in a practical respect" irrespective of whether that freedom is "really" real or can be shown to be so.[42]

> That is to say, all laws which are inseparably bound up with freedom hold for it just as if its will were proved free in itself by theoretical philosophy.[43]

As rational beings we are conscious of our own causality and also conscious of a law of action, the categorical imperative.[44] So far, we have no understanding of why we ought to subject ourselves to this law, but we do take an interest in it because we recognize that if we had holy wills, the "ought" would properly be a "would" valid for every rational being. But we do not see on what grounds the moral law obligates us. "Freedom and self-legislation ... are both autonomy" and therefore neither can be used to furnish a ground for the other.[45]

But if we were to "assume a different standpoint when we think of ourselves as causes a priori efficient through freedom from that which we occupy when we conceive of ourselves in the light of our actions as effects which we see before our eyes,"[46] it might be possible to break out of the circle of freedom and self-legislation to find a motivation for morality. "Reason ... as a pure spontaneous activity ... is elevated even above understanding,"[47] which exists only to "bring sensuous conceptions under rules."[48] Understanding requires sensuous intuitions in order to work, but reason is a "pure spontaneity." Therefore, a being which possesses reason, *as* such a being, must necessarily regard itself as belonging to the world of understanding, the intelligible world, and not to the world of the senses, the world of appearances.

> Thus he has two standpoints from which he can consider himself and recognize the laws of the employment of his powers and consequently of all his actions: first, as belonging to the world of sense under laws of nature (heteronomy), and, second, as belonging to the intelligible world under laws which, independent of nature, are not empirical but founded only on reason.[49]

When we understand ourselves as belonging to the intelligible world (when we think of ourselves as free and recognize the autonomy of our will and its consequence, morality[50]) as well as to the world of nature, we understand the "ought" and why it applies to every rational being (who is a member of both worlds) and thereby why it applies to us. But to explain how our freedom is possible is not given to us.

> Since no example in accordance with any analogy can support it, it can never be comprehended or even imagined. It holds only as the necessary presupposition of reason in a being that believes itself conscious of a will.[51]

This inability to explain free will is of a piece with the "impossibility of discovering and explaining an interest which man can take in moral laws."[52] In the concept of "interest" is encapsulated the most mysterious question of all, "Why ought I to be moral? ... What are the conditions ... that make it possible for [a rational being] to take an interest in the law or to have the law as his incentive?"[53]

The concept of an interest derives from the concept of an incentive; it "indicates an incentive of the will so far as it is presented by reason."[54] It is contrasted with "natural instincts" and is "that by which reason becomes practical." Reason can take a direct or an indirect interest in an action:

> ... if reason can determine the will only by means of [an] object of desire ... reason takes merely an indirect interest in the action ... A direct interest in the action is taken by reason only if the universal validity of its maxim is a sufficient determining ground of the will.[55]

An "indirect" interest in the action must be empirical, since "reason without experience can [not] discover objects of the will," and such an interest must "without exception ... belong under the principle of self-love or happiness."[56] A "direct" interest, however, is one that "find[s] satisfaction in the law of the action"[57] and "since the law itself must be the incentive in a morally good will, the moral interest must be a pure nonsensuous interest of the practical reason alone."[58] Further, the concept of a maxim, the subjective rule of action, derives from that of an interest; a maxim is "morally genuine only when it rests on the mere interest in obedience to the law,"[59] i.e., when the subjective maxim agrees with the objective law. These concepts apply "only to finite beings, for ...

they presuppose a limitation of the nature of the being."[60] A holy will for whom "ought" would be "is" has no need of interest, incentive, or maxim.

In the second Critique, Kant probes the issue of respect for the moral law, that by which we are motivated to obey it. Kant observes that the effects of the moral law can be either negative or positive, and either objective or subjective. By either pathway – the negative constraint of inclination by the moral law, or its positive humiliation, as Kant puts it, of self-conceit (the "inclination to regard one's own subjective maxims and interests as having the authority of law"[61]) – the effect of the working of the moral law within us generates or awakens respect for the moral law. Respect can be likened to a kind of feeling, a moral feeling.[62] In its form as submission to law, or constraint, it evokes "displeasure proportionate to constraint."[63] But because "this constraint is exercised only through the legislation of one's own reason, it also contains something elevating ... [which] can also be called self-approbation."[64] The "interest which is subjectively produced by the law .. has a very special name, viz., respect."[65]

Discussion

Within this framework, it is clear that the motivation to duty (called "respect" by Kant) must involve motivation to impartiality, a concept which takes on a depth of meaning and a clarity that seemed unavailable to impartiality in its previous expressions as sympathy or detachment. Both the (desirable) "sympathetic ideal observer" and the (undesirable) "view from nowhere" – the fullest, albeit abstract and hypothetical theoretic expressions of the modalities of sympathetic and detached impartiality – can be seen in this framework to have been subject to the "critical" error that Kant persistently warns against in his philosophy: reason transcending its proper bounds. Both ideals are

theorized as humanly unattainable end points on a conceptually unbroken continuum originating with the individual as he is in the moment, desiring to act morally. (The usual caveat here is that although one cannot in fact take the standpoint of either of these ideal beings, nevertheless one ought to be able to imagine how someone "who could occupy such a standpoint *might* reason"[66] [emphasis added].) The individual then mobilizes either a presumed natural sympathy or a rational self-interest to attain a state of impartiality, a standpoint from which, either through seeing the world through another's eyes, or by rationally taking account of presumably shared human interests, he feels most confident that his moral deliberation will yield desirable results.

But Kant conceptualizes the human being – in fact, any rational being – as occupying two distinct standpoints. The first is in the world of appearances and nature, where either sympathetic feeling or cognitive detachment from the particularities of the phenomenal ego might have their proper place. But in Kant's view, the entire enterprise of morality takes place from the second standpoint -- the human person, *as* person, is a noumenal being, a dweller in the intelligible world, for whom feelings of sympathy and consideration of worldly interests are irrelevant. In fact feelings and considerations of these sorts would immediately shift the locus of deliberation from the unconditional recognition of and regard for the moral law, back into the phenomenal world of regard for feeling and interests – from categorical to hypothetical. In this view, the counterpart of the ideal observer and the viewer from "nowhere" is the holy or divine will, that will for whom "ought" is "would." But for Kant the holy will seems not to be at the end-point of a continuum but is self-contained. The "would" belongs to a will that is unhampered by anything in the world of appearances – not inclinations, not feelings, not interests – and hence fully free to express its nature without interference or the need for reflection. But

finite, rational beings, who have a foot in each world, are in a different position altogether. For these beings, "the moral necessity is a constraint" upon the inclinations of their phenomenal being and is thus "not ... a manner of acting which [they] naturally might favor."[67] From this standpoint there may indeed be a continuum of sorts, but it is an empirical continuum consisting of the degree to which an individual may be willing to set aside his inclinations in order to do his duty. However, he can *know* his duty only insofar as he is a member of the intelligible world.

Impartiality must then be reconceptualized in order to be understood in the context of this scheme. As concerns the holy will, impartiality is once again irrelevant, as it was with both the sympathetic ideal observer, omniscient and perfectly sympathetic, who has no need of it, as well as with the viewer from "nowhere" who sees all connections and actions and their inevitability in the relentless determination of things, and for whom moral constructs such as impartiality are also irrelevant, for different reasons.

For the finite rational individual, however, impartiality has both a subjective and an objective aspect, in Kant's terminology. In its objective aspect, impartiality is embedded in the respect for the moral law which becomes evident as one's self-conceit is humiliated. The workings of the moral law within us, on this level, are largely unknown to us, but operating from its base in the intelligible world, this striking down of self-conceit causes the moral law to be experienced with "the greatest respect and [is] thus the ground of a positive feeling which is not of empirical origin [and] can be known a priori."[68] Thus, in a practical sense, it is the movement toward impartiality that is motivated by respect for the moral law.

Subjectively, the primacy of our inclinations, unchecked, can lead to the

pathologically-based self-pretensions of our maxims to universality; "this propensity ... in general can be called self-love; when it makes itself legislative ..., it can be called self-conceit. ... The moral law ... forever checks self-conceit" and "if anything checks our self-conceit *in our own judgement*, it humiliates."[69] [emphasis added] It is our *awareness* of something we might have put forward as the determining ground of our will being humiliated by the moral law that *subjectively* awakens respect for it in us and motivates impartiality.

The form taken by Kantian impartiality is a noumenal humility before the moral law, universal and best seen in his conception of finite rational beings as ends in themselves. As an end in himself, the individual is necessarily morally bound to other persons by virtue of his recognition of their equal status as ends in themselves, thereby constituting a realm or kingdom of ends. In Kant's own terms, it hardly matters whether this recognition is conscious on an individual basis or not; it is a defining characteristic of personhood that all rational beings are ends in themselves, that is, persons, and as such, the moral binding, and therefore the motivation to impartiality in the recognition of our equality with others, is presumably already in place.

Rawls: Kantian Interpretation and Kantian Constructivism

Given this framework, Rawls' *A Theory of Justice* can be reconsidered in a different dimension from the largely psychological framework considered earlier. Previously, it was unclear how a concept of impartiality could be understood in Rawls' theoretical framework. Either it was to be understood and defined as an explicit artifact of the coherentist scheme of the original position, or a qualitatively different concept of

impartiality was to have been invoked as a critical mechanism from "outside" the theoretical framework, but it was not evident how to do that. In the first case, analysis of impartiality seemed to be blocked by the fact that the same theoretical mechanism seemingly ended up both defining impartiality and generating the principles of justice to be evaluated by that same impartiality. In the second case, it was unclear what sort of impartiality could be called upon to evaluate a theoretical framework which included a specific conception of impartiality as part of its existing structure, implying the necessity of another, larger theoretical framework from which to evaluate it.

From a Kantian perspective, however, some of these issues can be fruitfully addressed from precisely within the existing framework, without losing sight of its coherentist aspect but digging out an analysis from a different place than "outside" itself.

Rawls draws upon Kantian concepts to reinterpret the theoretical framework of justice as fairness. First, he brings out Kant's idea that "moral principles are the object of rational choice"[70] and concludes that "moral philosophy becomes the study of the conception and outcome of a suitably defined rational decision."[71] This leads to the consequences that Rawls later also discusses in his "Kantian Constructivism in Moral Theory"[72] – that the principles derived from such a procedure must be public and that the primary condition under which these decisions occur is a characterization of the decision-makers as "free and equal rational beings."[73] Rawls describes the original position as an "attempt to interpret this conception."[74] The original position therefore expresses the autonomy – the self-legislation – of the deciders and can be conceptualized as a perspective from which noumenal beings may see the world.[75] He further conceptualizes the original position as supplying the "missing part of [Kant's] argument" that morality stems from the phenomenal self trying to express its true (noumenal) nature as a free and

equal rational being – the missing part being the "concept of expression": the original position becomes the cognitive-metaphysical impartial "venue" from which one may express one's true self in the phenomenal world.[76] The principles of justice thereby derived may then be considered as "categorical imperatives in Kant's sense," since for Kant a categorical imperative is "a principle of conduct that applies to a person in virtue of his nature as a free and equal rational being."[77] Rawls concludes his "Kantian Interpretation" by viewing "the original position ... as a procedural interpretation of Kant's conception of autonomy and the categorical imperative."[78]

So far, there is not much more to be said about the concept of impartiality in the original position than has been said earlier. The deciding parties, in a stripped-down condition of knowledge and information concerning their particular selves, are in a forced state of neutrality regarding their personal ends from which they must generate principles that will define the basic structure of a just society. The "fairness" of "justice as fairness" refers to this very state of restricted personal knowledge, situating all parties "fairly with respect to one another,"[79] prior to their engaging in rational deliberation to choose principles. Still, this concept of fairness bears little relationship to the commonsense ideas of impartiality that arise when one is attempting an analysis of what impartiality is and how it works, since the dual problems concerning impartiality in the original position have not yet been resolved: the tendency of the original position to be culturally biased in its construction, and its explicit coherentism within this potential bias. In his lectures on Kantian constructivism, however, Rawls gives some deeper insights into the nature of the original position, insights that break through the problem of how to understand impartiality in the light of these two characteristics.

There are two aspects of constructivism as used in his theory to which Rawls

draws particular attention. One is his emphasis on the specific conception of the person that plays the vital part of agent of the construction, and second is his distinction between what is Reasonable and what is Rational.

> The procedures of constructivism are contrasted with the search for moral truth interpreted as fixed by a prior and independent order of objects and relations, whether natural or divine, an order apart and distinct from how we conceive of ourselves.[80]

It is in this contrast that the analysis of impartiality in Rawls is most interesting. Attending to a common-sense understanding of impartiality would likely reveal that what one aims for in seeking an impartial stance in moral deliberation is precisely a perspective from which to be able to "get the right answer," one that is revealed when we lay aside our parochial perspectives. The idea of such a "right" answer leads naturally to a belief that it might be possible to apprehend, however dimly and imperfectly, the outlines of a system of moral truth that is assumed to be "out there" somewhere. What else could the sympathetic observer be about, or the sufficiently-detached rational decider? The God's-eye view is emblematic of a theoretical being who "sees" the "truth" and thus the existence of a "truth" is almost taken for granted. But in Rawls' Kantian constructivism he argues squarely against such a view:

> ... the idea of approximating to moral truth has no place in a constructivist doctrine: the parties in the original position do not recognize any principles of justice as true or correct and so as antecedently given; their aim is simply to select the conception most rational for them, given their circumstances. This conception is not regarded as a workable approximation to the moral facts: there are no such

> moral facts to which the principles adopted could approximate.[81]

This view is radical, honest, devoid of any kind of Platonism, and reflects Hume's assessment of the place of other people's opinions in moral theory. However, it draws the concept of impartiality into another realm altogether. If impartiality is not a state wherein we can "approximate to moral truth," then it must be explicitly defined as nothing more or less than the state of agreement among the parties to the original position. And in that case it is vital to have a deeper characterization of who these parties are or should be, especially when there is in principle no independent way to evaluate their worth as deciders who will generate morality from within themselves.

There are three "model-conceptions" that Rawls develops in his theory: the well-ordered society, the moral person, and the original position, a "mediating model-conception" between the first two, whose role is to

> establish the connection between the model-conception of a moral person and the principles of justice that characterize the relations of citizens in the model-conception of a well-ordered society.[82]

The well-ordered society has the following characteristics: first, its regulating principles of justice are public and known to be so; its basic structure is believed by everyone to satisfy these principles; and the principles are founded on the publicly-shared and well-established beliefs of the society "as established by the society's generally accepted methods of inquiry."[83] Second, the members of such a society are free, moral, and equal persons: they are moral in that they have and believe other members have both a sense of justice and a conception of their good; they are free in that they recognize and affirm that

their conceptions of their personal goods and ends, chosen in the light of their higher interests, may change; and they are equal in that they regard others as on a par with themselves as having an equal deciding say in determining what the basic structural principles of justice for their society will be.[84]

Rawls also identifies three different perspectives from which the well-ordered society may be viewed for the purposes of assessing its basic structure, and he develops these three perspectives explicitly as a response to the standard objection to his coherentism. The perspectives are

> ... that of the parties in the original position, that of the citizens in a well-ordered society, and that of you and me who are examining justice as fairness to serve as a basis for a conception that may yield a suitable understanding of freedom and equality.[85]

In the mediating original position, the parties to deliberation are not the actual citizens of the society, in possession of their full autonomy, but they are "artificial agents" in the construction of the society, their autonomy characterized more narrowly as "rational autonomy," the kind that is typically employed in pursuit of ends specified in hypothetical imperatives. Thus, the parties are

> ... rational in their deliberations to the extent that sensible principles of rational choice guide their decisions. Familiar examples of such principles are: the adoption of effective means to ends; the balancing of final ends by their significance for our plan of life as a whole and by the extent to which these ends cohere with and support each other; and finally, the assigning of a greater weight to the more likely consequences; and so on. ... the rational is interpreted by the original position in

reference to the desire of persons to realize and to exercise their moral powers and to secure the advancement of their conception of the good.[86]

Rawls contrasts this conception of the Rational with another, fuller conception that he terms the Reasonable. Where the Rational expresses "each participant's rational advantage," the Reasonable incorporates Rationality into a framework of reciprocity and mutuality, expressing the "fair terms of cooperation" of the original position – in other words, explicitly adding a dimension of morality. The sole relevant criterion of the participants in the original position is that the veil of ignorance assures that all parties are represented solely as possessing the "minimum adequate powers of moral personality," defined as "the powers that equip us to be normally cooperating members of society over a complete life."[87]

In short, the perspective from which to view the problem of impartiality in a coherentist and culturally biased framework is that of the parties to the original position – moral persons whose morality is embodied in, and for the purposes of Rawls' theory, exhausted by, the capacity for social cooperation over a complete life. It is their "coherence" with each other – their mutual agreement – that structures the society according to the principles selected and agreed upon by them, and that also incidentally vitiates the importance of any potential cultural "skew" to the description of the original position. And it is the veil of ignorance that forcibly elicits from them the particular kind of impartiality that drives the engine of social choice: the impartiality of the moral person, defined by his desire and ability to enter into lifetime relationships of reciprocity and mutuality, informed by rational considerations of advancing the agreed-upon good.

The connection of the impartiality of the participants with the objectivity of the

principles chosen is then made clear by Rawls:

> [t]he rational intuitionists' objection, properly expressed, must be that no hypothetical agreement by rationally autonomous agents, no matter how circumscribed by reasonable conditions in a procedure of construction, can determine the reasons that settle what we as citizens should consider just and unjust; right and wrong are not, even in that way, constructed. ... If on the other hand, such a construction does yield the first principles of a conception of justice that matches more accurately than other views our considered convictions in general and wide reflective equilibrium, then constructivism would seem to provide a suitable basis for objectivity.[88]

"Deliberative Fields"

Up to this point, consideration of various forms and venues of impartial moral deliberation has not yielded a usable conception of impartiality that can either escape from, or make use of, the confines of the moral deliberator's own biases and predilections. This is unsurprising, but disappointing, since the danger of exclusion seems always to be lurking as a potential outcome of deliberation. Regardless of the deliberator's awareness of this problem and intention to minimize it as much as possible, the outcome ends only by demonstrating that such biases and predilections can exist on a deeper and more basic level than is perhaps possible to overcome. It appears fruitless to search for a "truly" impartial stance "outside" of any such frame of reference, because then we seem immediately to find ourselves in one of the extreme and ideal positions: either as the ideal sympathetic observer or taking in the view from nowhere, neither of which is useful for the ordinary individual deliberating about moral concerns. On the other hand, anything

short of this throws us immediately back into a situatedness in particular intellectual or social contexts which reinforces itself and tends, apparently inexorably, toward exclusion.

Rawls confronted this difficulty directly and tried to eliminate it as a problem by acknowledging without embarrassment that this situated view of impartiality is in fact the only view possible to us as finite and imperfect rational beings and addressing the problem of inclusion by a scenario of impartiality that would presumably yield his two principles of justice, subsequently giving an interpretation of his theory that would give support to this view from a Kantian perspective. Yet the problem does not thereby appear to be resolved, only explained. There is still the problem of evaluating his view of impartiality from a perspective outside of it.

Barbara Herman's "deliberative fields" do not on the surface offer such a perspective, but on the contrary, give a different Kantian explanation of the necessity for deliberative situatedness in relation to another issue, thus attempting not to override biases and particular interests, but to acknowledge and make use of them as an essential part of a Kantian framework. Paradoxically, in this view, a genuine "outside" impartiality emerges at the end as a real possibility. The deliberative context in her "Agency, Attachment, and Difference"[89] is different from that in Rawls and other theorists. She is concerned to respond to criticisms of Kantian ethics that are focused on the role of impartiality; these criticisms[90] question whether the use of impartiality in Kantian theories in fact leaves room for morality in "relationships of attachment between persons"[91] and whether it in fact "devalues the affective life – the life constituted by feeling, intimacy, connection."[92]

Herman acknowledges that Kant's ethics are considered

> the standard model of an impartial ethical system. Persons have moral standing in virtue of their rationality, and the morally dictated regard we are to have for one another reflects this deep sameness: we are never to fail to treat one another as agents with autonomous rational wills. This yields impartial treatment of persons and impartial judgment across cases.[93]

However, if relationships of attachment have "distinctive moral claims that impartiality disallows"[94] then there is a tension between these relationships and the rest of morality: "[t]o the extent ... that impartiality defines the moral perspective, partiality creates tension with and within morality."[95] Since each of these elements – attachments and morality – is essentially constitutive of the human person, it is extremely painful to contemplate them as being in principle potentially at war, and further, to believe that in actual cases of conflict, motives toward care must in principle give way to motives of impartiality. Herman wants to "accept that it is reasonable to expect a moral theory to give (noninstrumental) expression to the role that sociality and the partiality of connection play in a human life ... [and to] show that ... Kantian ethics does this."[96]

She surveys and discards as insufficient various means that previously had been tried to accommodate partiality of connection as a *moral* element. One possibility is to argue that motives of attachment can be compatible with Kantian ethics, so long as there is in addition a primary motive of duty regulating the agent's volition. But being compatible with Kantian ethics is not the same as establishing that motives of connection have their own moral character and can be moral motives in their own right. Another possibility is that by adopting a morally required end, one may act from motives of connection in order to achieve that end; the agent's "complete maxim then includes not only the motive of connection but also the underlying moral commitment to the required

end (from the motive of duty)."[97] But once again, the motives of attachment serve only instrumentally as a means to a morally required end but have not thereby gained moral status in their own right. A similar dead end is reached when attachments are seen as the means to securing and maintaining one's happiness, to which there is an indirect duty – once again the motives of attachment are a means to a morally required end but do not serve as moral motives in their own right.

The question may arise as to why it is deemed so important to establish that attachments have *moral* value, rather than some equally important but different kind of value. But in that case, relative value weights would have to be compared between the two kinds of value, and if moral value "trumps connection, the value of connection is diminished."[98] And the Kantian "cannot accept" the possibility that values of connection could in their turn outweigh the value of motives of duty.[99]

Last, the idea that Kantian morality is associated with rationality, while attachment is associated with feeling, is troubling:

> the assignment of (at best) subordinate value to the motives of connection supports the idea that our affective nature is not essential to our moral agency.[100]

The question is whether affect and feeling can or do incorporate rationality in any way, thus opening an avenue to consideration of motives of connection as having moral status.

Having surveyed and put aside various attempts to argue, from a Kantian perspective, that motives of attachment can be interpreted morally, Herman undertakes to shift the terms of the discussion by changing the perspective from which the question is asked. She asks not what style of impartiality is being used in Kantian deliberation, but

what role impartiality plays according to the way an agent "represents its place in deliberation"[101] and suggests that there are two distinct models of deliberation. "Failure to recognize [them] leads to serious distortion of the problem thought to be posed by impartial morality to concerns of trust and connection."[102]

In the first model, the "plural interest" model, the agent is conceptualized as confronting a "bundle" of various different kinds of interests which have been initially sorted by weight; deliberation amounts to a further judicious weighing, balancing, and adjustment among competing interests of similar weights. Impartial morality is the supreme interest of all of these, weighing most heavily because of its regulative capacity. Therefore, when morality is seen to be in conflict with one's attachments, there is a deep experience of tension:

> The problem arises when it looks like "over here" is what I care most about, what I want to happen (and cannot not want to happen), but "over there" is what impartial morality demands. ... And when impartial morality wins, it is not only at the expense of what I most care about, it provides no deliberative space even to acknowledge my concerns.[103]

There is an either/or quality to this model; in a conflict between my attachments and the requirements of morality, either I act morally but then perhaps suffer attachment losses, or I act on my affective interests but then must view myself outside of morality. I accept that on either scenario I suffer losses, because this is the price I must pay for my principled moral commitments on the one hand, or the important value of my attachments on the other. Herman suggests that if this is the view of impartial morality that is paradigmatic, it is no wonder that there are serious objections to it, and therefore the

tendency to want to throw impartiality out altogether. But Herman contends that there is another deliberative model which maintains impartiality intact, but does not imply the potential for such serious losses; she calls this model the "deliberative field" model.

On this model, the requirements of morality are not set against equally compelling but presumably non-moral interests, but are integrated fully into one's life as a whole:

> According to this ... model, the practical self does not have as its major task negotiating a settlement among independent competing claims. ... Deliberation addresses a field only partially shaped by those commitments, concerns, and relationships that determine my conception of the good. They stand there as myself – as interests that I need no further ... reason to care about. ... A human life is not the resultant of a 'bundle' of competing interests, (among which is an interest in morality). One's interests are present on a deliberative field that contains everything that gives one reasons.[104]

What this amounts to is an integrated, non-static complex characterized by varying interests, commitments, attachments, and principles in deep and mutually modifiable relationship to one another, all constitutive of the self that I am and "the good as I see it."[105] There is acknowledgement of unforeseen consequences of actions taken with even the most careful deliberation according to the clearest moral principles; this naturally can lead to modification of ends and means if one of these turns out in the event to be morally unacceptable. The Good is not to be conceptualized as a static bundle of objects of desire, but as an integrated complex capable of continuing transformation in view of circumstances and changing conceptions developed through experience of a life lived according to principles that are regulative in our lives.

The difference between the two models is conceptualized as a difference in

activity or passivity in relation to desires. The agent of the plural interest model is seen as passive in relation to the desires that make up his conception of the good:

> If we take the paradigmatic deliberative situation to be either means-end calculation or the resolution of conflict between ends, it will look as though our starting point is the pursuit of discrete goods (the objects of desires or interests) whose compatibility is a matter of luck. ... sometimes this is just the way things are ... but focusing on these cases reinforces a sense of our passivity as agents: what Kant meant, I believe, by a heteronomy of the will.[106]

On the other hand, the agent of the deliberative fields model is seen as active in relation to the desires that constitute his good, which is not viewed as a complex object of desire, but of practical agency. Ends are not adopted and abandoned in isolation from each other; either may alter what ends are already there and may affect future ends. Because of the "complexity of relations among ends, ... deliberation itself will then reshape or reconfigure the deliberative field."[107] Further, if deliberation is "structured by *substantive* regulative principles"[108] – such as treating people as ends – then this affects and alters the ends of friendship and other relations of attachment:

> It is not that I must replace motives of connection with moral motives; I will have *different* motives of connection. Perhaps I will be more sensitive to problems of exclusion or of fairness. Perhaps I will be less tempted to interfere 'for the best.'[109] [emphasis in text]

It is in the context of treating people as ends that the argument is focused most sharply on the situatedness of the agent and therefore on the role and function of impartiality in

this model. If treating people as ends is paradigmatic of Kantian impartiality, then it is vital to know what that in fact amounts to – how to actually treat people as ends – and that cannot be known unless impartiality takes the specific situatedness of both the deliberator and the other well into account:

> Without knowledge of how intimacy engages vulnerabilities, I cannot see that or how certain behaviors which could be acceptable among strangers are impermissibly manipulative among intimates (and vice versa). ... Absent such knowledge, moral judgment is not possible.[110]

Far from moral imperatives, such as treating people as ends, being potentially in conflict with relationships of attachment, they are instead fully integrated into those relationships, and necessarily so, or there would be no understanding of how to proceed on those very imperatives. The somewhat dry formality of (for example) the categorical imperative is such precisely because its practical application requires the acknowledgement of the infinite variety of situatedness of both agent and patient, and therefore of the role of impartiality, in order to apply it at all, and its application will necessarily look different in different circumstances, with different players. There is a "mutual practical dependence between formal moral principle ... and the structure of attachment."[111]

A further distinction between the two models of deliberation is the sense that in the plural interests model, the deliberative field was somehow "empty" prior to the arrival of agents on the scene. This sense of emptiness that is to be filled with competing agents' interests is reinforced by deliberative models such as Rawls' original position. In this model, agents come together, each with his or her own future potential interests, and

structures of fairness are utilized to negotiate among them in order to yield beginning principles of justice as fairness. As a schematic, it serves the purpose of isolating the actual deliberative mechanism and its motives so that it can be inspected and recognized; as a model for actual deliberation, even though Rawls indicates that an agent may enter this position at any time, it is unclear how it can be made practically effective. He has also distinguished the three kinds of agents: the "parties in the original position, ... the citizens in a well-ordered society, and ... you and [I] who are examining justice as fairness to serve as a basis for a conception that may yield a suitable understanding of freedom and equality,"[112] but it is unclear which perspective the ordinary deliberator is to assume and when.

In the deliberative field model, however, the ongoing difficulty of achieving a "true" impartiality is significantly diminished, but this is not because a way has been found to rid ourselves of our biases and predilections once and for all. On the contrary, the biases and predilections of individual persons, both as agents and patients, are in constant view, having been co-opted, as it were, in the very service of a genuine and practical Kantian impartiality, rather than viewed with suspicion as "Idols" or submerged in a self-conscious attempt to achieve impartiality. In the deliberative field model, impartiality can come about as a natural byproduct of the decision to adopt as one's own the essential recognition of Kantian morality, that as rational beings we are all of equal moral status. This recognition can become the touchstone of actual decision-making, sometimes requiring a formal impartiality or disinterestedness, as when, in adjudicating a playground dispute that my child is involved in, I see that it is she who must apologize for some transgression. Sometimes there is a recognition that equity issues must be addressed in which formal impartiality would be out of place; for example, when there are

extenuating circumstances surrounding the playground transgression. At other times a distinct partiality of care and concern is appropriate, as when it is my own child I send to college instead of my neighbor's poorer but equally deserving child.

All these examples show that in this model there need be "no tension between [relationships of] trust and the morality of impartiality";[113] Herman explains that a commitment to the equal moral status of all the parties in one's deliberative field requires that each be treated always as an end, the essence of Kantian impartiality; but what being treated as an end actually amounts to can and often does vary from situation to situation:

> From the fact that my child trusts that my concern for him will lead me to guard and preserve his well-being as I can, it does not follow that I violate his trust if I refrain from doing some things that will benefit him (because they are wrong or unfair) or if I act for someone else first, as when I tend to the younger child hurt in the playground, or expend finite resources on a needier sibling. What my son has reason to trust is that I am committed to his well-being ...[114]

At the root of all moral decision-making lies the equality of moral status of all parties, thereby offering the best chance to avoid the pernicious tendency toward exclusion justified by the dictates of other kinds of impartiality.

A Contemporary Utilitarian Theory: R.M. Hare

Earlier, we saw that sympathy theories of impartiality suffer from the problem that there seems to be little opportunity or motivation to overcome already-existing affiliations that would make sympathetic identification difficult to accomplish. A contemporary sympathetic utilitarian theory, however, has moved well beyond that

difficulty to make sympathy between persons unlike one another easier to accomplish. The theory identifies human critical reasoning with that of the ideal observer, and at the same time presents this kind of reasoning as an avenue to sympathy. R.M. Hare, in *Moral Thinking*, wants to show that for some words, their meanings are due entirely to their logical properties; if this is also so for deontic terms like "ought" and "must," then a moral calculus can be developed from analysis of the logical properties of those terms and some others, and therefore, one "ought to be able in principle to learn *all about* the canons for [moral] thinking"[115] or "rational thinking about moral questions."[116] A major element in the derivation of universal moral prescriptions using this moral calculus is the proper understanding and use of sympathy.

Sympathy enters into Hare's methodology when he proposes to derive substantive moral conclusions from both linguistic and substantive premises, using "canons of reasoning ... established by ... linguistic intuitions."[117] As far as substantive premises are concerned, this obviously raises the question of what sorts of substantive premises would serve in such a derivation (Hare stresses that it would not be a "deduction" as traditionally understood[118]). If they are factual in nature, they would present an "is/ought" problem; the derivation would then seem to violate Hume's Law. But if they were prescriptive in nature, then there is no way to account for where they came from; there would have been no advance made over intuitionism, a major target of Hare's theory. Hare aims to show that the resolution of this problem lies in the logical and moral requirement to "universalize our prescriptions ... which amounts to treating other people's prescriptions as if they were our own"[119] – in short, through a methodology of sympathy:

> So it is not that we are going by logic from facts to prescriptions; it is rather that logic compels us, having ascertained *facts about what others are prescribing* or will prescribe, to treat these prescriptions on a par with our own original prescriptions in our moral reasoning.[120] [emphasis added]

The method for doing this involves several things: the capability to ascertain the relevant features of a situation where a moral decision is required; a recognition that what the effects on others of our moral decisions might be *is* such a relevant feature; and therefore a capacity to enter into a sophisticated process of identification with those possible experiences of others, in order to test our potential preferences in those situations against what our current preferences are.

There is required, in addition, a recognition that moral thinking takes place on two entirely different levels. First is the level of *prima facie* principles, which is often the repository of our received opinions, our upbringing, and therefore of our intuitive and generally unreflective moral responses, principles, and dispositions – working principles for practical use in everyday life – but that may also be arrived at by the process of critical thinking, the other level of moral thought. This is the same kind of reasoning as that performed by an ideal observer, or an "archangel," in Hare's terms. Critical thinking "proceeds in accordance with canons established by philosophical logic and thus based on linguistic intuitions only."[121]

> Critical thinking consists in making a choice under the constraints imposed by the logical properties of the moral concepts and by the non-moral facts, and by nothing else.[122]

All archangels reasoning in this way will come out the same "at the end of their critical

thinking."[123] Philosophical error in practical reasoning or theorizing usually stems from a failure to distinguish between these two levels of reasoning about moral issues, whereas a correct appreciation of this distinction, combined with proper sympathetic identification, will yield universalizable moral principles in accordance with canons of moral reasoning generated by the logic of moral concepts. These moral principles will be – must be – expressed in prescriptions that are utilitarian in nature,[124] because they are generated by reducing all interpersonal conflicts of desires to intrapersonal ones, thereby allowing each individual reasoning morally to compare correctly the relative internal desires and preferences and decide upon an action based on the stronger desire.

This point is brought out in Hare's discussion of how critical thinking can generate a correct understanding of the mechanism of sympathy in order to yield universalizable prescriptions. Essential to the enterprise is the recognition that moral thinking must be done in light of the facts; Hare is clear that answers to moral questions are not only prescriptive and evaluative, but also descriptive,[125] in the sense that there are facts about actions or people concerning their properties of goodness or badness that must be taken into account in moral judgement.

> In making moral judgements we are purporting to commend or condemn actions or people because they have some properties which make them right or wrong, good or bad; and therefore it would be obviously irrational to make the judgements without ascertaining whether or not they in fact had the properties.[126]

In light of this requirement, and since we are not archangels and therefore cannot know all the relevant facts, we must be able to determine within the scope of our capabilities which facts are indeed relevant to the decision required – "[w]e need, that is to say, to make

judgements of relevance."[127] Hare describes a method of arriving at such judgements that is very similar to the methodology used by Rawls in his discussion of reflective equilibrium. His definition of a morally relevant feature of a situation is that a moral principle may be applied to that situation which mentions the feature.[128] But at the start of moral thinking we do not yet have any moral principles, so guesswork is needed:

> If we think that a feature of a situation might be relevant, we experiment with principles mentioning the feature; to accept the principles will be to accept the relevance of the feature, and to reject them will be to reject *those* reasons why it might be relevant, though there may yet be *other* principles which make it relevant.[129]

A particular class of such features is "the likely effects of possible actions on people"[130] including both others and ourselves. This is established by trying to see "what it is like to be those people in that situation,"[131] thus introducing the discussion of sympathy.

Correct sympathy depends initially upon understanding the relationship between the cognitive and affective states involved in a particular situation, for example, one that can potentially cause suffering to someone. It is a conceptual truth, according to Hare, that if one suffers, one must also know that one is suffering, and vice-versa. Added to that is a particular conative state, in that when one suffers, one necessarily has a motive for ending the suffering. If any of these elements is missing, then there would be no suffering, as a matter of conceptual necessity.

The difficulty now lies in trying to determine what it is like for someone else to be suffering. In the ensuing discussion, Hare picks up the puzzles about "who" is identifying with whom in the practice of sympathy:

> Can I properly be said to know what it is like for him (not just to know that his neck is being broken), unless I myself have an equal aversion to having that done to me, were I in his position with his preferences?[132]

Here, he distinguishes two conceptualizations of the experiencing "I" to help resolve the myriad problems that arise from contemplation of whether we can "know" what anything is "like" for another. On one level is the idea "that 'I' is not wholly a descriptive word but in part prescriptive," which means that "in identifying myself with some person ..., I identify with his prescriptions."[133] This means that I would take on the preferences of that other person as if they were my own, which in the hypothetical case of identification they would be. "Insofar as I think it will be myself, I now have in anticipation the same aversion [to some proposed suffering] as I think he will have."[134]

This in turn enables universalization:

> The effect of the argument ... is to facilitate one step in [the] progression [in the use of the property of universalizability], namely the step from prescriptions which I accept for my own experiences to prescriptions which I must accept for experiences I should have, were I to be in someone else's position with his preferences.[135]

It is here that the argument concerning interpersonal conflicts of preference being converted to intrapersonal conflicts is constructed. His example is of two persons, one a cyclist parking his bicycle where someone else wants to park his car. In thinking about this situation, I (the car-driver) fully recognize the cyclist's desire to leave his bike where it is; I even acknowledge that were it my bike, I might also have a preference to leave it

where it is. But in that case I would also have to acknowledge that the relative inconvenience of moving the bike as against the inconvenience of not being able to park the car is much slighter, and I would opt to move the bike.

> ...if I have full knowledge of the other person's preferences, I shall myself have acquired preferences equal to his regarding what should be done to me were I in his situation; and these are the preferences which are now conflicting with my original prescription [to leave the bike where it is]. So we have in effect not an interpersonal conflict of preferences or prescriptions, but an intrapersonal one; both the conflicting preferences are mine. I shall therefore deal with the conflict in exactly the same way as with that between two original preferences of my own.[136]

Conflicting preferences would play out in just the same way were I to mentally reverse the roles played by the biker and the driver; since the universal properties of the situations remain the same, while only the individuals differ in their respective roles, I can still make the same decision that, all things considered, moving the bike is the thing to do. This is how "the requirement to universalize our prescriptions generates utilitarianism."[137]

The major problem here, as in Smith, is the epistemological puzzle of how I can come to know another person's experiences well enough to be able to identify with them; here is where Hare brings in the other conceptualization of the experiencing "I," using Zeno Vendler's ideas from his essay "A Note to the Paralogisms." In that essay, Vendler focuses on what he calls the "transcendental 'I'" – that "I" which has "no content and no essence; ... a mere frame in which any picture fits; ... the bare form of consciousness."[138] As such, when "I" transfer my consciousness in imagination to the person of another, I transfer only this non-essential framework of consciousness, not the essential qualities that make me who I am in the world, seen from outside – my "empirical self" – a person

with a certain name and a certain specific situatedness in the world. To the objection that this transference is not possible, Vendler notes gently that were it not possible, it would also not be possible to read history or literature; one would simply not understand what is being said without the capacity to identify on this transcendental plane with the players and characters. And it is not that there are two components to the self, the empirical, essential self and the transcendental "I"; the latter does not exist as a "thing":

> ...existence, like possibility, is a 'category' operating in the field of experience. The transcendental aspect of my being consists in nothing else but in the realization that I, as a subject of experience, am only contingently tied to the senses of this body, that is to say, that the world could be experienced through other eyes, perceived in other perspectives, in one word, it consists in my ability to perform feats of transference.[139]

Therefore, I can easily imagine being someone else and I can correctly imagine the experiences that person might have:

> ... the 'I' itself, i.e. the mere form of consciousness, is 'mine' only according to the particular content it has, and since it has that content only contingently, I indeed could be that man, and thus ... could feel the very same thing he feels. ... Thus wide range of experience, and the knowledge of physiology and psychology enables one to compensate for the difference found among us in bodily state and personal history. Once the possibility of transference is recognized, the sharing of human experience becomes a manageable task.[140]

With the power to correctly make someone else's prescriptions one's own, one is now in a position to universalize the prescriptions developed after correct application and

understanding of sympathy. Universalization depends upon the capacity to make interpersonal comparisons of strengths of desires (rather than summing units of pleasure), and these are reducible to intrapersonal comparisons after correct application of sympathy:

> ...insofar as I fully represent to myself the strengths of other people's preferences, I have preferences, myself now, regarding what should happen to me were I in their positions with their preferences. ... I have to disregard entirely my own antecedent preferences; the preferences which I form have to consist entirely of replicas of the other people's preferences.[141]

The universalization will play out along utilitarian lines, since we are weighing preferences and each person, including oneself, is to count for one and none for more than one:

> We have to treat everybody as one, including ourselves: to do unto others *as* we wish they should to to us ..., and love our neighbours *as* (not more than) ourselves.[142] [emphasis in text]

The question now is whether Hare's sympathetic utilitarianism has succeeded in overcoming the problems of bias and exclusion that plagued previous sympathy theories. The tensions, as always, are between the theoretical and the practical, the ideal and the actual. Impartiality is certainly built into the constraint of utilitarianism that each is to count as one and none as more than one, which makes utilitarianism one of the fairest theoretical moral methodologies; the critical thinking of the would-be archangel is meant to assure that this indeed is what happens. But as Hare himself points out, our human situation is dangerous because of "our lack of information and our proneness to self-

deception."[143] Also, the theoretical recognition of the distinction between the empirical and transcendental self is meant to assure that we may walk, in imagination, in another's footsteps and correctly feel what they feel, making their experiences and prescriptions our own and comparing intrapersonally the strengths or degrees of our respective preferences in the situation; but to what extent is there a risk of our empirical self stowing away on this transcendental transfer of consciousness and making its own personal preferences key in moral decision-making?

The defense of utilitarianism presented by Hare depends on the recognition of yet a further distinction: that between far-fetched hypothetical cases which may require, according to opponents of utilitarian theory, counterintuitive outcomes (this being presumably a definitive objection against a utilitarian moral perspective), and events and situations in the real world, situated either actually or hypothetically within the range of ordinary human experience (that toward which both *prima facie* principles and critical thinking are, and should be, oriented). Given that the situations under consideration will be of the latter kind, is there now warrant for saying that this particular version of a sympathy theory of impartiality will have a better chance to yield a moral community that is inclusive, rather than exclusionary in the same ways as previously? The answer is a guarded "yes," and for similar reasons that were brought out for Kantian theories.[144] What can save this particular sympathy theory from falling into exclusionary practices is the previously mentioned principle, attributed to Bentham by Mill, that each is to count for one and none for more than one,[145] combined with the requirement to access archangel-like critical reasoning in times of conflict of duties or other troubles that intuitive or *prima facie* principles cannot handle on their own. This fundamental principle of equality, to be brought to bear during sympathetic transferences and also serving as a motivation for

them, serves the same purpose as Kant's ideas about respect – it reminds us that *everyone* is our neighbor, not just some people whom we happen already to approve of. As seen earlier, the prior disapproval of certain persons is a danger in sympathetic identification, because the alienation or disaffection that may be experienced naturally when someone is very different from oneself may block motivation for sympathetic transfer to begin with; if sympathetic transfer proceeds nevertheless, the disaffection may block a full transfer, leaving room for the preferences of the existing "empirical self" to have greater valence than those of the target of the transfer. These in turn may not come as fully to view as they might have, had the empirical self not been permitted so much access.

In this way, the moral equality of persons is assured in much the same way as in Hobbes – persons are equal to begin with and they remain so without further qualification of who is to count and who is not. Hare emphasizes this moral equality of persons by hypothesizing, in his discussion of rights, that there is really only one right that is not confined to the intuitive or *prima facie* level of thinking but emanates from the critical level: the "right to equal concern and respect," which for Hare is "nothing but a restatement of the requirement that moral principles be universalizable"[146] and is therefore an expression of Bentham's principle.

But Hare's theory does not depend merely upon the recognition of the moral equality of persons combined with a relatively clean and workable methodology of sympathetic identification; it also aims to develop workable and universalizable moral principles derived from certain kinds of premises, based on a methodology of moral logic. This part of his methodology threatens to wipe out the power of the impartiality of his sympathy theory, since it opens the door for the introduction of the same hidden biases

that have threatened other impartiality methodologies, both sympathetic and detached.

A major objection that Hare has against intuitionists, and against Rawls in particular[147] (although Rawls explicitly places his theory in opposition to intuitionist theory[148]) is that according to him, they make the truth of their arguments depend upon agreement with other opinions.[149] For Hare, intuitionism is a "form of disguised subjectivism," and he notes that he himself has been " ... often falsely accused of this sort of subjectivism."[150] He sees a major difference between him and Rawls as Rawls' dependence on consensus and received opinion about moral intuitions at every step of his argument, whereas he, Hare, begins with the logic of moral concepts as checked against the *linguistic* intuitions of native speakers. He is very concerned not to have the enterprise of moral argument and theorizing deteriorate from the level of philosophy to that of anthropology[151] and he sees Rawls' reliance on moral consensus as a starting point as just such a slide downward. The questions to be answered here are whether the distinction between the two methodologies is as clear as Hare makes it, whether the answer to this question can impact upon the conceptualization of sympathetic impartiality that Hare develops, and then whether this would influence the major question of inclusion/exclusion.

In the discussion of Hare's methodology above, we began with his response to possible difficulties concerning the substantive premises he means to pair with linguistic premises to derive substantive moral conclusions from "canons of moral reasoning established by linguistic intuitions."[152] Turning now to the other part of the practical argument, the linguistic premises and intuitions, it will be seen that here is where the greatest impact on impartiality will be found.

Hare's purpose is to generate substantive moral principles on the basis of analysis

of the meaning of moral terms – specifically, to ascertain whether the logical properties of moral terms exhaust their meaning or whether something else is needed to understand them:

> ... [T]he first step towards answering a question rationally is to understand it ..., which entails understanding the words in which it is posed. The reward for learning about their meanings or uses is that we are at the same time learning something of the canons for thinking logically about questions containing them. For they, like all other words, owe their meanings partly or wholly to their logical properties.[153]

His examples are "all" and "some" – for these words, "their logical properties exhaust their meaning."[154] For other words, such as "blue" and "red," "something other than their logical properties" is needed to establish their meaning – "no amount of logic will show us the difference between blue and red."[155] Hare contends that the deontic terms "ought" and "must" are more like "all" and "some" than like "blue" and "red." Subsequently, various other moral terms are taken up by Hare and examined for their meanings and logical properties.

Hare uses a hypothetico-deductive procedure in which the hypotheses concern "what people do mean when they use ... moral words."[156] In one use of this kind of procedure both the hypotheses and the data against which they are checked are linguistic – native speakers can recognize whether "certain locutions are deviant or non-deviant."[157] In another use, the hypotheses are also linguistic, but the data against which they are checked are morally relevant facts about human desires and aversions. If the linguistic hypotheses are about moral words about desires and aversions, and the check turns up

facts about these desires and aversions that were predicted by the use of these words, then certain substantive predictions concerning moral reasoning by humans may appropriately be made. In this way the original hypotheses have "escaped refutation."[158] Moral opinions, therefore, while not proved correct, are nevertheless accounted for by explanation.

Hare notes and responds to several objections that may be made to such a procedure, among them the question of whether this "linguistic method" would "make us the slaves of our existing conceptual scheme." But Hare responds that his

> strategy has been to expose the logic of the moral concepts as we have them, and show that they generate certain canons of moral reasoning which will lead to our adopting a certain method of substantial normative moral thinking.[159]

He then invites anyone who disagrees to set out the logic of an artificial language and proceed from there; his contention is that the difference in results between the two would not be great. The meanings of our moral terms apparently are what they are, regardless of the language in which they are expressed.

There are two problems with the way Hare has set out his theory. The first problem is that he does not always make clear what is to count as a moral concept suitable for such neutral conceptual analysis and what is not. The deontic terms "ought" and "must" that he identifies are clearly suitable in this way; but other terms he uses can be ambiguous – for example, "suffering." Linguistic intuitions may accurately account for the correct use of the former but analysis of the latter may need the "something more" he identifies for correct understanding of "blue" and "red."

This introduces the more serious problem that linguistic intuitions are not always

morally neutral, as he acknowledges; they are often fueled by existing *moral* intuitions, and because of this, Hare's methodology in fact does come close to the reliance on moral consensus that he decries in Rawls, particularly if the class of moral concepts is widened to include terms beyond the purely deontic ones mentioned above. Because of this, the charge of subjectivism leveled against him turns out, in fact, to be not far off the mark.

He does acknowledge that the use of certain terms in our moral language carries this sort of moral weight – he calls them "secondarily evaluative words, whose very use, in their full evaluative sense, commits us to substantial moral evaluations,"[160] for example, "lazy."

> It is possible, by concentrating on such words, to create the impression that our conceptual scheme, and the very meanings of our words ... commit us to the adoption of certain norms of conduct. The view is thus easily propagated that all linguistic intuitions concerning any moral words incorporate moral intuitions, or that no distinction can be sustained between the two kinds of intuition.[161]

But by not clarifying the class of moral terms that would lend themselves to the development of a moral logic, and by not distinguishing them clearly from the class of the so-called "secondarily evaluative" terms that would not so lend themselves, because of their imported moral commitments, Hare opens himself up to exactly this problem and this kind of misunderstanding of his theory. The problem is exemplified by his conceptual choices concerning two moral terms, "suffering" and "liberty." He treats the former, apparently, as one that bears no moral intuitions and can be analyzed neutrally, thus having a formative role in the development of a moral logic, but treats the latter as apparently "secondarily evaluative," without providing guidance as to how each of these

terms fell into the class it did and how to tell which class other moral terms may fall into. An examination of his conceptual decisions concerning these terms will make the point.

The first example is the above-mentioned "suffering." Hare begins his discussion by once again directing us to look for the facts of a situation before embarking on deliberation:

> We are required to ascertain the facts before making factual statements ... [e]ven if moral judgements cannot be called truth-claims without qualification, they are subject to a similar requirement to ascertain the facts before pronouncing morally upon them. It is the function of moral principles to provide universal guidance for actions in all situations of a certain kind, and one of the most important functions of singular moral judgments is to make clear what our principles are. All this would come to nothing if our moral judgements were unrelated to the facts about the situations on which we were commenting.[162]

He establishes the conceptual truths about "suffering" – conceptual necessity requires that if one suffers, one *know* that one is suffering and one be *motivated* to end or escape from the suffering; if either of these elements is missing, there is and can be no suffering. He then immediately notices a difficulty about this conceptualization of suffering, namely "that a being, if there were such, ... who lacked *self*-consciousness might suffer without knowing that it was he who suffered."[163] He further admonishes the reader not to confuse the concept of suffering with the concept of pain, because "in unusual cases it is possible to have pain without suffering, and without having a motive for ending or avoiding the pain, even *ceteris paribus*."[164]

Hare takes care of the first problem by ignoring it: "To avoid this difficulty, let us confine our attention to beings who have self-consciousness."[165] Without disputing the

correctness of this conceptualization of suffering, it must be noted right away that this is not a neutral linguistic analysis that can be a building block for a universal moral logic; neither is the associated conceptual decision to simply ignore a subclass of potential sufferers. Both together import into the analysis the requirement that the suffering individual must be one who knows he is suffering and ignore those who are not self-aware in this way. But it is clear that, given his analysis, the absence of self-awareness, combined with the admonishment not to confuse suffering with pain, has opened the door by conceptual fiat to the derivation of substantive moral principles that would support inflicting pain upon non-self-aware humans, such as fetuses, anencephalic infants, arguably even normal infants, or persons in a persistent vegetative state, not to mention animals, with the justification that they would not be – *could* not be – experiencing suffering. This is hardly neutral, and it is far from factual that these human and animal beings do not experience suffering; it is a matter open for interpretation, and interpretation would likely proceed on a basis consistent with the other moral commitments a native speaker may hold. Moreover, to claim that the ordinary linguistic intuitions of the majority of persons would support such an analysis is simply incorrect (otherwise the dispute about the morality of abortion would be able more easily to leave the question of the suffering of the fetus aside).[166]

Another example is the concept of "liberty" Hare mentions in his discussion of rights and justice. His discussion is brief; it begins by noting that

> [i]t might have been contingently the case that societies could prosper under tyranny and slavery. ... I firmly believe that as a matter of fact no likely society is going to be better off with a system allowing slavery; and I also firmly believe that all societies which are capable of operating a democratic system (an important

> qualification) would be well advised in their own interests to adopt one. But my reasons for these judgements are beliefs about contingent matters of fact. If these were shown to be false, then the same philosophical views about the nature of the moral argument involved might make me advocate slavery and tyranny.[167]

Here, Hare makes the again non-neutral philosophical decision that "liberty" is *not* to be made an object of conceptual analysis in its own right, in the way that "suffering" was. Liberty and slavery are instead to be viewed through the lens of judgements stemming from "beliefs about contingent matters of fact."[168] "It is not a priori true or self-evident" that "liberty ... is an overriding principle."[169] But *a priori* or self-evident truth is a different matter from deciding whether a moral concept belongs in a set of terms that will be used to develop a moral logic or is merely "secondarily evaluative," as Hare seems to treat this one. This once again is a philosophical decision made by one philosopher; others have made precisely this concept of liberty the object of fundamental conceptual analysis upon which *subsequently* to build theories about just societies, not to allow the existing facts on the ground to determine whether liberty or tyranny is the way to go.[170] Again, this is not to dispute whether it is correct to argue morally in one way or the other, but to point out that such procedural decisions are not in themselves morally neutral. Conceptual analysis of liberty as, say, the *sine qua non* of human personhood, and a correlative idea that societies ought to provide the groundwork for the flourishing of human personhood, may yield a far different understanding of the "justice" of a just society than starting with facts on the ground and then making empirical judgements about whether slavery or liberty is the better route.

A related problem refers to theoretical implications he does not discuss, namely, that this methodology does not provide safeguards against the possibility (even

likelihood) that it can be used to support an entire language and practice of prejudice. Perhaps this is not a requirement of moral theorizing, that it should be safe against misuse; but this loophole does incidentally also highlight the extreme subjectivity of the moral prescriptions derivable by this method. The example I will use is "normalcy." Normalcy can be analyzed as a statistical concept, where minority groups that fall outside a large statistical majority in some set of practices or beliefs can be seen as "deviant." But it is to be noted that there is already heavy evaluative baggage accompanying the terms "minority" and "deviant," even if one intends to use them strictly neutrally, as statistical descriptions; there is also a substantial moral commitment involved in the prior decision to treat a certain set of practices and beliefs as something to be "normed" in the first place. When analysis of normalcy is shifted from straightforward statistics to linguistic intuitions of native speakers, the door to imported moral commitments is opened even wider. A conceptual analysis of normalcy in, say, sexual preferences and practices, based on what can be presumed to be the ordinary linguistic intuitions of native speakers concerning human sexual normalcy and deviancy, in the not-too-distant past yielded the conceptualization – by scientifically trained members of the medical community – that homosexuality was not only deviant, but actually pathological. Homosexuality was listed in the Diagnostic and Statistical Manual of the American Psychiatric Association (DSM) as a psychiatric pathology, illustrating how linguistic intuitions can infiltrate even areas reliant on a presumably objective scientific methodology. It has subsequently dropped out of the DSM due to a shift in linguistic intuitions based on changing facts, such as increased physiological understanding by the medical community and systematic political consciousness-raising by the homosexual community.

While Hare acknowledges that linguistic intuitions are open to revision upon

careful analysis in light of facts about humans and human reasoning (see below), it is confusing to simultaneously claim, as he does more than once, that careful use of the canons of reasoning based on these kinds of intuitions will yield the same conclusions from everyone – approaching, as he puts it, the unflawed reasoning of the ideal observer. But once again, this example shows that linguistic intuitions can be the bearers of heavy moral implications, and are not as neutral as they would have to be in order to develop such a reliable moral "logic" from them. It is certainly true that substantive moral prescriptions can be derived from them. Hare's problem here, however, is the same as he points up in Rawls: who is to say what the right thing is to do, in distinction from what "everyone" *thinks,* in the form of their linguistic intuitions, is the right thing to do? This, of course, is a central question in ethical theory; the virtue of Rawls is that he acknowledges up front the place of community consensus as an inescapable part of moral theory-building within that community and develops a conceptualization of objectivity to work as well as possible within that framework.[171]

Hare responds to criticisms that have been made against his theory on a somewhat different ground:

> I still often meet philosophers who think that it is an argument against my views to show that what is right or wrong cannot depend on what somebody prescribes. ... For if it did so depend, then we should be able to derive the statement 'It is wrong' from a factual statement about the utterances or perhaps the thoughts of some prescriber; and I have repeatedly made clear that I do not think that such a statement can be derived from *any* statement of fact.[172]

Hare's theory, however, is not so crude, as shown above. It is not facts about the

"utterances or thoughts of some prescriber" that serve as the link between factual statements and universalizable prescriptive conclusions;[173] it is much more subtle than that. It is the linguistic intuitions that serve as the link between the facts on the ground and the prescriptive conclusions; the moral intuitions that he is opposed to can in fact be unconsciously laundered through these linguistic intuitions and that is what gives his theory a problem similar to that of Rawls and other intuitionists. Hare does acknowledge the possibility that people may have fallen into fallacies in their reasoning which can lead to certain linguistic intuitions having to be revised upon reflection, but immediately rejects that possibility:

> [W]e do not have ... to assume ... that the moral opinions of ordinary people are correct; we have to assume only that they are the natural outcomes of the fact that people are as they are, on the hypothesis that words have certain meanings. It might even be that there were various fallacies in reasoning into which it was easy to fall. ... It has been my aim, without any appeal to moral intuitions as certificates of the correctness of moral judgements, to find out what people do mean by the moral words. Having done this, we should be at liberty to recommend the adoption of a different use; but I shall not be doing this, because all the work I have done in moral philosophy ... has convinced me that our ordinary moral concepts, once their use is clarified, are serviceable.[174]

To be fair to Hare's critics, his theory leaves at best a lingering confusion and unclarity about what, exactly, Hare is saying, and this confusion facilitates the interpretation of him as a subjectivist. On the one hand, moral logic developed from conceptual analysis of moral terms approaches the consistency and universality of those of the ideal observer; on the other, given that we are only human, it could in principle be

subject to revision, but this is not likely to be necessary. Again, on the one hand, analysis of moral terms will yield a logic with which to make universalizable moral prescriptions; on the other hand, it will yield only non-refuted hypotheses about the meanings of words and accurate predictions about how people will behave, given those meanings.

The further problem of how to classify moral terms adds to the confusion and opens the door for the importation of moral intuitions and their associated prescriptions, and it is this that ultimately undoes most of the good done by the egalitarian sympathy methodology he so carefully delineates. It remains to be assessed empirically whether the admonishment to treat each as one and none as more than one can stand up to the prejudices and exclusionary biases that can so easily come in through the door that Hare has left open.

In the final chapter I will return to some of the dimensions within which I have explored the concept of impartiality and use these same dimensions to show how the dreaded outcome of the use of unexamined impartiality, exclusion, actually affects not only the persons who are excluded but the very fabric of community itself.

NOTES

Chapter Seven

[1] In *Whose Justice? Which Rationality?*

[2] Ibid., 4.

[3] Ibid., 12.

[4] Ibid., 332 - 333.

[5] Ibid., 367.

[6] Ibid., 355.

[7] This term first came into common use with Carol Gilligan, *In a Different Voice* (Cambridge, MA: Harvard University Press, 1982). Nel Noddings developed an ethic of care in *Caring: A Feminine Approach to Ethics and Moral Education* (Berkely, CA: University of California Press, 1984).

[8] Gilligan, *Voice*, 62-3. Also Gilligan, "Moral Orientation and Moral Development," in *Ethics: Classical Western Texts in Feminist and Multicultural Perspectives*, ed. James P. Sterba (NY: Oxford University Press, 2000), 553.

[9] Gilligan, "Moral Orientation," 553-556.

[10] Barbara Herman, "Agency, Attachment,and Difference," *Ethics* 101 (July, 1991): 775-797.

[11] Ibid., 5.

[12] Ibid.

[13] Ibid.

[14] "... anthropology observes the actual behavior of human beings and formulates the practical and subjective rules which that behavior obeys, whereas moral philosophy alone seeks to formulate rules of right conduct ... what ought to happen." (Immanuel Kant, *Lectures on Ethics*, trans. Louis Infield (NY: Harper Torchbooks, 1963 [1775-1780]), 2.

[15] Kant, *Foundations*, 8-9.

[16] Ibid., 9.

[17] Ibid., 16.

[18] Ibid.

[19] Ibid.

[20] Ibid., 18.

[21] Ibid., 24.

[21] Ibid., 25.

[22] Ibid.

[23] Ibid., 29.

[24] Ibid., 30.

[25] Ibid., 30-31.

[26] Ibid., 33.

[27] Ibid., 34.

[28] Ibid., 39.

[29] Ibid.

[30] Ibid., 46.

[31] Ibid.

[32] Ibid., 47.

[33] Ibid.

[34] Ibid., 49.

[35] Ibid.

[36] Ibid., 50.

[37] Ibid., 51.

[38] Ibid., 63.

[39] Ibid., 64.

[40] Ibid., 65.

[41] Ibid.

[42] Ibid., 66.

[43] Ibid.

[44] Ibid., 67.

[45] Ibid.

[46] Ibid., 69.

[47] Ibid., 70.

[48] Ibid., 71.

[49] Ibid.

[50] Ibid., 72.

[51] Ibid., 78-79.

[52] Ibid., 79.

[53] Lewis White Beck, *A Commentary on Kant's "Critique of Practical Reason"* (Chicago, IL: University of Chicago Press, 1960), 216.

[54] Immanuel Kant, *Critique of Practical Reason*, trans. Lewis White Beck (NY: MacMillan Publishing Company, 1956 [1788]), 82.

[55] Kant, *Foundations*, 79 n.4.

[56] Kant, *Critique of Practical Reason*, 21.

[57] Beck, *Commentary*, 216.

[58] Kant, *Critique of Practical Reason*, 82.

[59] Ibid.

[60] Ibid.

[61] Beck, *Commentary*, 219.

[62] Ibid., 83.

[63] Ibid.

[64] Ibid.

[65] Ibid., 83-84.

[66] Stark, "Decision Procedures," 491.

[67] Ibid., 84.

[68] Ibid., 76.

[69] Ibid., 77

[70] Rawls, *Theory of Justice*, 251.

[71] Ibid.

[72] John Rawls, "Kantian Constructivism in Moral Theory", *Journal of Philosophy* 77 (1980): 515-572.

[73] Rawls, *Theory of Justice*, 252.

[74] Ibid.

[75] Ibid., 255.

[76] Ibid.

[77] Ibid., 253.

[78] Ibid., 256.

[79] Rawls, "Kantian Constructivism," 250.

[80] Ibid., 248.

[81] Ibid., 259.

[82] Ibid., 249.

[83] Ibid.

[84] Ibid., 250.

[85] Ibid.

[86] Ibid., 251.

[87] Ibid., 251-2.

[88] Rawls, "Kantian Constructivism," 261.

[89] See Note 9.

[90] She lists Bernard Williams and Carol Gilligan, among others, as being in the "vanguard of this complaint." (775)

[91] Ibid.

[92] Ibid., 777.

[93] Ibid., 775.

[94] Ibid., 776.

[95] Ibid.

[96] Ibid., 779.

[97] Ibid., 778.

[98] Ibid.

[99] Ibid.

[100] Ibid.

[101] Ibid., 782.

[102] Ibid.

[103] Ibid., 783.

[104] Ibid., 784.

[105] Ibid.

[106] Ibid., 786.

[107] Ibid., 787.

[108] Ibid.

[109] Ibid.

[110] Ibid., 788.

[111] Ibid.

[112] Ibid.

[113] Ibid., 782.

[114] Ibid.

[115] R.M. Hare, *Moral Thinking – Its Levels, Method, and Point* (Oxford: Clarendon Press, 1981), 3.

[116] Ibid., 4.

[117] Ibid., 16.

[118] Ibid.

[119] Ibid., 16-17.

[120] Ibid.

[121] Ibid., 40.

[122] Ibid.

[123] Ibid., 46.

[124] Ibid., 111.

[125] It should be clarified here that Hare is emphatically not a "descriptivist" in the accepted sense, but he does recognize the element of "descriptive meaning" in moral conventions, which he says are very much like linguistic conventions in the criteria of their application, "except that to observe them is to adopt substantial moral opinions, which adopting a merely linguistic convention would not be." (70)

[126] Ibid., 88.

[127] Ibid., 89.

[128] Ibid., 63, 89.

[129] Ibid., 89-90.

[130] Ibid., 90.

[131] Ibid., 92.

[132] Ibid., 94.

[133] Ibid., 96

[134] Ibid., 97. Hare acknowledges that he is taking the direct, short route from one's own preferences in the present moment to hypothetical cases where one is placed into someone else's position. The longer route has traditionally involved lengthy speculations about one's present preferences regarding one's future experiences, and from there to one's present preferences regarding someone else's future experiences. Hare discards that route and discusses various problems that arise in the longer journey to identification.

[135] Ibid., 108.

[136] Ibid., 110.

[137] Ibid., 111. He notes that other philosophers have made use of this concept of "extended sympathy." (128).

[138] Zeno Vendler, "A Note to the Paralogisms," in *Contemporary Aspects of Philosophy*, ed. Gilbert Ryle (Stocksfield: Oriel Press, 1976), 117.

[139] Ibid., 118.

[140] Ibid., 119-120.

[141] Hare, *Moral Thinking*, 128.

[142] Ibid., 129.

[143] Ibid., 147.

[144] As Hare points out, "the understanding [of the formal, logical properties of moral words] which we owe above all to Kant, yield[s] a system of moral reasoning whose conclusions have a content identical with that of a certain kind of utilitarianism." (4).

[145] J.S. Mill, *Utilitarianism* (London: J.M. Dent & Sons, Ltd., 1947 [1861]), 58.

[146] Ibid., 154.

[147] R.M. Hare, "Rawls' Theory of Justice," in Norman Daniels, ed., *Reading Rawls* (Stanford, CA: Stanford University Press, 1989 [1975]).

[148] See, e.g., Rawls' "Kantian Constructivism in Moral Theory."

[149] Hare, "Rawls' Theory of Justice," 83.

[150] Ibid.

[151] Ibid., 86.

[152] Hare, *Moral Thinking*, 16.

[153] Ibid., 2.

[154] Ibid., 3.

[155] Ibid.

[156] Ibid., 13.

[157] Ibid.

[158] Ibid., 14.

[159] Ibid., 20.

[160] Ibid., 17.

[161] Ibid.

[162] Ibid., 88.

[163] Ibid., 92.

[164] Ibid., 93.

[165] Ibid., 92.

[166] It might be worth exploring whether the fact that this particular conceptualization of "suffering" serves such a useful and simplifying purpose in certain debates could account for a good part of the motive driving this supposed "linguistic intuition."

[167] Ibid., 167.

[168] Ibid.

[169] Ibid., 166.

[170] For a small sample: Kant, Mill, and Rousseau, not to mention Rawls.

[171] See Rawls, "Kantian Constructivism in Moral Theory."

[172] Hare, *Moral Thinking*, 208-9.

[173] Although, once again, a casual reader may be justifiably confused by Hare's linking of moral words with people's desires and aversions in his discussion of how to test whether moral words have the meanings they have. (14)

[174] Ibid., 14, 15.

CHAPTER VIII

SYNTHESIS: A MEDITATION ON EXCLUSION

The elements of impartiality that are generally agreed upon are (1) treating like cases alike and recognizing the requirement to give reasons if under some circumstance like cases are treated unlike; (2) broadly speaking, the quality of disinterestedness, however it is played out in any given modality of impartiality; and (3) the recognition of the equal moral status of all persons as fundamental to morality. The first element has not been the focus of concern here. The second and third elements have emerged as deeply intertwined, with some evidence that there is a causal relationship between them, in that when disinterestedness fails, recognition of the equal moral status of persons is threatened as well. This then leads to the specter of exclusion from consideration of membership in the moral community – a depersonalization – which then can then play out, at a minimum, as exclusion from sympathetic consideration as an individual or from the distributive benefits of an otherwise just society. Specifically, hidden partialities can compromise the trustworthiness of disinterestedness; the major problem has been that many types of partialities are deeply internalized and can become even more submerged as one is the more aware of striving conscientiously toward an impartial state of mind. Part of the struggle has been to work out a method of recognizing these partialities and bringing them out into the light so they can be acknowledged, and then either disregarded or utilized.

But the major problem in sympathy and detachment theories has been that the

partialities, whether consciously identified or not, probably cannot be adequately compensated for even in principle because the other major theoretical element has been repeatedly slighted: the concept of the moral equality of persons. Even when theories begin with an explicit (even moving, as in Locke) recognition of equality, one finds that as the theories develop, the concept of moral equality can become more and more qualified until at the end there can be a large contingent of persons who are not actually recognized as such. To some degree this is because of prevailing cultural understandings of personhood, which is then one of the hidden, internalized partialities affecting the composition of the resulting moral community; but even taking appropriate precautions against anachronistic approaches to older theories does not adequately explain similar problems in contemporary theories with more contemporary understandings of personhood.

Kantian theories, on the other hand, including the utilitarian theory discussed in the previous chapter, are fully predicated on the moral equality of persons;[1] one cannot lose sight of this foundation because every part of such theories refers back for its justification directly or indirectly to this element. Consequently, it is difficult, if not impossible, to justify any sort of exclusionary qualifications of what moral personhood consists in. Therefore, although moral deliberators will still have all sorts of internalized partialities upon entering the deliberative situation, these biases will have correspondingly less power to influence the deliberative process improperly the more the foundations of this kind of deliberation are kept in mind.

Having observed how hidden partialities and biases can operate to exclusionary ends, it may be worthwhile to close with some reflections on the nature of exclusion, how it may be experienced in sympathy and detachment theories, and why exclusion must

necessarily be considered an evil to be avoided. The reasons depend on the idea of moral personality or personhood.

"In the realm of ends everything has either a *price* or a *dignity*."[2] Kant's distinction between a human being having price, or being replaceable by an equivalent, and having a dignity, or being beyond all price, is what is at stake in exclusion from the moral community. Rational beings are ends in themselves (beyond all price) and therefore

> [i]t follows incontestably that every rational being must be able to regard himself as an end in himself with reference to all laws to which he may be subject, ... and thus as giving universal laws. ... It also follows that his dignity ... entails that he must take his maxims from the point of view which regards himself, and hence also every other rational being, as legislative. (The rational beings are, on this account, called persons.)[3]

Moral personality, or personhood, is free from the determinism of nature and is

> regarded as a capacity of a being which is subject to special laws (pure practical laws *given by his own reason*), so that the person as belonging to the world of sense is subject to his own personality [personhood] so far as he belongs to the intelligible world.[4] [emphasis added]

For Rawls, moral personality is characterized by the capacity both for a conception of the good and a sense of justice:

> Thus a moral person is a subject with ends he has chosen, and his fundamental preference is for conditions that enable him to frame a mode of life that expresses his nature as a free and equal rational being as fully as circumstances permit.[5]

But Rawls makes it clear that framing an effective mode of life as above requires the unity of the person ("manifest in the coherence of his plan"[6]), and it is precisely the unity or integrity of the person that is at risk when the threat of exclusion looms. Under severe and prolonged deprivation of the conditions for unity of the self, moral personality itself can crumble into irretrievable fragments. A Kantian interpretation of this process can show, in metaphorical terms, what is at stake.

Kant described the self-legislative capacity, or autonomy, of the person as stemming from his rootedness in two different worlds, the noumenal or intelligible, and phenomenal, or the world of appearances. He conceptualized the autonomy of the person as his noumenal self striving to manifest itself in the phenomenal realm, this enabling the full expression of moral personality. If this is an accurate, albeit metaphorical, description of moral personality, then (continuing the metaphor) it is likely that upon exclusion, a depersonalization process will ensue, whereupon the moral personality will begin to lose its footing in, or forget, the noumenal realm and what its task in that realm really is.

As one becomes more firmly "phenomenal" and correspondingly less of a "person," one becomes thereby increasingly susceptible to the determinism of nature. In this case it is psychological nature that will assure his deepening internalization of others' evaluation of him as a non-person and therefore unworthy of full (or even any) participation in a moral community. It is not unreasonable to suppose that after a while, the individual will reach a point of no return, where his autonomous state of mind has receded so far into the distance of memory that it is no longer retrievable. The literature of social psychology and personality development research has many examples of the terrifying ease and rapidity with which this transformation can be accomplished.[7]

In less metaphorical terms, an individual who is not regarded as a person is on the outside looking in. He has been removed from consideration; he does not belong. He is not like others; he is permanently and categorically different from others who resemble him in every respect but this one. His power to effect his plan of life is significantly diminished, if not done away with altogether. There is an impenetrable, but transparent, shield between himself and others, such that he can observe others in their lives as persons and knows himself to be outside of that possibility. No matter how strong the individual's natural self-esteem is, it cannot stand up for long against the continual dismissal of himself and his worth as a person – over time, self-doubt sets in, loss of confidence ensues, and in the worst cases, the individual internalizes the world's perception of him as unworthy and gives up on the task of being human. "Human good turns out to be activity of soul in accordance with virtue ... in a complete life"[8] – but how can a fragmented, disunified soul act in accordance with virtue?

This fragmentation and disunification occurs at different levels in sympathy and detachment theories. If the operative mode of impartiality is sympathy, and the agent feels no sympathy for the patient because of cultural, class, racial, intellectual, or other differences, the patient will likely be aware of at best being not quite noticed; there may be courtesy and politeness, but this is likely to be accompanied by blank incomprehension at the expression of one's ideas, values, difficulties, or dilemmas. To the extent that one wants one's life and its experiences to be validated, or at least witnessed sympathetically, by one's peers, repeated experiences of incomprehension can and will take their toll and the understanding that one is not actually a peer of one's fellows will gradually sink in. At worst, such incomprehension can be accompanied not by politeness but by hostility and perhaps even violence. It is likely that much prejudice is rooted in a

failure of the capacity to be sympathetic to someone who is "other" in some significant way.

However, sympathy can be educated, in part because it is a natural capacity, as has been noted earlier. As the artist's eye can be trained to "see" aesthetically, an individual's sympathy can be trained to "see" morally. A systematic exposure to the idea of the moral equality of persons, the justification of this idea, and examination of what it actually amounts to to "Love thy neighbor as thyself" can open closed eyes and minds sufficiently to reestablish the excluded individual as a moral person in the eyes of the agent. To a large extent, this is possible because of the natural sympathy that one has for oneself that enables sympathy for others. Putting oneself into another's place is the operative procedure in an impartiality based in sympathy; the reason this works is that when one sees oneself in that circumstance, one "feels" the consequences in oneself and responds accordingly.

That education to a broadened sympathy (one that includes all moral persons) is not only possible but relatively commonplace has been demonstrated over the generations as first one, then another formerly excluded group has had their personhood reinstated, arguably due at least in part to persistent and intense consciousness-raising endeavors by women, persons of color, and others. The "love" in "Love thy neighbor" is then not the love of inclination (there may never be liking or warmth in the recognition of the other as having moral equality with oneself) but that of respect.[9]

In "equality" hypothetical contract theories, the difficulties of exclusion are largely those of implementation. Moral equality is a fundamental element of these theories, although it is explicated with differing degrees of attention to how structural issues may affect outcomes of inclusion (for example, Rawls' coherentism may adversely

affect such outcomes, although the two principles, if they are still the ones to be selected, would compensate a great deal for exclusionary tendencies and in fact, under the original position and its conditions, would be selected precisely for the purpose of excluding as few people as possible from the ensuing just society).

However, it is not so easy with an impartiality which relies on Lockean detachment, for these hypothetical contract theories and the style of impartiality implicated in their construction are not based upon *equality* but upon *productivity*. It is in these theories that Kant's distinction between price and dignity takes its ugliest form, for in these theories, human beings do have price, not dignity, the price being set by their degree of productivity; those whose price is too low are not persons. With increasing productivity, price becomes worth, with the corollary that those of low price have low or no worth. None of these theories quite puts it that way; in fact Kant's respect for persons plays a small but significant part in Nozick's justification of his theory,[10] but the outcomes, most dramatically exemplified by Gauthier, are plainly anti-Kantian, dividing human beings along the boundary between cans and cannots, haves and have-nots. All haves and cans are persons, included in and protected by the moral community and the society it establishes; all the rest are not.[11]

This particular form of exclusion has proved remarkably resistant to reeducation and consciousness-raising. Indeed, it is not clear how, if at all, an individual who is highly talented and productive and who adheres to the sincere and reflective belief that such capabilities mark the worth of a person, could be reeducated. There would appear to be little or no incentive to alter beliefs that have been effective not only in his life but in the lives of others similar to himself and to society as a whole – the benefits of such a belief system are manifest and unquestionable. To ask such a person to put himself into the

perspective of another, a have-not or cannot, is likely to draw one of those blank stares of incomprehension, because his primary touchstone is productivity; he literally cannot comprehend the moral status of someone who is not productive and would not know how to put himself into that person's place. Further, he would not understand why he should trouble himself to do so.

An attempt to educate this individual to the moral equality of persons is likely to founder upon the same rock. He recognizes persons like himself, productive and capable ("Industrious and Rational"), as moral equals; to extend that recognition to someone not productive or capable would likely require an experience that would for some reason put him on the other side of the divide, such as a serious illness, a child's developmental disability, or some other personal catastrophe. Barring such a transformative experience, his sincere belief is that one *earns* one's worth as a moral person by one's deeds; one is not *given* moral worth to start with as a birthright merely by virtue of being human. "Deeds" is then interpreted relatively narrowly to reflect deeds of productivity of one sort or another. In that view, have-nots are that way primarily because of their lack of industry (this view is particularly clear in Locke) and so they have *earned* their lack of worth and low price (this view is clear in Nozick and Gauthier).There is a perverse respect for a shallowly-interpreted autonomy hidden in these beliefs; one could, after all, if one really wanted to, overcome all obstacles and produce *something* and then this would earn respect. Clearly, then, if one is not producing anything at all, one must not really want to and then ought to be left to his own devices. But autonomy is far deeper than that; it reflects the self-legislative nature of the moral person, a nature that is founded on freedom and equality. Persons can act autonomously under conditions that "express their nature as free and equal rational beings";[12] this means minimally that respect for their

freedom and equality must be in place prior to their being fully able to act autonomously, as a default position, in a manner of speaking. The demand here, however, is that persons are to act "autonomously" first and then earn respect for the freedom and equality they display in so doing, with subsequent admission to the moral community when they have done so.

In this regard, revisiting the rootedness of the moral person in dual realms could shed light on the structure of such a belief system. It is not alone by the operations of exclusionary mechanisms that one may "lose" one's rootedness in the noumenal; one may not ever have recognized it to begin with. Much depends on upbringing, the prevailing social and cultural atmosphere, and so on. But one is no less disconnected than if one had been subject to exclusion. In this case one is in fact *not* excluded, but is a fully integrated member of society. But note how the unacknowledged fragmentation of moral personhood plays out here: exclusive residence in the phenomenal realm is likely to facilitate and enable ever-increasingly a valuation of *both oneself and others* as having a "price" – having one's worth, in short, be dependent upon one's price. And recall that having a price at all is synonymous with being interchangeable with another.

Kantian impartiality, namely the conscious and repeated reflection upon the noumenal dimension of moral personality, holds out the best hope for ordinary mortals to achieve a reasonable state of impartiality that accounts for biases and respects persons, and therefore is the least likely to have outcomes of exclusion. This is so even if particular Kantian theories, or Kant's own theory, are theoretically intricate; it might be thought that ordinary persons are not likely to make a study of such theories in order to improve the quality and outcome of their moral deliberative practices. But this is not a problem. Of the three types of impartiality, it is the justification of Lockean detached impartiality

which is the most artificial and most difficult to understand, having the counterintuitive premises and outcomes that it has. Sympathetic impartiality is by and large natural in most persons; it is only that without continuous reflective attention to the element of moral equality, as provided, for example, by "Bentham's dictum," it tends to operate predominantly among like or similar persons, leaving others out of consideration.

But Kant developed his theory upon his observations of what ordinary persons actually do in moral deliberation:

> I do not ... need any penetrating acuteness in order to discern what I have to do in order that my volition may be morally good ... I ask myself only: Can I will that my maxim become a universal law? If not, it must be rejected ... because it cannot enter as a principle into a possible universal legislation, and reason extorts from me an immediate respect for such legislation.[13]

The ordinary moral deliberator does not understand the philosophical foundations of this question and will leave it to the philosophers to work it out; also, more than likely he does not phrase it in just that way, but as a variant of "What if everybody did that?"[14] But the question, regardless of how it is phrased, reflects an "estimation of the worth which far outweighs all the worth of whatever is recommended by the inclinations"[15] and points the deliberator toward his duty. "Thus," says Kant, "within the moral knowledge of common human reason we have attained its principle ...[as] the standard of its judgments."[16]

Kant was clear that the ordinary moral deliberator has it well within his power to know the right thing to do, and is unsurprised at that fact – the right thing to do is "within the reach of everyone, even the most ordinary man."[17] It should therefore also be

unsurprising that Kantian impartiality should also have an intuitive ease and the flexibility for an ordinary person to handle effectively all sorts of moral dilemmas. As Barbara Herman has shown,[18] Kantian impartiality also need not interfere with relationships of caring; in fact morality in care relationships fares best under Kantian auspices, for the particularities and partialities of specific situations all depend on respect for persons to get them straightened out appropriately, with everyone's dignity intact.

Care for the mentally ill, the elderly, and the developmentally disabled – these also fare best when considered in a Kantian framework, or in that of a sympathetic utilitarianism, and again for the same reason. One need not feel any particular sympathy for any of these persons in order to respect them *as* persons and provide appropriate attention to their needs. Nor do these individuals need to "produce" anything, or have production as the end point and purpose of their care, in order to recognize their moral equality with us. They would not have a "price" assigned to them, and their worth would correspondingly not be keyed to that price. As for criminal justice, the criminal may also not be used as a means to anyone's end, and although he has vitiated his membership in society by his actions, this does not excuse our negating his personhood. Even executing him for murder is a way of respecting him as a person, according to Kant. The principle of equality guides the punishment to be imposed; what this means is that if you kill someone, you kill yourself.[19] These options, with their clear theoretical backing, are not available to us in any of the sympathy or detachment theories examined here.

But there is a conundrum here. The prevailing *zeitgeist* described above, wherein every individual has a price and a corresponding worth, seems to suggest by its very prevalence that it is not only sympathy and Kantian impartiality that are natural, but that this kind comes naturally as well, appearing to contradict the earlier idea that its

justification is the most difficult to grasp because of its counterintuitiveness.

Here what is needed is to attend once again to the different ways in which "natural" can be understood. Earlier it was suggested that a Kantian impartiality is where both sympathetic and detached impartiality would tend if given their head and allowed full development. Indeed, the operative question within Kantian impartiality, "What if everyone could do that?" (or "Can I will that my maxim become universal law?"), is one that is in fact asked within all styles of impartiality, by virtue of seeking out an impartial stance to begin with. The problem is not failing to ask that question, but failing to recognize the universality of "everyone." It is not to be overlooked, moreover, that the motivation to impartiality of any variety stems from a desire to fair or just in a given circumstance. It is then a matter of distinguishing among the different understandings of "natural" in relation to the varieties of impartiality and how these various understandings affect the perceived scope of "everyone."

I have already noted that sympathy comes "naturally" to human persons in the sense that the *capacity* for sympathetic engagement appears inborn. But children are not born with the capacity to fully *experience* sympathetic engagement; they are born self-oriented, and as young children have limited understanding of the social graces requiring an acknowledgement of others' feelings. The ability to experience sympathetic engagement requires the cultivation that attends good upbringing.

Similarly, the capacity for Kantian impartiality appears inborn as well, as the self-legislative capacity of the individual whose self-conceit has been humiliated and who then experiences the dawning of respect for the moral law within. But again, individuals do not experience this until they at least approach adulthood and have matured, by virtue of an appropriate upbringing in their turn, into a recognition of the moral law not only within

themselves but within others as well. But these "Kantian" individuals begin their lives as children in the same developmental progression as everyone else – with a self-oriented view of the world that they must be gradually weaned away from and into a broader perspective on the relation between themselves and others.

With Lockean detachment, however, that maturation process seems not to have advanced to even the "sympathetic" stage. When stripped of the intricate trappings of their respective hypothetical contract theories, the spontaneous self-orientation of a young child is what is revealed. Events and people in the world are taken in and evaluated with reference to oneself and the benefits, advantages, and preferences one either comes to the table with or desires to obtain. Therefore, even within Lockean impartiality (taking impartiality of any kind to signify the frame of mind one strives for when confronted by the requirement to deliberate morally and one desires to be fair or just), when one does deliberate, it is reasonable to suppose that even here one asks the Kantian question. But due perhaps to an improper or incomplete upbringing, one has a severely restricted view, *already*, of who is entitled to be included for consideration as "everyone." Deliberation then cannot but reinforce this prevailing evaluative system. *Within* the system, however, outcomes are likely to be as fair and just as they would be under any other deliberative modality. It is only that this modality is likely to have the most seriously restricted population of "worthy" individuals to be fair and just with.

There is a difference, then, between what is "natural" in these various cases – a capacity and even a motivation for moral deliberation under various modalities of impartiality – and what is "naturalized" or "cultivated" (or "educated," as in Hobbes).[20] Clearly, the Lockean modality is the least cultivated and the most spontaneous – in the sense of being the most childlike – while the sympathetic modality is more advanced,

with Kantian impartiality as the full expression of adult moral personhood.[21]

To genuinely love one's neighbor as oneself, therefore, requires more than spontaneity and the natural capacity to do so. It requires an understanding of all its elements: love, my neighbor, myself, and also the understanding that one *shall* love one's neighbor as oneself – it is a command, not a suggestion. In this regard, it is Kant who will have the last word:

> ... love as an inclination cannot be commanded. But beneficence from duty, when no inclination impels it and even when it is opposed by a natural and unconquerable aversion, is practical love, not pathological love; it resides in the will and not in the propensities of feeling, in principles of action and not in tender sympathy; and it alone can be commanded.[22]

NOTES

Chapter Eight

[1] However, even Kant made it clear that only certain kinds of persons are eligible ("fit") to participate fully in the life of the community, although he hastened to add that this did not compromise their status as moral persons. *Metaphysical Elements of Justice*, trans. John Ladd (Indianapolis, IN: The Library of Liberal Arts, 1965 [1797]), 79-80..

[2] Kant, *Foundations*, Second Section, 53.

[3] Ibid., 57.

[4] Kant, *Critique of Practical Reason.*, Part I, Chapter III, 89-90.

[5] Rawls, *Theory of Justice*, §85, 561.

[6] Ibid.

[7] For example, the well-known Stanford Prison Study conducted by Philip Zimbardo et al., in which randomly assigned student volunteers lived as either prisoners or guards in a mock prison setting. The study, originally slated to last two weeks, had to be ended after six days, because of the dramatic and frightening changes in behavior of both "prisoners" and "guards." C. Haney, W. Banks, & P. Zimbardo, "Interpersonal Dynamics in a Simulated Prison," *International Journal of Criminology*, I (1983): 69-97.

[8] Aristotle, *Nichomachean Ethics*, trans. and with Introduction by Sir David Ross (London: Oxford University Press, 1966), 1098a17-18.

[9] "The possibility of such a command as 'Love God above all and thy neighbor as thyself' agrees very well with [respect for the moral law]. For, as a command, it requires respect for a law which orders love and does not leave it to arbitrary choice to make love the principle. ... That law of all laws ... thus presents the moral disposition in its complete perfection ... " Kant,*Critique of Practical Reason.*, Part I, Chapter III, 85-86.

[10] For example: "Side constraints upon action reflect the underlying Kantian principle that individuals are ends and not merely means; they may not be sacrificed or used for the achieving of other ends without their consent. Individuals are inviolable." Nozick, *Anarchy*, 31.

[11] It is worth remembering, at this point, that Kant has explicitly required of the general Will that it subordinate itself "to the internal authority of the state in order to support those members of the society who are not able to support themselves. ... Because [the wealthy's] existence depends on the act of subjecting themselves to the commonwealth for the protection and care required in order to stay alive, they have bound themselves to contribute to the support of their fellow citizens." *Metaphysical Elements of Justice*, "General Remarks – C", 93.

[12] Rawls, *Theory of Justice*, §78, 515.

[13] Kant, *Foundations*, First Section, 20.

[14] Singer, *Generalization*, 5.

[15] Kant, *Foundations*, First Section, 20.

[16] Ibid.

[17] Ibid.

[18] Herman, "Agency..."

[19] Kant, *Metaphysical Elements of Justice*, II, "General Remarks – E", 100-101.

[20] Barbara Herman speaks of "normalizing" natural motives to impartial principles. "Agency ...," 788-92.

[21] Recapitulating, in fact, the stages of moral progression delineated by Kohlberg.

[22] Kant, *Foundations*, 16.

WORKS CITED

Aristotle. "Poetics," *in The Complete Works of Aristotle.* Edited by Jonathan Barnes. Princeton, NJ: Princeton University Press, 1985.

_____. *Nichomachean Ethics.* Translated and Introduced by Sir David Ross. London: Oxford University Press, 1966.

Bacon, Francis. *The New Organon.* New York: Oxford University Press, 1986.

Baier, Kurt. *The Moral Point of View: A Rational Basis of Ethics.* Ithaca, New York: Cornell University Press, 1958.

Barry, Brian. *Justice as Impartiality.* Oxford: Oxford University Press, 1995.

Beck, Lewis White. *A Commentary on Kant's "Critique of Practical Reason."* Chicago, IL: Univerity of Chicago Press, 1960.

Butler, Joseph. *Fifteen Sermons.* Edited by T. A. Roberts. London: S.P.C.K., 1970.

_____. *Five Sermons.* Edited by Stephen Darwall. Indianapolis, IN: Hackett Publishing Company, 1983.

Cassirer, Ernst. *Rousseau, Kant, and Goethe.* New York: Harper Torchbooks, 1945.

Daniels, Norman, ed. *Reading Rawls: Critical Studies on Rawls' "A Theory of Justice."* Stanford, CA: Stanford University Press, 1989.

Darwall, Stephen. "Is There a Kantian Foundation for Rawlsian Justice?" in *John Rawls' Theory of Social Justice.* Edited by H. Gene Blocker and Elizabeth Smith. Athens, OH: Ohio University Press, 1980.

_____. "Sympathetic Liberalism: Recent Work on Adam Smith." *Philosophy and Public Affairs* 28, No. 2 (1999).

_____. *Impartial Reason.* Ithaca, NY: Cornell University Press, 1983.

Dworkin, Ronald. "The Original Position," in *Reading Rawls.* Edited by Norman Daniels. Stanford, CA: Stanford University Press, 1989.

Erickson, Milton. "Hypnotic Alteration of Sensory, Perceptual, and Psychophysiological Processes," in *The Collected Papers of Milton H. Erickson,* Vol. II. Edited by Ernest L. Rossi. New York: Irvington Publishers, Inc., 1980.

Firth, Roderick. "Ethical Absolutism and the Ideal Observer." *Philosophy and Phenomenological Research* XII, No. 3 (1952): 317-345.

Gauthier, David. *Morals by Agreement.* Oxford: Clarendon Press, 1986.

Gilligan, Carol. "Moral Orientation and Moral Development," in *Ethics: Classical Western Texts in Feminist and Multicultural Perspectives.* Edited by James P. Sterba. New York: Oxford University Press, 2000.

_____. *In a Different Voice.* Cambridge, MA: Harvard University Press, 1982.

Haney, C., W. Banks, and Philip Zimbardo. "Interpersonal Dynamics in a Simulated Prison." *International Journal of Criminology* I (1983): 69-97.

Harman, Gilbert. *The Nature of Morality.* New York: Oxford University Press, 1977.

Herman, Barbara. "Agency, Attachment, and Difference." *Ethics* 101 (July, 1991): 775-797.

Hobbes, Thomas. *Leviathan.* Middlesex, England: Penguin Books, Ltd., 1968 [1651].

The Holy Bible, Michelangelo Edition (King James Version). New York: Abradale Press, Publishers, 1969.

Hume, David. *A Treatise of Human Nature.* Garden City, NY: Dolphin Books, 1961 [1739-40].

_____. *An Inquiry Concerning the Principles of Morals.* New York: Bobbs-Merrill Company, Inc., 1957 [1752].

_____. *David Hume's Political Essays.* Edited by C. W. Hendel. New York: The Liberal Arts Press, 1953.

Kant, Immanuel. *Critique of Practical Reason.* Translated by Lewis White Beck. New York: MacMillan Publishing Company, 1956 [1788].

_____. *Foundations of the Metaphysics of Morals.* Translated by Lewis White Beck. Indianapolis, IN: The Bobbs-Merrill Company, Inc., 1959 [1785].

_____. *Lectures on Ethics.* Translated by Louis Infield. New York: Harper Torchbooks, 1963 [1775-80].

_____. *The Metaphysical Elements of Justice.* Translated and Introduced by John Ladd. Indianapolis, IN: The Bobbs-Merrill Company, Inc. 1965 [1797].

Kierkegaard, Soren. "You Shall Love Your *Neighbour*," in *Other Selves: Philosophers on Friendship.* Edited by M. Pakaluk.Indianapolis, IN: Hackett Publishing Company, 1991.

Kirk, G. S., et al., eds. *The Presocratic Philosophers.* Cambridge, England: Cambridge University Press, 1983.

Kohlberg, Lawrence. *The Philosophy of Moral Development.* San Francisco, CA: Harper & Row, Publishers, 1981.

_____. *The Psychology of Moral Development.* San Francisco, CA:Harper & Row, Publishers, 1984.

Locke, John. *Two Treatises of Government.* New York: New American Library, 1963 [1690].

Luce, R. Duncan, and Howard Raiffa. *Games and Decisions.* New York: John Wiley & Sons, Inc., 1957.

MacIntyre, Alasdair. *Whose Justice? Which Rationality?* Notre Dame, IN: University of Notre Dame Press, 1988.

MacPherson, C. B. *The Political Theory of Possessive Individualism.* Oxford: Oxford University Press, 1962.

Mill, John Stuart. *Utilitarianism, Liberty, and Representative Government.* London: J. M. Dent & Sons, Ltd., 1947 [1861-3].

Nagel, Thomas. "Rawls on Justice," in *Reading Rawls.* Edited by Norman Daniels. New York: Basic Books, 1975.

_____. *Equality and Partiality.* New York: Oxford University Press, 1991.

_____. *The View From Nowhere*. New York: Oxford University Press, 1986.

Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.

Rawls, John. "Kantian Constructivism in Moral Theory," in *Moral Discourse and Practice*. Edited by Stephen Darwall et al. New York: Oxford University Press, 1997.

_____. *A Theory of Justice*. Cambridge, MA: The Belknap Press, 1971.

Rousseau, Jean-Jacques. "The Social Contract," in *The Essential Rousseau*. Translated by Lowell Bair. New York: The New American Library, 1974 [1762].

Sacks, Oliver W. "To See and Not To See." *The New Yorker* (10 May 1993), 59-66.

Selby-Bigge, L. A., ed. *British Moralists*. With a new introduction by Bernard H. Baumrin. Indianapolis, IN: The Library of Liberal Arts, 1964 [1897].

Sidgwick, Henry. *The Methods of Ethics*, 7th ed. Indianapolis, IN: Hackett Publishing Company, 1981 [1907].

Singer, M.G. *Generalization in Ethics*. New York: Russell & Russell, 1971.

Smith, Adam. "The Theory of Moral Sentiments," in *British Moralists*. Edited by L. A. Selby-Bigge. Indianapolis, IN: The Library of Liberal Arts, 1964 [1897].

_____. *The Theory of Moral Sentiments*. New York: Augustus M. Kelley, Publishers, 1966 [1759].

Stark, Cynthia. "Decision Procedures, Standards of Rightness, and Impartiality." *Nous* 31:4(1997): 478-495.

Williams, Bernard. "The Idea of Equality," in *Moral Concepts*. Edited by Joel Feinberg. New York: Oxford University Press, 1969.

_____. *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press, 1985.

_____. *Morality*. Cambridge, England: Cambridge University Press, 1993.