8-1982

The Political Philosophy of John Rawls: An Examination of His View of Man and Social Stability in the Just Society

Calvin Jay Mouw
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses

Part of the Political Science Commons

Recommended Citation
https://scholarworks.wmich.edu/masters_theses/1669

This Masters Thesis-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Master's Theses by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
The thesis constitutes a critical examination of Rawls's psychological assumptions in *A Theory of Justice*. Rawls's psychological theory, derived from Rousseau and Kohlberg and characterized by his "Sense of Justice," depends upon overly altruistic assumptions about man's nature. These assumptions, along with resultant high levels of public motivation, lead to a high degree of social instability. Rawls fails to address this instability adequately.

The alternative model developed in this thesis takes as given the demonstrated ability of human beings to cooperate with one another and bases this capacity, in the tradition of Hume, on a mechanism of sympathy which brings together a wide variety of individual egos in a social situation. This position is based largely upon well-understood self-interest. This model is one in which the proper mixture of egoistic and altruistic motives are achieved, and more sufficiently than Rawls's psychology, provides for application of the concept of social stability.
ACKNOWLEDGEMENTS

Thanks are offered to Drs. Isaak and Mg Anaw for the time and assistance that were given during my studies at Western Michigan University, and special gratitude is expressed to Dr. W. A. Ritchie whose guidance and comments have resulted in a far better understanding of the world of political philosophy.

Calvin Jay Mouw
INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
MOUW, CALVIN JAY

THE POLITICAL PHILOSOPHY OF JOHN RAWLS: AN EXAMINATION OF HIS VIEW
OF MAN AND SOCIAL STABILITY IN THE JUST SOCIETY

WESTERN MICHIGAN UNIVERSITY M.A. 1982

University Microfilms International
300 N. Zeeb Road, Ann Arbor, MI 48106
PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark √.

1. Glossy photographs or pages 
2. Colored illustrations, paper or print 
3. Photographs with dark background 
4. Illustrations are poor copy 
5. Pages with black marks, not original copy 
6. Print shows through as there is text on both sides of page 
7. Indistinct, broken or small print on several pages 
8. Print exceeds margin requirements 
9. Tightly bound copy with print lost in spine 
10. Computer printout pages with indistinct print 
11. Page(s) _______ lacking when material received, and not available from school or author.
12. Page(s) _______ seem to be missing in numbering only as text follows.
13. Two pages numbered ________. Text follows.
14. Curling and wrinkled pages 
15. Other ____________________________

University
Microfilms
International

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
**TABLE OF CONTENTS**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ACKNOWLEDGEMENTS</strong></td>
<td>ii</td>
</tr>
<tr>
<td><strong>INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Chapter</strong></td>
<td></td>
</tr>
<tr>
<td>I. MORALITY AND HUMAN NATURE</td>
<td>8</td>
</tr>
<tr>
<td>Humans as Self-interested Creatures: Morality</td>
<td></td>
</tr>
<tr>
<td>Imposed Externally</td>
<td>12</td>
</tr>
<tr>
<td>Human Beings as Inherently Capable of Goodness</td>
<td>15</td>
</tr>
<tr>
<td>Morality Arising from Society and the Conditions of Social Living</td>
<td>22</td>
</tr>
<tr>
<td>II. RAWL'S VIEW OF MAN IN THE ORIGINAL POSITION</td>
<td>36</td>
</tr>
<tr>
<td>The Original Position</td>
<td>39</td>
</tr>
<tr>
<td>The Thin Theory of the Good</td>
<td>46</td>
</tr>
<tr>
<td>The Concept of Morality and Social Choice Functions</td>
<td>59</td>
</tr>
<tr>
<td>III. THE MAXIMIN PRINCIPLE AND THE PRINCIPLE OF AVERAGE UTILITY</td>
<td>66</td>
</tr>
<tr>
<td>Arrow-Harsanyi Critique of the Maximin Principle</td>
<td>68</td>
</tr>
<tr>
<td>The Principle of Average Utility</td>
<td>73</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Chapter

IV. Rawls's Moral Psychology ........................................ 80
   Rawls's Moral Psychology ........................................ 82
   Critique of Rawls's Moral Psychology ............................ 97
   Human Nature in Political Analysis ............................... 107

V. The Concept Of Stability ........................................ 114
   Rawls's Concept of Stability ..................................... 119
   An Alternative Model of Psychological Assumptions ............ 135

CONCLUSION ......................................................... 145

BIBLIOGRAPHY ....................................................... 149
INTRODUCTION

Rawls in his book *A Theory of Justice* (Rawls, 1971), raises a very interesting problem, so far largely neglected by moral and political philosophers. Suppose there is a society with a strong sense of justice among its citizens, and also just or mostly just institutions. Would such a society be stable? Rawls argues that it would be. He also states that a society based on his principles of justice would be much more stable than one based on the principles of utilitarianism.

It is important to realize that considerable portions of social contract theory and utilitarian moral theory can be interpreted as having been designed to offer a reason, at least from the point of view of human interest, for all human beings to be moral. Hobbes's account of the basis for political obligation tries to show that a political community, based on certain laws, is in the best interest of each and every individual. Hume's account of justice illustrates the difficulty in trying to show a satisfactory explanation of the motivation to be just. Hume provides a "justification" for justice based primarily on utilitarian grounds. (Hume, 1963) Rawls revives the social contract theory.
Hobbes in *The Leviathan* (1957) and Locke in the *Second Treatise* (1952) made psychological assumptions about the individuals involved in contracting into the society. One of these assumptions was that self-interest was a dominant human motive. In Rawls's contract he basically describes a situation in which principles of justice are to be chosen; the one important motivational assumption Rawls makes is that the parties are mutually disinterested. Rawls's man can be basically classified as one of limited altruism (or self-interest expanded over to the immediate family, close friends and neighbors). This is one of the weakest motivational assumptions he can make.

There are two important theoretical constructions in which Rawls's motivational assumptions are developed: in the description of the "Original Position," and in the section on the "Sense of Justice."

For Rawls it is important that a conception of justice should relate to actual human motivation in such a way that a "well-ordered" society would be stable with respect "to the justice of its institutions and the sense of justice needed to maintain this condition." Rawls gives an account of the development of the sense of justice in a three-stage process with a corresponding psychological principle for each one. The process is basically a derivation from philosophers such as Rousseau (1964) and psychologists such as Piaget (1951) and Kohlberg (1971).
Rawls's just society as described in this context, however, is not merely an improved version of the best societies that exist at the present; it is rather, unlike any society known to political scientists or historians. It is a society in which the citizens and legislators are never motivated by their own self-interest or, in the case of the legislators, by the selfish interests of their constituents. They are, in contrast, always motivated by this strong sense of justice. This society, as depicted by traditional political theory and classical utilitarianism and firmly established in much of the current literature; for example, Downs, *An Economic Theory of Democracy* (1957), and Buchanan and Tullock, *The Calculus of Consent* (1962).

It will be argued that Rawls is quite right in rejecting such motivational assumptions as a totally realistic view of human motivation. It is not true that the average citizens never care about anything but their own narrow economic self-interest, or that politicians do not care about anything except trying to get elected or re-elected. Moreover, it can be shown that under some conditions many rich people will support legislation benefiting the poor and increasing their own taxes. Also, many politicians and elected officials have made unpopular decisions, in many cases hurting their chances for reelection. It must be remembered, however, that
psychological egoism does not claim that this motivational assumption is literally true; all that it claims is that our political system operates as if these motivational assumptions were correct.

Therefore, the motivational assumptions of psychological egoism do come close to being true, and one should stop and think before accepting totally Rawls's theory of stability. It will be argued that Rawls's theory of man, dominated by this sense of justice, is a far too altruistic position to take; and in fact, does not contain the Aristotelian components that he proposes. The burden of proof is on Rawls to show that this altruistic position would not lead to instability. Rawls places the burden on altruism, but he fails to recognize the extent to which the burden of proof of social stability rests with this altruistic assumption.

It will be argued that a stable society needs a proper balance between egoistic and altruistic motivation. Without political leaders fighting for altruistic objectives, or without private citizens giving their political support, democratic societies would not have achieved even that, no doubt imperfect, level of social justice and government they currently have.

However, political movements based largely or completely upon well-understood self-interest are also an essential component of a political system. Citizens
working for their own sectional economic interests may be biased judges of the public good or interest, but at least in some cases they are well-informed. On the other hand, citizens pursuing highly altruistic objectives might often work for causes they know little about, or of which they have very one-sided information. The working citizenry pressing for its own economic interests will at least know what it is talking about. However, some far-off altruistic millionaire fighting for this same worker's interests might have a wrong idea of what this worker really needs or wants. A society in which everybody neglects his own interests, and is looking out for everybody else's interests, probably would not be a very stable society.

Therefore, it seems that a society with a reasonable prospect for social stability would not be a society where ordinary citizens and legislators would be primarily motivated by their sense of justice. Instead, it would be a society where most people would be motivated by a mixture of egoistic and altruistic interests. It should be mentioned that that society would have to maintain a strong sense of justice, but it does not mean that the pursuit of justice should take all priority, or be the main preoccupation of life. It only means that citizens and legislators would have to show enough respect for justice so as to stop pressing their own egoistic and altruistic
objectives beyond the point where they would violate the legal and moral rights of other people; also so as to restore these rights if they have been violated by injustices in the past.

Principally, this thesis will constitute a critique of Rawls's moral psychology based on the fact that Rawls misconstruits his theory of moral psychology and the motivational assumptions of man. With this misconstruction, Rawls ends up with a theory of justice in which the level of stability would be lacking to the point of society being chaotic and devoid of social order.

There will be five chapters in this thesis. The first will constitute a look at the philosophical views of the psychological conception of human nature. The second will look at Rawls's conception of human nature and justice, with emphasis placed on the first of the two major constructions in which Rawls explains his motivational assumptions, the Original Position. The third chapter will analyze the result of Rawls's misconception in the Original Position, his concept of the Maximin Principle. A critique will be offered of this strategy, and the argument for the Principle of Average Utility will be presented. The fourth chapter analyzes the second major source of Rawls's moral psychology and motivational assumptions, the Sense of Justice. The first part of this
chapter will analyze Rawls's moral psychology and his Rousseauian-Kohlbergian derivation; the second part will offer a critique of Rawls's position and also show how this concept of moral psychology affects his principles of justice. The fifth chapter constitutes a look at the aspect in which Rawls's moral psychology fails the most, the concept of stability. Finally, an alternative model of psychological assumptions will be offered, a model based on the Humean construction of self-interest and the Arrow-Nagel circumstances of altruistic behavior. (Hume, 1964; Arrow, 1972; Nagel, 1970) This model will be presented as one in which the problem of stability would be properly addressed, and on which a theory of justice could be more securely built.
CHAPTER I

MORALITY AND HUMAN NATURE

It is usually assumed that a conflict exists between self-interest and the interests of others, between egoism and altruism: one has to act contrary to one's self-interest in order to act altruistically. Comte (1975) said that in order to act altruistically, individuals need to overcome self-interest and develop a concern for the welfare of others. It was thinking like this that led to the notion that for an act to be altruistic, there should be no positive reinforcement at all for the actor. However, most psychologists would say that this is an unreasonable motivational assumption. Psychological theories have assumed that behavior is influenced by its consequences. If there were no reinforcement for behavior that required sacrifice on the part of the actor, such behavior would not have existed over time. As the concern for others develops in people, they will acquire a sense of satisfaction from this feeling of serving the needs of others. Therefore, instead of there being a conflict between the needs of others and the self, it can be said that they often identify with one another.

The concepts of morality and human nature of course,
have been a traditional subject of concern of philosophers. "Morality is a set of rules, customs, or principles that regulate people's conduct in relation to other people, conduct that affects human welfare" (Staub, 1978, p. 10).

There have been several stages of morality distinguished by philosophers; these stages are characterized by individuals on the one hand, and society or culture on the other. The lowest level of morality has been called "prerational or group-enforced morality," in which the person acts morally out of fear of negative consequences for not following the moral standards and conversely, craves the positive consequences for following them. The highest level of morality has been called "personal, rational autonomous morality." At this level the moral standards "belong" to the individual; they are not derived directly from the rest of society, but are the result of individual, rational self-examination. It is on this level that an individual can hold beliefs or values which differ from the rest of society. A middle or intermediate level may be called "internalized morality." This level is characterized by individuals who learn the values and standards of their culture and proceed to accept them as their own; they either rationally or irrationally internalize them.

Moral philosophy in its most mature form begins to play a role at the level of autonomous morality. The standards and values of a society are not necessarily the moral standards and values. There are several reasons for this.
First, some of a society's rules may be unjust or immoral. Second, moral values tend to vary in different cultures. What may be just or acceptable conduct in one society may not be in another. Third, some standards or values within a society may contradict each other. By allowing exceptions to some rules, or by not enforcing others, conflicts over what the accepted values are may occur.

Within philosophy, morality is often related to the concept of obligation. Moral values imply obligation, immoral values do not. Moral values usually refer to the consequences of moral conduct on individuals or the group as a whole. Philosophers have developed two different approaches to morality. The first approach is that the consequences of an act are the basic criteria of morality; i.e. the "maximization of beneficial consequences." An individual should act morally in order to bring about the greatest amount of positive or beneficial consequences. The second approach says that moral conduct must be based on certain basic rules or principles. People who advocate this approach have developed rules and principles within a wide spectrum of ideas.

As for the concept of justice, it was regarded as a basic virtue by Plato, and by many other philosophers. Kant offered his "categorical imperative" as the basic virtue of morality. One should act only on that maxim which you can at the same time will to be universal law; meaning simply, one should act in a certain way only if
one is willing to have everyone else act in that way in the same situation. With justice as the basic virtue, other virtues can be derived from it, according to the situation. Traditionally, some of the virtues have included temperance, justice, and prudence. Prudence, for example, implies the capacity to evaluate the consequences of action. It is only when the consequences are correctly evaluated that one can then apply the appropriate value or principle in action. Some philosophers, such as Frankena in *Ethics* (1963), have suggested that the basis of morality lies in benevolence — the desire to do good, and justice — the distribution of this good to affect the greatest number of individuals.

In the classical Greek conception of morality the emphasis was placed on moral judgment, or the ability to decide between what is right and what is wrong. Generally, it can be said that moral philosophy is concerned with the determination of moral values, and the way it does this is through reason. However, there has long been another motivational force and that is the will, and this also has to be considered part of morality; another important element is feelings, important because they connect thought and action. Most psychologists believe that feelings of guilt for not doing the right thing or satisfaction for doing the right thing motivate moral action. Action itself is another element of morality, the adherence or disadherence to moral values or conduct.
Humans as Self-interested Creatures: Morality Imposed Externally

The assumption that man is inherently evil has long been present in philosophy, and its influence still exists today in some psychological theories. This assumption was heavily promoted by many theologians, those who believed that man was inherently sinful. The strongest and most well-known secular pessimistic view of human nature came from Thomas Hobbes.

Hobbes looks at the individual in a state of nature as a combination of desires and appetites. Contrary to Aristotle and the whole later medieval tradition, Hobbes's man never looked at society as an end but always as a means. Men are not political or social animals, but rather individual beasts.

Felicity is a continual progress of the desire, from one object to another; the attaining of the former, being still but the way to the latter. The cause whereof is, that the object of man's desire, is not to enjoy once only, and for instant of time; but to assure forever, the way of his future desire.... So that in the first place, I put for a general inclination of all mankind, a perpetual and restless desire of power after power, that ceaseth only in death. And the cause of this, is not always that a man hopes for a more intensive delight, than he has already attained to; or that he cannot be content with a moderate power: but because he cannot assure the power and the means to live well, which he hath present, without the acquisition of more. And from hence it is, that kings, whose power is greatest, turn their endeavours to the assuring if at home by laws, or abroad by wars: and when that is done, there succeedeth a new desire; in some, a fame from new conquest; in others, of ease and sensual pleasure; in others, of admiration, or being flattered for excellence in some art, or other ability of the mind.
Competition of riches, honor, command, or other power inclineth to contention, enmity, and war: because the way of one competitor, to the attaining of his desire, is to kill, subdue, supplant, or repel the other. (Hobbes, 1957, Part I, Ch. 11)

All men, says Hobbes, are egoistic; and it seemed to him self-evident that Egoism was the fundamental law of human nature. Everything one does has reference to the result of the action for oneself. Hobbes's Egoism was, moreover, partially reinforced by a tendency toward a solipsistie theory of knowledge. Solipsism claims that the only objects which any of us can possibly know are in one's inner state of consciousness. If a person can only know his own state of mind, nothing other than his own state of mind will concern him. This solipsistic tendency in Hobbes's conception reinforces his psychological theory of man as egoistic. Every individual is subject to the conceptions of his mind, and these in turn are the result of the desires and appetites of his body.

The difficulty arises when men in the state of nature realize that there are other individuals who are equal to themselves. This difficulty is seen in the struggle for power, each individual dominated by his lust for control over the other. This realization that there are other powers to be reckoned with comes as a great shock to the individual in Hobbes's state of nature. The state of nature is thus characterized by constant warfare.

It should be noted that throughout Hobbes's account of human psychology within the state of nature, he always
assumes that individuals will use discretion in their power struggles. An emphasis should be placed on individuals; it is never groups of individuals who are involved. This is not only characteristic of Hobbes but also many other early modern thinkers.

Within psychological theory, the pessimistic view of man's nature can be seen in the psychoanalysis of Freud (1966). Freud states that the basic impulses of human nature are sexual and aggressive, and that morality derives from the necessity to restrict the non-moral, self-interested character of man. Freud thought that man developed ethical norms in order to make social life possible. Conversely, Freud also thought that society sometimes develops values and principles that are too restrictive, and thus inhibit the expression of man's basic impulses. Man can develop an overly strict conscience in the developing or in the "internalizing" of societal standards, thus leading to psychological sickness.
Human Beings as Inherently Capable of Goodness

Staub, in *Positive Social Behavior and Morality: Social and Personal Influences* (1978), states that many conceptions of justice started with Plato. Plato described the Socratic method of developing virtue, a method that can be basically developed in two parts. The first part is the negative aspect, the part aimed at getting rid of all the preconceptions of the mind, including such things as principles, moral values and so on. The positive part involves the concept of self-examination. This process was thought to lead to the knowledge of virtue and included in this process was not only the self-consciousness but also the belief in some sort of principle within a person that would prevent action contrary to the laws of human nature. To know the good is to do the good.

Staub (1978) lists several important characteristics of the Platonian view of human nature. Staub (1978) states that the Platonian view "has within itself the elements of man's goodness. Another principle is that morality is rational, and that man can become moral through self-examination that leads to understanding. A third, less direct implication is that man can be corrupted by circumstances; this is suggested by the emphasis on man's divesting himself of societal convention if he is to reach self-knowledge. Finally, the conception of an internal monitor foreshadows the conception of conscience; both are internal guides to the right action" (p. 14).
The inherent goodness of man can also be found in the later writings of many other philosophers and psychologists. Most of these writers say that man is capable of good if certain conditions are met. Aquinas, (1970), relied heavily on Aristotle and thought of the conception of the will as important in the determining of human action. Aquinas also believed in the use of reason along with the will as a guide toward the perfection of every individual. He believed that goodness was inherent, that it was natural, and that not to do good was unnatural and thus sinful.

John Locke and His State of Nature

According to some of the naturalist writers, human nature is or could be good. If man lived according to the natural law, there would be no evil. The environment plays an important and influential role. John Locke in The Second Treatise is an example of this.

Locke begins with a discussion of the nature of man and of a state of nature. There are basically two characteristics in his state of nature: first, the state of nature is a "state of perfect freedom" in which individuals can do what they want as long as it is within the limits of the law of nature. Secondly, it is a state of equality. No one has any more rights than anyone else; men are equal in the rights they possess. Up to this point, Locke does not differ with Hobbes; they begin to differ in that for Hobbes
the equality and freedom of the individual in a state of nature leads directly to war, as mentioned previously. For Locke however, his state of nature is no utopia, but it is certainly no state of war either. The difference appears in Locke's use of the natural law.

The State of Nature has a Law of Nature to govern it, which obliges everyone: and Reason, which is that Law, teaches all Mankind, who will but consult it, that being all equal and independent, no one ought to harm another in his Life, Health, Liberty, or Possessions. (Locke, 1952, Ch. 6)

Because Locke's man is assumed to be a rational creature before the contract, it makes it much easier for him to accept the law of nature. Freedom is not "license." While each individual has the freedom to "dispose of his person or possessions," he does not have the freedom to "destroy himself, or so much as any creature in his possession, but where some nobler use than its bare preservation calls for it" (Locke, 1952, Ch. 6). The key words are "some nobler use than its bare preservation." Differing here with Hobbes, Locke sees the preservation of human life coming from that combination of natural law and reason. Within this reason, man values not only his own life, but also the preservation of the lives of others. Hobbes on the other hand bases his law of nature on the fear for our own lives.

One could say then that Locke's state of nature has two distinctive features. Men in the state of nature generally respect one another, with reason telling them that they should not "harm another in his Life, Health, Liberty, or Possessions." Secondly, men, in contrast to Hobbes,
do not base this respect on self-interested or egoistic motives; rather, they value each other as means to an end.

Men then, are governed in the state of nature by natural law; and since men are rational creatures, they are capable of understanding this law. Evil in man was brought about by human institutions and customs that went against the natural law. When these conditions began to occur, undesirable behavior or behavior against the law of nature also occurred. Clearly, in the overall view of Locke's state of nature, man was potentially good.

Rousseau's Man in the State of Nature

For Rousseau, man in the state of nature would be a savage, just like any other animal. But Rousseau classifies him as a noble savage, one that lived outside of society, and was therefore good. However, if man is a noble savage, where does he acquire this nobility? Ramon Lemos (1977), in Rousseau's Political Philosophy, states that nobility would be derived from primarily two sources — from his independence and from his natural compassion. The independence results from man being in the pure state of nature, for this is a state of independence; there are no permanent relationships or associations with anybody, these only come in the development of a society. In the pure state of nature, no one is dependent on anyone else, each individual is self-dependent. It is with this notion of self-dependence
and self-preservation that man maintains his nobility within the state of nature, as against for example, the moral nobility that is attainable by man in society. The result is two different kinds of nobility, and Rousseau states that no man can possess both kinds. Natural nobility can be possessed only by man in the pure state of nature, whereas rational or moral nobility can only be achieved by a man in society. By man moving into society, he loses some of his natural compassion, and also terminates his natural independence. Thus, both sources of nobility are effected.

According to Rousseau, prior to or independent of reason, there are two principles that are operating inside us. "One of them deeply interesting us in our own welfare and preservation, and the other exciting a natural repugnance at seeing any other human being, and particularly any of our own species, suffer pain or death" (Rousseau, 1964, p. 157). The first principle is self-love, "a natural feeling which leads every animal to look to its own preservation, and which, guided in man by reason and modified by compassion, creates humanity and virtue" (Rousseau, 1964, p. 182). There is a difference between self-love and egoism. Self-love, as the quote mentions, is a "natural feeling" that is present in every animal, including man, both in the pure state of nature, and in society. Egoism, states Rousseau (1964), "is a purely relative and factitious feeling, which arises in the state of society, leads each individual to make more of himself than of any other,
cause all the mutual damage men inflict one on another, and

is the real source of the sense of honour" (p. 182). Moreover,

In our primitive condition, in the true state of na-
ture, egoism did not exist; for as each man regarded
himself as the only observer of his actions, the only
being in the universe who took any interest in him,
and the sole judge of his deserts, no feeling arising
from comparisons he could not be led to make could
take root on his soul; and for the same reason, he
could know neither hatred nor the desire for revenge,
since these passions can spring only from a sense of
injury; and as it is the contempt or the intention
to hurt, and not the harm done, which constitutes the
injury, men who neither valued nor compared them-
selves could do one another much violence, when it
suited them, without feeling any sense of injury.
In a word, each man, regarding his fellows almost
as he regarded animals of different species, might
seize the prey of a weaker or yield up his own to a
stronger, and yet consider these acts of violence as
mere natural occurrences, without the slightest emo-
tion of insolence or despite, or any other feeling
than the joy or grief of success or failure.
(Rousseau, 1964, p. 182)

Thus, while self-love and natural compassion are the
two fundamental principles of man's soul, egoism is not.
Egoism can take place only when man is in society, and it
is not natural to man, like self-love and natural compassion
are. It is at this point that Rousseau's state of nature
becomes diametrically opposed to Hobbes's state of nature.

The notion that human nature is good or that man will
develop to become good under the right circumstances has
also been implied by several contemporary psychologists.
The concept of self-actualization was used extensively
by Rogers and Maslow, meaning essentially that under cir-
cumstances in which man is allowed to actualize his own
nature, this man will end up being both a psychologically healthy person, and also a good person. Erich Fromm in his earlier writings also stressed the concept of self-actualization and the potential for man being good.

Kohlberg's (1971) "cognitive developmental" theory of moral development also assumes that the potential of man is to be morally "good", rather than "bad". In Kohlberg's view, varied experiences and relationships between people result in a progressive moral development of a person. These experiences result in the further development of a person's differentiation of "right" and "wrong", and with this development comes a new stage. This moral reasoning becomes further advanced at each stage, each stage being more moral than the one before. Initially, man is in a primitive form of morality, where what is "right" is defined by those in a position of authority; whereas in the most advanced or autonomous stage of morality, men can formulate their own principles of what is right or wrong. This last stage is significant in that it is characterized by "beliefs in justice and in the sanctity of human life" as the basis for judging the principles of morality. In summary, Kohlberg's theory assumes that man is potentially good, and that it is through interaction with other people and experiencing different roles that man will be able to progressively achieve higher levels of morality.
Morality Arising from Society and the Conditions of Social Living

In examining the sources of morality, one must consider the results of people living in social groups, and also the role that social organization plays in the overall influence of social existence. Even Hobbes, along with many Enlightenment thinkers such as Adam Smith, suggested that there be some kind of social balance that controls the selfishness of man. Hobbes believed in a model system in which forces could be set up which would lead each man, for his own selfish reasons, to contribute to the overall common good and social stability of a society. Smith, about one hundred years later, suggested that the natural law in economics was that each man was to act in his own interest and thus benefit all of society.

The Moral Philosophy of David Hume

When speaking of the moral philosophy of Hume, one should be understood to mean his theory of man in relation to his fellow man. Hume's ethical theory is closely connected to his idea of society, and one cannot be conceived without the other. Morality, thus, begins with the consequences of people living in social groups.

Hume bases his theory of the psychological development of man on the passions, of which there are two categories, the direct and the indirect. The direct passions are those
that derive immediately out of pleasure and pain; the indirect are built upon the direct. In his *Treatise on Human Nature* (1964) Hume lists among the direct passions "desire, aversion, grief, joy, hope, fear, despair, and security." The indirect passions include "pride, humility, ambition, vanity, love, hatred, envy, pity, malice, generosity, with their dependents" (II,II-1,1). Hume, of course, is concerned to explain human conduct within the terms of the interaction between these passions.

By undertaking this course, Hume is trying to explain human behavior much differently than did those of the ancient Greek tradition. Aristotle, as mentioned previously, emphasized human personality as the development of potentialities. Hume offers an explanation which is much more mechanistic, for he looks at the human psyche through the interactions between men in society. While Hume does not claim that man can be conceived only on these interactions, he does tend to think that man can best be understood in terms of his psychological origins. On this particular subject, Hume would reject the view of Hobbes; but while rejecting Hobbes's position, Hume would still be much closer to his thinking than that of the classical Greeks. The basic difference between the psychology of Hume and the psychology of Hobbes lies in the fact that Hume denies the existence of the extreme egoistic qualities that Hobbes has given to man. While Hume was one
of the more well-known expositors of the belief that man's selfish motives could lead to the overall development of a positive social order, he does not in fact offer psychological qualities as extreme as those of Hobbes.

Hume's Theory of Sympathy

Hume deals with sympathy both in his Treatise and in the Enquiry (1963), and it is interesting to contrast the two. In the Treatise, Hume states that sympathy is a "propensity of human nature," and it is through this that man is able to share the pains and pleasures with his fellow men. It is this same tendency that also gives us the same social feeling that brings us to form the very basis of our social life. Mall in Hume's Concept of Man: An Essay in Philosophical Anthropology (1967), states that "sympathy is the faculty of our approbation and disapprobation of virtues and vices respectively" (p. 93). Man has the inherent tendency to promote the good of a society. Sympathy is not a virtue or an emotion, it is just the way our human nature works. According to Mall, sympathy also generates disinterested passions of which if they did not exist, no moral life would be possible. He also says that they are devotional in nature, meaning that they lead us to look at other individuals and society as a whole with a much "greater and richer value." (Mall, 1967, p. 93)

Sympathy, according to Hume, is a natural and
altruistic principle that can be found in all of us; and it is this feeling of sympathy that generates concern and love of others, and concern for all of society.

Now we have no such extensive concern for society but from sympathy; and consequently it is that principle, which takes us so far out of ourselves, as to give us the same pleasure of uneasiness in the character of others, as if they have a tendency to our own advantage or less. (Hume, 1964, p. 315-316)

In comparing Hume's view of human nature with those of Hobbes and Locke, we can say that society is not the result of certain manipulatory interactions among certain individuals. Hume does not agree with the social contract theorists who take human intellect or reason as the basis for both society and morality. As mentioned, for Hume, human nature is something that is naturally sympathetic and social; for sympathy is a feeling that is inherent. Society, for the general contract theorists, is the "artificial result of an equally artificial contract." Hume calls Hobbes and Locke "arm-chair moralists," because they work out an abstract moral system to explain the more concrete facts of moral life. As Mall (1967) suggests, "man's social nature is not an attainment of a particular stage in his socio-cultural development. Man is born in a society, as he is born in a language" (p. 93).

It is this innate feeling of sympathy that makes life worth living. It transcends our egoistic feelings of pleasure and pain, and as a result, we naturally believe that others have these same feelings; feelings of sorrow
and pity, and feelings of joy and hope. In the Treatise, Hume (1964) states that: "We are certain that sympathy is a very powerful principle in human nature.... All lovers of virtue (and such we all are in speculation, however we may degenerate in practice) must certainly be pleased to see moral distinctions derived from so noble a source, which gives us a just notion of both the generosity and capacity of human nature" (p. 579).

Hume deals with sympathy differently in his Enquiry Concerning the Principles of Morals than he did in the Treatise. In the Treatise, the principle characteristic could be defined as Hume's tentative belief that he is able to analyze this feeling even more deeply. Mall states that this becomes clear when Hume tries to explain sympathy as depending on different relations, and that he thinks sympathy can be explained as a result or causality of these different relations. The discussion continues that if man was to be more sympathetic to his own relations, it would lead to feelings of egoism and not natural sympathy. In the Treatise: "Sympathy depends on the relation of objects to our persons to ourselves; since we are most uneasy under the contempt of persons who are both related to us by blood, and contiguous in place. Hence we seek to diminish this sympathy and uneasiness by separating these relations, and placing ourselves in a contiguity to strangers, and at a distance from relations" (Hume, 1964, p. 323).
Mall thinks that Hume here seems to be interested in working out some form of mechanistic theory of sympathy, and that in his Treatise he makes too much of the theory of association, and also that Hume is still not seeing the futility of his attempt. "Before one is conscious of the different relations as father, mother, etc. one sympathizes with them, i.e. to say sympathy precedes relation. All that results due to these relations is an influence" (Mall, 1967, p. 95). In the Enquiry, Hume does not seem to try a more mechanistic explanation of his theory of sympathy. He continues to argue that sympathy is something that is inherent in human nature; however, while the methodology in the Treatise is an "explanatory psychology," or that which is used more in the natural sciences such as chemistry or physics; in the Enquiry he seems to be more prescriptive in his approach, thus possibly drawing a relationship between the concept of "enquiry" and a more descriptive methodology. Hume uses different synonyms in place of sympathy, for example "humanity" or sometimes "fellow-feeling." It is known as the feeling in us that causes us to be concerned about the well-being of others. Thus, it can safely be said that the feeling of sympathy is always an altruistic feeling, and never egoistic.

In summary, sympathy is a natural feeling on which Hume bases his moral and ethical theory. The social interest, or the public interest, is served by the "moral
approbations and disapprobations" which are the concerns of sympathy. Everyone must be concerned with the welfare of others. Therefore, Hume says that it is sympathy which gives us our moral ideas and also lets them develop in us.

Hume's Refutation of Egoism

Hume seeks to refute egoism by stating that the masses at least believe that there are both altruistic and selfish passions, and that these altruistic passions can in no way be interpreted as selfish. Hume then puts the burden of proof on those who say that altruistic passions do not exist. He believes that no one, including Hobbes, has shown at any time that any form of selfish feelings could be interpreted as altruistic. "Until, therefore, someone offers better evidence that the common distinction between self-love and love of others is illusory, we are bound to accept the general view of mankind as having some congruence with reality" (Sibley, 1970, p. 416).

Hume takes his argument further by stating that in the observation of animals we find that they are capable of kindness towards others of their own kind and to humans alike. Hume wonders whether Hobbes would account for this as "sentiments, too, from refined deduction of self-interest" (Sibley, 1970, p. 416). If the "reductionists" admit that there is this altruistic feeling among lower life animals, then they would have a hard time trying to convince anyone
that this feeling does not exist among humans.

Hume also maintains in his Enquiry that it is this outward feeling of benevolence, or feelings for others, that makes possible a common moral life. Without it, instability arises, the conventions of society are not durable enough to maintain some form of stability.

Hume makes one other important observation in the Treatise, and it is one of the more well-known aspects of his writings. "Reason is, and ought to be, the slave of the passions, and can never pretend to any other office than to serve and obey them" (II, III, 3). What is important here is that he is rejecting the classical Greek notion that the passions can somehow be made subordinate to reason. Hume (1964) continues: "A passion is an original existence or, if you will, modification of existence, and contains not any representative quality which renders it a copy of any other existence or modification" (II, III, 3). What Hume is saying here is that certain passions cannot be in conflict with reason. Sibley writes in Political Ideas and Ideologies (1970), "the passion of anger, for example, cannot be in conflict with reason which has to do with discovery of truth. Only that can be contrary to reason which has reference to it" (p. 417). In response to the idea that reason gives us the basis for morality, Hume states that "it's not contrary to reason to prefer the destruction of the whole world to the scratching of my finger" (p. 417). Reason has nothing to do with the
formulation of our moral principles, for they are the result of our passions.

The Nature of Morality in Society as it Relates to Hume's Concept of Justice

Passions, therefore, are the basis of morality for the individual. There are, however, certain "universals" that underlie all these individual feelings, that are visible on a societal level. As pointed out earlier, Hume maintains that every human has a certain level of sympathy for other human beings, and this feeling of sympathy is also dominant on the level of the social group. The smaller and more closely related the social group, the stronger the feeling of sympathy is. The more loosely organized and less intimate larger group will have a lesser intensity of sympathy. Sympathy therefore, for Hume, serves as the basis for stability within a social group.

This "moral sense" for Hume is by no means cognitive, it does not give us knowledge. What he basically is trying to do in using the term is to show that it is within this "moral sense" that man can determine his feelings of right and wrong and be able to determine what the goals of conduct should be. Hume would say that what is moral, or what is good, are those actions which tend to promote the greatest pleasure, and conversely, those actions which tend to promote pain would be labeled not good, or immoral.
Hume is quick to point out, however, that one must not equate moral goodness with the greatest pleasure principle; because in determining what is right or wrong, we are looking at the motivations of these actions, not the consequences. Hume (1964) in the same section of the Treatise writes that the consequences are "signs or indications of certain principles in the mind and temper" (II,III,3). By studying the consequences, we can learn about motivation; but it is not a clear indication. One should study the consequences only when one cannot get at the motivations. This is important, says Hume, because it is too easy to identify good actions with well-meaning motives.

What concerns us the most is Hume's account of justice, or the performance of these actions out of a sense of duty or obligation without regard for the consequences. Sibley states that these have come to be equated peculiarly with justice, because it is by no means a certainty whether these actions are performed for our individual happiness or for the public as a whole. Sibley continues with his analysis of Hume's meritorious man that is found in the Treatise, Section II-III,2.

Thus when a man of great merit returns a fortune to a miser or a seditious bigot to whom it belongs, he has presumably acted justly, but it is difficult, to see how he has benefited either himself or the public. The fortune would be better administered by himself than by the miser or the bigot, from the viewpoint of public interest.... Yet we call his act just.
If moral value is somehow related to happiness and happiness to pleasure, how can we account for the fact that we approve his conduct?

Hume answers by saying that through the years, mankind has discovered that on the whole it is better off to distribute economic goods, even in scarcity, than to hold them in common for society or for their group. While men do have a natural feeling of sympathy, it lessens when it is applied to those who are not closely related to us; and as experience has shown, we are "disquieted" when scarce goods are distributed on some overall rule or regulation rather than shared directly.

If the act were an isolated one and unrelated to the whole experience of the human race, it would obviously be better, from the viewpoint of net happiness (pleasure), for the money not to be restored. But when we look at the act in the perspective of the universal primary passions combined with the fact of scarce goods and man's tendency to attenuate his sympathy as its objects become remote, the particular act becomes just. That is to say, while the immediate results of the act may not be useful, the observance of the principles upon which the act is based helps to support the whole body of rules which appear essential for stability and social order. And stability and social order are, given the experience of mankind, essential if either individuals or the group are to gain as much happiness as possible and reduce pain and disquiets to the lowest proportions. (Sibley, 1970, p.417)

Basically, Hume (1963) is trying to show that "public utility is the sole origin of justice, and that reflections on the beneficial consequences of this virtue are the sole foundation of its merit" (Sec. 3, p. 147). Imagine a situation that would look like this:
material goods have ceased to be scarce, while at the same time human beings retained substantially the same nature which Hume has sketched out. Under these circumstances, men would not have to labor for their livelihoods, for nature would provide all they could desire. Would there be any need to provide any partition of goods? (Sibley, 1970, p. 418)

This situation to Hume would be superfluous, for by Hume's hypothesis, everybody already has enough. If there was no injury, there would be no need for property, and if the institutions of property would be superfluous, so would the whole concept of justice.

Another situation, says Hume, would be one where the economic goods continue to be limited but where the "mind is so enlarged, and so replete with friendship and generosity, that every man has the utmost tenderness for every man." In this situation, sympathy would be as strong for everyone as it would be in a closely integrated group. In this situation also, justice would be useless, for the sentiments for those outside the intimate group would be at the same level of sympathy for those inside; thus no form of distribution of economic goods is needed.

The concept of justice is thus instrumental. The basic use is to "procure happiness and security by preserving order in society." Human nature is not egoistic, as with Hobbes, "but a contrast between strength of sentiment and benevolence for those remote from us. It also presupposes human desires for material goods far greater than those which nature and human ingenuity can supply. If either
of these factors be absent, justice is without utility" (Sibley, 1970, p.418).

This analysis of morality and human nature has tried to develop a basis on which further discussion will take place. It divided the study into three main areas:

1. Humans as Self-interested Creatures
2. Humans as Inherently Capable of Goodness
3. Morality Arising from Society and the Conditions of Social Living

In each of these sections, classifications within moral philosophy and psychological theory were attempted. Hobbes is the most famous example of a philosopher who classifies man's human nature as being extremely egoistic; because of this, he also was subject to heavy criticism by later philosophers. Rousseau is the classic example of a philosopher who considered man as inherently good; and his critique of the Hobbesian theory of human nature is well versed and documented. Hume was chosen for the third section because his philosophy is characterized by the understanding of his theory of man in relation to his fellow men.

A further reason for this construction is that for the remainder of this thesis, the moral philosophy and psychology of Rawls will be understood to be a derivation of the thoughts of Rousseau and the psychologist Kohlberg. The standing that this writer will maintain will be one of
Humean background, with proper emphasis placed on his concept of self-interest as a basis for the study of moral psychological principles and also its role in both political and economic theory.
CHAPTER II

RAWL'S VIEW OF MAN IN THE ORIGINAL POSITION

In the book, A Theory of Justice (1971), Rawls has revived the contract tradition in moral and political philosophy. His idea of the social contract, a hypothetical choice situation called the Original Position, was first proposed in an article in 1958, and is presently expressed with much more detail in the book. This is Rawls's most mature conception of how to deal with the question of social choice. He treats this problem of social choice "through the proxy of a specially constructed parallel problem of individual choice, which can be solved by the more reliable intuitions and decision procedures of rational prudence" (Nagel, 1973, p. 220).

It is important to realize that most of social contract theory and utilitarian theory have been designed to offer a reason for all human beings to be moral. Hobbes's account of political obligation tries to show that a political community, based on certain precepts and laws, is in the best interest of each and everyone within the community. Hume's theory of justice points out the difficulty in trying to give a sound reason for the motivation to be just. Hume provides a "justification" for being just based
primarily on utilitarian grounds. Rawls is turning back to the social contract doctrine for his theory of justice. "My aim is to present a conception of justice which generalizes and carries to a higher level of abstraction the familiar theory of social contract as found, say, in Locke, Rousseau and Kant" (Rawls, 1971, p.11).

As a rule, the concept of the social contract in the past was basically the start of civil society, the state of nature. Hobbes and Locke both made psychological assumptions about the individuals in this state of nature. The primary psychological characteristic was that men were basically creatures of self-interest; self-interest being the dominant motivational assumption. In Rawls's contract however, he basically describes a situation in which the principles of justice he has developed are to be chosen; the primary motivational assumption that he proposes is that men in this original position are "mutually disinterested." Combined with the prudential element that Rawls reintroduces in the construction of the "veil of ignorance" and the maximin strategy; Rawls's man will be classified in the original position as being one of limited altruism. Rawls's economic man is based on self-interest and is essentially collectivistic, whereas his political man is represented by a "Sense of Justice" and is essentially individualistic. There are two important places in which Rawls's motivational assumptions take place: the first is
in the original position, the second is in his concept of the "Sense of Justice."
The Original Position

The original position, as stated, is only hypothetical; it is a philosophical tool that serves as a backbone for a more complex and systematic theory. The primary end of Rawls's theory is "justice as fairness," the major principles of which are determined by the unanimous consent of the free and equal parties within the original position. These principles are "general," binding to those who made the contract, known to the public, and when applied to certain situations, must conform with the sense of justice.

The conditions of the original position can be divided into two kinds: those which concern knowledge and those which concern motivation. Knowledge is limited, individuals do not know their social position, their talents, or their "conception of the good." This is what Rawls calls the "veil of ignorance." This introduction of a veil of ignorance represents quite a departure from traditional contractarianism, making Rawls's theory more unique, but also more controversial. This is Rawls's description of the veil of ignorance:

It is assumed, then, that the parties do not know certain kinds of particular facts. First of all, no one knows his place in society, his class position or social status; nor does he know his fortune in the distribution of natural assets and abilities, his intelligence and strength, and the like. Nor, again, does anyone know his conception of the good, the particulars of his rational plan of life, or even the special features of his psychology such as his aversion to risk or liability to optimism or pessimism. More than this, I assume that the parties do
not know the particular circumstances of their own society. That is, they do not know its economic or political situation, or the level of civilization and culture it has been able to achieve. The persons in the original positions have no information as to which generation they belong. (Rawls, 1971, p.157)

What is it, then, that those parties in the original position do know? Again we quote from Rawls (1971):

It is taken for granted... that they know the general facts about human society. They understand political affairs and the principles of economic theory; they know the basis of social organization and the laws of human psychology. Indeed, the parties are presumed to know whatever general facts affect the choice of the principles of justice. There are no limitations on general information, that is, on general laws and theories, since conceptions of justice must be adjusted to the characteristics of the systems of social cooperation which they are to regulate.... (p. 137-38)

The main reason for using this concept of a veil of ignorance is to employ the "notion of pure procedural justice as a basis of theory. Somehow we must nullify the effects of specific contingencies which put men at odds and tempt them to exploit social and natural circumstances to their own advantage" (p. 136). This nullifying "the effect of specific contingencies" leads one to believe that partial and individual interests will have no input.

One other element, to be more thoroughly discussed later, is Rawls's assumption that those in the original position will follow the "maximin rule" or some modified version of it. "The maximin rule tells us to rank alternatives by their worst possible outcomes: we are to adopt the alternative the worst outcome of which is superior to the worst outcomes of the others" (p. 152-53). The meaning
and interpretation of this rule will be clearer when it is discussed with the two principles of justice which emerge from the original position; however, for the present, it is important to note that this maximin rule is where the greatest departure takes place from the utilitarian principle of the greatest happiness for the greatest number. It should be emphasized, one of the main purposes of Rawls's theory is to develop alternative principles of justice than those offered by utilitarianism. Applied to the First Principle, the Priority of Liberty, Rawls maintains that liberty may not be restricted save to secure the maximin liberty possible under concrete circumstances. He also stresses that limitations on liberty and inequalities of freedom cannot be justified on the ground that they promote the general interest. By contrast, the utilitarians are committed to serving the general welfare, not to securing equal or maximin liberty. In applying maximin to the second principle, the Difference Principle would not permit the general interest (to which utilitarianism is committed) to require that the good of some persons be sacrificed in order to secure the greater good of others. According to Rawls, social inequalities are permissible only if the least advantaged group benefits from them. As a result, it can be said that there are two main themes in Rawls's entire work; the development of his own theory of justice, and the destruction of the utilitarian basis for a theory
of justice.

Daniels writes that Rawls uses the original position in two different capacities. First, he uses it as an analytical device; it helps us to "understand the degree to which our conception of justice depends on various formal requirements like the conditions of generality, finality, and publicity that apply to the principles" (Daniels, 1975, p. xix). Another aspect of its analytic abilities, is that it can be used as a model for reducing complex problems, for example, the social choice function of the principles of justice, into more simplified, workable problems, like the rational individual choice of principles. If one is able to reduce problems like this by means of the original position, then it can be effectively utilized as an analytic device.

The second capacity the original position can be used for is as a justificatory device. It is in this capacity that Rawls places greater emphasis, primarily because he wants "the choice of a set of principles in the original position to constitute a justification of them." The question that should be asked, however, is why should people who are not in the original position believe that it is important that those subject to the constraints that Rawls places on them, happen to choose certain principles? Or to put it more succinctly; Why should the choice of these people "justify" our adopting them? Rawls's answer is that the conditions and constraints that are placed in
the original position lead to a "model of procedural fairness," and therefore should be acceptable to everyone. In other words, says Daniels (1975), "Rawls believes he can show there is a justificatory role played by the original position if he can show the moral persons who are parties in it are treated fairly" (p.xix).

The position this paper will take is that the original position can be useful as an analytic device, but as a justificatory device it is subject to severe shortcomings.

The Principles of Justice

Rawls proposes both a "general" and a "special" conception of justice. The general conception deals with Rawls's argument for the maximin criterion — the maximization of the minimin share of the goods to be distributed. The general conception states that:

All social primary goods — liberty and opportunity, income and wealth, and the bases of self-respect — are to be distributed equally unless an unequal distribution of any or all of these goods is to the advantage of the least favored. (Rawls, 1971, p. 303)

The special conception is comprised from the general conception. Rawls develops an argument for two principles of justice, which actually result in three separate claims, all of which derive from, and appear to be based on, the maximin criterion. In the original position, Rawls does not allow for probabilities, but for the adoption of the maximin rule in the social choice function. The three
claims are: the "priority of liberty"; the "difference principle"; and the doctrine of "fair equality of opportunity". This makes up the special conception of justice that Rawls describes this way:

(I) 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.

(II) 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.

Priority Rules:

(I) First Priority Rule (The Priority of Liberty)
The principles of justice are to be ranked in lexical order and therefore liberty can be restricted only for the sake of liberty. There are two cases:
(a) a less extensive liberty must strengthen the total system of liberty shared by all;
(b) a less than equal liberty must be acceptable to those with the lesser liberty.

(II) Second Priority Rule (The Priority of Justice over Efficiency and Welfare)
The second principle of justice is lexically prior to the principle of efficiency and to that of maximizing the sum of advantages; and fair opportunity is prior to the difference principle. There are two cases:
(a) an inequality of opportunity must enhance the opportunities of those with the lesser opportunity;
(b) an excessive rate of saving must on balance mitigate the burden of those bearing this hardship.
The Three Features Favoring the Maximin Strategy

Within the choice strategy of the original position, Rawls argues that there are "three chief features". As mentioned previously, Rawls disdains from using probabilistic strategies; and it is as an argument against these strategies as much as it is an argument in favor of the maximin that Rawls employs these three features. Fishkin, in his article "Justice and Rationality: Some Objections to the Central Argument in Rawls's Theory" (1975), writes that "if maximizing an individual's utility is accepted as an account of his personal good — in preference to the good determined for him by his rational plan — then the principle of average utility turns out to be the principle most likely to secure that good for any individual from the perspective of the original position" (p.616).

Fishkin's statement presents the argument for the average utility principle best exemplified by the economist John C. Harsanyi (1953,1955), and argued for in much more detail in Section Three of this paper. At this time, however, we must look at the first of the three features favoring maximin over the argument for average utility. Quoting Rawls, "the situation is one in which a knowledge of likelihoods is impossible, or at best extremely insecure" (p.154). Rawls argues that this rule is too artificial an assumption to make in a time when so much is at stake. He writes that the calculation of probability from
behind the veil of ignorance would have to be based on the principle of insufficient reason; and that the average utility rule, with probabilities is supposed to be justifiable in the social choice function. Therefore, "the parties discount estimates of likelihoods not supported by a knowledge of particular facts and that derive largely if not solely from the principle of insufficient reason" (p.173).

Fishkin states that the second and third features that Rawls gives concerns what he calls the "qualitative anatomy" of the choice problem. (Rawls, 1971, p. 157)

The person choosing has a conception of the good such that he cares very little, if anything, for what he might gain above the minimin stipend that he can, in fact, be sure of by following the maximin rule. It is not worthwhile for him to take a chance for the sake of a further advantage, especially when it may turn out that he loses much that is important to him. This last provision brings in the third feature, namely, that the rejected alternatives have outcomes that one can hardly accept. The situation involves grave risks. (p.154)

The Thin Theory of the Good

The first feature comes out of the development of the original position. Rawls claims that any particular information which would give an "objective" basis for estimates of probability would also lead toward the further development of special interests. When you develop the probabilities of a person becoming a member of one group or another, you then lead that person to develop principles that would be to the advantage of that group, and also to develop
principles that would lead to disadvantage other groups. Thus the veil of ignorance; an individual in the original position would only have the knowledge of the general facts of human psychology, of economic theory, and of other laws of human behavior; and not have access at all to information concerning society itself, or groups within this society for which he is designing principles. In Section II ("Institutions"), Rawls develops a four stage process through which political institutions are developed in accordance with the principles of justice. As each stage develops, the veil of ignorance is gradually lifted, until, in the final stage of a fully developed well-ordered society, it is lifted entirely. When you reach this level of development Rawls proceeds then to discuss such things as civil disobedience and the majority rule. For the present stage of discussion, the veil of ignorance is as described in the original position, and there is no basis for estimations of probability apart from the principle of insufficient reason which, as mentioned, Rawls claims to be inadequate. The other two features, however, are based on what Rawls (1971) calls "the qualitative structure of the possible gains and losses in relation to one's conception of the good" (p.155). The choice is to be made on the "thin theory of the good," which Rawls says we need "to explain the rational preference for primary goods and to explicate the notion of rationality underlying the choice of principles in the original
position" (Rawls, 1971, p. 397). This thin theory is what provides the basic premises about the rationally self-interested individual about whom nothing in particular is known. The thin theory goes from what Rawls calls the "doctrine of rational plans of life" to the "doctrine of primary goods." It is the doctrine of primary goods on which is based the premises that make it possible to make a choice on the proposed principles. Another premise that Rawls (1971) proposes is the "Aristotelian Principle;" this principle however "does not assert that any particular kind of activity will be preferred. It says only that we prefer, other things equal, activities that depend upon a larger repertoire of realized capacities and that are more complex" (p. 429-430). The Aristotelian Principle will come into focus in this paper, in the discussion of Rawls's moral psychology.

Basically, Rawls's argument is that a rational individual in the original position may try to further his plan for life by providing himself with the primary goods which he will rationally want. Since the good, for Rawls, is the development of a rational plan for life, then the principles guarantee an individual an equal or equitable share of the primary goods necessary for this plan are certain to be in the best interests of this individual. Therefore, it is in this way that Rawls can argue that we can choose the principles that are in the best interests of a person.
without knowing anything about him that does not apply to anyone else.

Rawls also points out that this argument of a person's interests based on this "rational plan" does not depend at all on any tendency of a person to develop some sort of a coherent or predetermined way to plan his life. In Rawls's conception there is no conscious formulation of plans, rather: "a rational plan is one that would be selected if certain conditions were fulfilled. The criterion of the good is hypothetical in a way similar to the criterion of justice" (Rawls, 1971, p. 421).

For Rawls (1971), rational plans are distinguished from the other plans of life in that rational plans conform to certain "principles of rational choice" and also that these rational plans would be chosen by a person with "full deliberative rationality" (p. 408). Rawls continues by saying that the principles of rational choice do not characterize only certain aims as rational but instead "define rationality as preferring,... the greater means for realizing our aims, and the development of wider and more varied interests assuming that these aspirations can be carried through" (p. 413). As for the concept of full deliberative rationality, it says that whatever plan we choose is rational "with the full awareness of the relevant facts and after a careful consideration of the consequences" (p. 408, 416). In summary, then, plans are rational to the
extent that they involve an "effective consideration of means for realizing our aims" and a choice of these aims "made with the full knowledge of conditions and consequences" (Fishkin, 1975, p. 618).

This is, of course, an ideal; in actuality, our plans of life do not reach this level of rationality. Thus, on Rawls's criterion, man is seldom in a position to make a perfect judgment about his own interests, since the requirements are that these plans would be those that are rationally chosen, and not necessarily those that are actually followed.

Finally, man is in a situation in the original position as being as unaware as those in real life as to what the ideal plan is that they would choose with this full deliberative rationality. The veil of ignorance covers up knowledge of this hypothetical rational plan as well as it covers up the knowledge of aims and circumstances. As Fishkin (1975) points out, "they do not know either the plan of life they are to live nor the most rational plan for them under those conditions" (p. 618). The bottom line is that an individual in the original position must choose principles in the interest of maximizing his own personal good without having any idea of the rational plan which determines the sole criteria for the determination of that good.
From the Thin Theory to the Special Conception

Rawls's argument for the three features which support maximin and the general conception are directly related to his claim for the specific primary goods which lay support to the principles in the special conception. And once again, his argument is based upon the thin theory of the good whose purpose is to serve the premises about the primary goods that are required in order to choose the principles of justice.

These premises about primary goods which would lay support to the special conception are of two kinds: first, they concern the amount required of each primary good, and second, they concern the rational worth of each good in a "lexical" order. (Rawls, 1971, p. 42) Supposedly, both of these premises are supported by the thin theory of the good, in general, and specifically, they are supported by the doctrine of rational plans.

The Special Conception

The special conception of justice is made up of three separate principles: the priority of liberty, fair equality of opportunity, and the difference principle. Each one of these is a special case of the general conception applied to one of the three primary goods. An individual in the original position will try to make the smallest share of the good into the largest share; because being behind the
veil of ignorance, he has no conception of whether or not he will be the one who will receive the smallest share. In relating this to the first principle, Rawls argues that no one would want to accept anything less than the greatest equal share of liberty; for the second principle he argues that no one would want to accept anything less than fair equal opportunity; and thirdly, in arguing for the difference principle, no one would want to accept anything less than his share of wealth and income.

Within this total argument for the maximin principle, each of these variants depends on one more assumption, the "threshold assumption." This assumption relates to the primary good and corresponds with the second and third features which supported the general conception. For example, the argument for the first principle depends on there being an equitable share of liberty, or, "the most extensive equal share compatible with an equal share for everyone else", which an agent in the original position would wish to guarantee for himself and below which he would find his share of liberty intolerable. (p.156) This threshold assumption applies in the same way to the level of fair equal opportunities and to wealth or income.

Lexical Order

We have seen now how the three variants of the maximin argument corresponding to the three principles depend on the
applicability of the threshold assumption to the primary goods. Another assumption is developed, however, in order to support the special conception. Each primary good has to correspond to the "lexical priority" which Rawls develops between their three corresponding principles. Rawls (1971) proposes that the principles be ordered like "words in a dictionary ... numerals for letters, putting "1" for "a", "2" for "b" and so on, and then rank the resulting strings of numerals from left to right, moving to the right only when necessary to break ties" (p. 42). Rawls goes on to say that a lexical ordering cannot take place in a continuous real-valued utility function, for it violates the assumption of continuity.

The result of this ordering is that we are required "to satify the first principle in the ordering before we can move on to the second, and the second before we can consider the third, and so on" (Rawls, 1971, p. 43). The first principle has "an absolute weight," meaning that no improvement, however large, a second principle might conceive, can have more than the improvement of the first principle. Likewise; any decrease, however small, the first principle might conceive, has to outweigh a decrease, however large, the second principle might have. This works for the initial principles, or all those that take place before another in the lexical ordering.

When one applies the lexical priority to Rawls's
principles of justice, one can see the far ranging implications this ordering takes. The first principle (the priority of liberty) has lexical priority over the second half of the second principle (the principle of fair equality of opportunity) which in turn, has lexical priority over the first half of the second principle (the difference principle). According to this set up, there has to be a "qualitative division among primary goods" so that liberties have an absolute weight with respect to fair opportunities which have, in turn, an absolute weight with respect to wealth and income.

This absolute weight of each primary good only holds up to a certain extent however. Rawls (1971) states that lexical ordering "presupposes that the principles in the order be of a rather special kind. For example, unless the earlier principles have but a limited application and establish definite requirements which can be fulfilled, later principles will never come into play" (p. 43). What we have called the threshold assumption would appear then to have an added function, that being of allowing each principle to be "fully met" so that the next principle can come into being. (p. 43). As mentioned, liberties have absolute weight with respect to opportunities, but only until the level defined by the liberty principle is reached: "the most extensive total system of equal basic liberties compatible with a similar system of liberty for all." (p. 302)
Likewise, an improvement in the fair equal opportunity principle has absolute weight with respect to wealth and income but only until the threshold level, or the level defined by the second principle is reached. Therefore, "not only must each of the levels be defined for the next principle to come into play, but the absolute weight of each one in turn must be justified for all the conditions which may arise once the veil of ignorance is removed" (Fishkin, 1975, p. 621).

Thus, we can see that the first important place in which there is a motivational assumption working in Rawls's theory is in the original position. The parties must have some motivational assumption because they are being asked to choose from among alternative principles of justice. People like Hobbes and Locke made psychological assumptions as evident in Section I. The main assumption was that self-interest was dominant in human nature, and would go on to say that it might be the mainspring of all action.

Presently, Rawls states, we have a situation familiar in social theory. "That is, a simplified situation is described in which rational individuals with certain ends and related to each other in certain ways are to choose among various courses of action in view of their knowledge of the circumstances. What these individuals will do is then derived by strictly deductive reasoning from these assumptions about their beliefs and interests, their
situation and the options open to them. Their conduct is, in the phrase of Pareto, the resultant of tastes and obstacles" (p. 119). It is within this situation that one would like to prescribe a strictly deductive argument for the principles of justice. Rawls does not provide such a deductive argument, but he has briefly offered something that resembles it. His argument has received heavy criticism; the basic critique being of his description of the participants in the original position as being put in a position that is extremely unrealistic. This is true; however, one must remember that Rawls did not intend his original position to be realistic. It is intended to provide a basis for his principles that will be chosen on the basis of minimal assumptions of human motivation, and where the choice is made within a situation in which the constraints of having a morality are enforced.

It is clear, then, that the original position is a purely hypothetical situation. Nothing resembling it need ever take place, although we can by deliberately following the constraints it expresses simulate the reflections of the parties. The conception of the original position is not intended to explain human conduct except insofar as it tries to account for our moral judgments and helps to explain our having a sense of justice. Justice as fairness is a theory of our moral sentiments as manifested by our considered judgments in reflective equilibrium. These sentiments presumably affect our thought and action to some degree. So while the conception of the original position is part of the theory of conduct, it does not follow at all that there are actual situations that resemble it. What is necessary is that the principles that would be accepted play the requisite part in our moral thought and action. (Rawls, 1971, p. 120)
There are two parts of this statement that deserve repeating; first, "the constraints it expresses simulate the reflections of the parties;" and secondly, "the principles that would be accepted play the requisite part in our moral thought and action." (p.120) It is on these grounds that criticism of the original position can be well-founded.

There is, however, one question that this entire argument does not answer. That is, Why will human beings act as justice requires? If the motivational assumptions of the original position were all we knew about human motivation, there is no indication that men would lead a just life outside the original position. For Rawls states:

For the fact that in the original position the parties are characterized not as interested in one another's concerns does not entail that persons in ordinary life who hold the principles that would be agreed to are similarly disinterested in one another. Clearly the two principles of justice and the principles of obligation and natural duty require us to consider the rights and claims of others. And the sense of justice is a normally effective desire to comply with these restrictions. The motivation of the persons in the original position must not be confused with the motivation of persons in everyday life who accept the principles that would be chosen and who have the corresponding sense of justice. In practical affairs an individual does have a knowledge of his situation and he can, if he wishes, exploit contingencies to his advantage. Should his sense of justice move him to act on the principles of right that would be adopted in the original position his desires and aims are surely not egoistic. He voluntarily takes on the limitations expressed by this interpretation of the moral point of view. (Rawls, 1971, p. 147-148)

The question to be asked is: How does Rawls avoid the exact same problem that Hobbes had — Why will people who would choose the principles of justice in the situation
described by Rawls act in the ways required by his principles of justice when the limitations expressed in the original position are taken off? If we understand the principles of justice as earlier described (assuming here that Rawls's argument for the principles is valid), the question that still must be asked is: Why will men do what justice requires? Or, perhaps better for our purposes: What are the reasons for acting justly?

The answer to this question is quite complex, and parts of the answer can be found in various parts of A Theory of Justice (1971). The portion that will be emphasized in this analysis is the chapter entitled "The Sense of Justice," which is an amplified version of an earlier article by the same title. (Rawls, 1963) It is here that we find the second important place in which we can determine the motivational assumptions of Rawls's theory. It will be looked at in detail in Chapter IV.
The Concept of Morality and Social Choice Functions

Richards in *A Theory of Reasons for Action* (1971), states that the concept of morality should be stated as an "ideal agreement or contract which rational men in a certain defined situation would unanimously make" (p. 79). This concept is very similar to what several economists and political scientists have suggested as a way to theorize about the concepts of morality and justice. For example, Harsanyi (1955), who has developed a more acceptable form of the utilitarian argument, suggested that a way of understanding impersonal value judgments of right distribution is by comparing them to the choice of principles governing the distributive effects of certain institutions which each rational man would agree to when he was in complete ignorance of what his own relative position (and the position of those near to his heart) would be within the system chosen. Harsanyi goes on to say that in such a situation under the constraint of uncertainty, rational egoistic men would adopt a Laplacean strategy, which would result in their agreement to a principle which maximizes average, not total utility.

In this same vein, Buchanan and Tullock (1962), in their book *The Calculus of Consent*, try to make clearer the choice function, specifically the choice of a constitutional voting procedure like majority rule, in terms of the unanimous choice of self-interested rational maximizers from an original position in which each person is "uncertain as
to what his own precise role will be in any one of the whole claims of later collective choices that will actually have to be made." By setting up the argument this way: a choice of institutions under uncertainty, when the choosers can only assume random distribution in terms of the effects of the particular institution, Buchanan and Tullock (1962) claim to have delineated a view where "the purely selfish individual and the purely altruistic individual may be indistinguishable in their behavior" (p. 314). For Buchanan and Tullock, the self-interest assumption serves an empirical function, for the only way to determine whether an assumption of this sort is realistic is to compare some of the positive analytical implications of the theory with observable real-world facts.

The conceptions of Hume best coincide with the views of Buchanan and Tullock. Their basic analysis of the individual calculus that is involved in the choosing among alternatives has demonstrated that it will often be to the rational self-interest of the individual to select a particular rule that can be predicted to produce results on occasion that run counter to the self-interest of the individual calculated within a shorter time span. By shifting the choice backward from the stage of the specific collective decision to the stage of the constitutional decision, we have been able to incorporate the acquiescence of the individual to adverse collective action into a calculus that retains an economic dimension and that can be analyzed in non-moral terms. (Buchanan and Tullock, 1962, p. 314)

Hume's conception was similar in that he was able to ground political obligation, not on moral principle or
social contract, but on self-interest. The self-interest of each individual in the community dictates the observance of conventional rules of conduct; and these rules are necessary for both political and social order. It is to the initial interest of parties to agree on conventional rules if such rules do not exist, but it is also to the interest of individuals to abide by the conventional rules already in existence. For Hume, a person's self-interest would best be served by all other people adhering to the conventional but being free to violate these rules himself. This position is very similar to the position taken by Rawls; but what Rawls fails to realize, and what Hume does realize, is that because these rules are socially derived, they must apply generally. In order to maintain some degree of social order and stability, each individual must recognize the fact that if he were free to violate conventional rules, others must also be free; and upon recognizing this chaotic state of affairs, the individual will rationally choose to accept restrictions on his own behavior.

The notion of unanimous agreement, according to Richards (1971, p.79) is meant "to express the basically distributive character" of the concept of morality, where each person's interests are given equal weight, and where no person's interests are sacrificed for those of another (unless that person consents). When the contract is conceived of in
this way, it can be usefully compared with the notion of the ideal observer, which basically views the concept of morality as depending on the approval of one person, under the certain ideal conditions of perfect sympathy, impartiality, full-knowledge, etc. This conception has been used by Hume, specifically in Book III of the Treatise on Human Nature (1964). As stated in the first chapter, Hume's concept of morality suggests that the choice of principles depends on whatever principles such a "perfectly sympathetic observer finds most pleasurable, after sympathetically identifying with the pleasure and pains of all affected." Therefore, Hume ends up with a form of the utilitarian principle, since the ideal observer identifies with the maximin amount of pleasure over pain.

When the rational contractors of the original position consider which principles they will adopt, they will have put before them the circumstances which are likely to arise in their actual lives; a set of circumstances concerned with problems of justice.

In trying to analyze Hume's view, one can translate it into a problem of rational choice. A choice, in which an individual is trying to develop a system of rules which must satisfy "certain constraints on their generality and lack of complexity," so that people can easily understand and be able to apply the rules, which, when they are accepted, would lead to a certain stability, a system of
cooperation which develops under the circumstances of justice.

In looking at Hume's theory of governance, many writers, for example Hobbes and Locke, have suggested that men surrendered power to the magistrates in order to gain protection from other men. According to Hume, man's main reason for surrendering power to the magistrates is in order to protect him against himself. It is because of man's tendency to be short-sighted and passionate, a tendency he is unable to alter, to prefer near to distant advantages, that he does not act on rules of justice, and so harms himself by depriving these rules of any utility. For this same reason, man needs a magistrate to enforce the rules, and to decide impartially when they have been broken. The fact that man needs a magistrate because he foresees that other men will also break the rules is, according to Hume, only secondary.

This aspect of Hume can be subjected to criticism; it is safe to say that all men need rules of justice to be enforced, both for the short-term reason that they need to be protected against themselves and other people, and for the long-term reason that these rules must be obeyed by everyone if they are to be of advantage to anyone. Hume is correct however in stating that it is in man's long-term interest to have rulers who impose justice on man.

For Hume, all those governments are legitimate which
manage to survive over a long period of time. A sense of loyalty develops along with this legitimacy and leads to stability. Hume is not clear as to whether this sentiment of loyalty precedes the firm establishment of a government or whether its stability comes prior to the sentiment. There is quite possibly an interaction: tendencies to stability evoke feelings of loyalty, and feelings of loyalty promote stability.

It is Hume, who sees the legislator as a sort of governmental ideal observer, who would seem to have the correct solution to such problems of rational choice, especially when they deal with the creating of rules to stabilize the already existing rules of justice. (Hume, 1964, p. 502ff) Hume argues that the legislative power will not always be able to change the positive law. He lists five general principles in the Treatise from which governments can derive authority: (1) long possession; (2) present possession; (3) conquest; (4) succession; and (5) public laws. For Hume, the legislative power is one step removed from the principles, and the positive law is two steps removed from them. One could suppose that the legislative power would quite easily be able to change the positive law. However, what Hume has in mind is that, when the legislative power is changed, it will not always be able to change the positive laws. The force that the legislative power receives from the above principles will not necessarily be
transformed to any new law it may make, because of the extra step in the passage from the original principles, to the legislative power, to the authority of the positive law. The basic premise is that positive law retains some authority of its own, and is not subject to change upon the appearance of a new legislative power. The United States constitution is an example of this.

In so far as the correct solution to such a decision problem will best satisfy the desires for all, people will tend to support that government, so that, when there is an ideal government, there will be what Richards (1971) calls this "stable, self-supporting feedback mechanism" (p. 129). A mechanism of this sort is very similar to what Downs has offered in his analysis of the function of government.
CHAPTER III

THE MAXIMIN PRINCIPLE AND THE PRINCIPLE OF AVERAGE UTILITY

Rawls is chiefly concerned to defend his principles of justice; he claims that his principles take precedence over any others, and the result is something that strongly differs from utilitarianism. In his Priority of Liberty Principle Rawls maintains that liberty may not be restricted save to secure the maximin liberty possible under concrete circumstances. Rawls stresses the fact that limitations on liberty and inequalities of freedom cannot be justified on the fact that they would serve to promote the general interest, say by improving the overall aspect of life. Utilitarianism, on the other hand, is concerned only to serving the general welfare, not to serving equality or maximin liberty. In certain circumstances, utilitarianism could justify restrictions on freedom. Although there are many instances in which Rawls diverges from utilitarianism, there are on occasion areas in which these utilitarian conceptions are used for the advancement of his own theory; for example, the use of market economic mechanisms.

It is sometimes thought that normative theories simply "account for" certain more specific judgments or principles, and Rawls can be partly interpreted in this way. Rawls

66

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
believes that his principles are most congruent with the strong moral convictions that are present in man. Admitting that the congruence argument cannot be conclusive, Rawls constructs a more direct "contractarian" approach, stating that according to this argument, those principles of justice are those principles which "free and rational persons concerned to further their own interests would accept in an initial position of equality as defining the fundamental terms of their association" (Rawls, 1971, p.11).
The maximin rule is mainly an argument for maximin as against what Arrow would call the sum-of-utilities criterion. It should first be noted, however, that the concept of an original position had previously been put forth by Harsanyi and W.S. Vickrey (1960); they used it, however, to supply a contractarian foundation to a form of utilitarianism. They started from a position set by J. von Neumann and O. Morgenstern (1947), that choice under risky conditions can be described as the maximization of expected utility. In the original position (Harsanyi and Vickrey did not call it this), each individual may with equal probability be any member of society. Therefore, if there were n members of society and if the ith member has $u_i$ under some given allocation decision, then the value of that allocation to any individual can be described as $\Sigma u_i(1/n)$, since $1/n$ is the probability of being individual i. As a result, when each individual is choosing among alternative allocations of goods, they will want to maximize this expectation, or looking at it for an entire population, maximize the sum of utilities. (Arrow, 1973, p. 250)

Rawls, starting from the same premises, denies the statement that society should maximize $\min u_i$. His position can be analyzed in two parts. First, that in the original position, it is reasonable to have a high degree
of aversion to risk, and being concerned with the worst possible outcomes is most definitely an extreme form of risk aversion. The second part is that the probabilities are not well-defined and therefore should be eliminated from considerations within the calculations. This first point raises some questions about the meaning of the utilities, and as Arrow suggests, does no justice to the fact that, at least in Narsanyi and Vickrey, the utilities are already so measured as to reflect risk aversion. The second part talks about unresolved controversy in the theory of behavior under uncertainty; are all uncertainties expressible by probabilities? There have been many replies to both sides of the question; Arrow and Hurwicz (1972), for example, have given a set of axioms which imply that choice will be based on some function of both maximin and minimum utility.

Arrow states that for a long time people have found the implications of the maximin theory unacceptable. It implies that any benefit, however small, to the worst-off member of society, will outweigh any loss to a better-off individual, provided that it does not reduce the second below the level of the first. To show that the difference principle, a direct result of the maximin strategy, would lead to unacceptable moral implications, let us take example number one.
First Example

Consider a society consisting of one doctor and two patients, both of them critically ill of pneumonia. Their only chance to recover is to be treated by an antibiotic, but the amount available suffices only to treat one of the two patients. Of these two patients, individual A is a basically healthy person, apart from his present attack of pneumonia. On the other hand, individual B is a terminal cancer victim but, even so, the antibiotic could prolong his life by several months. Which patient should be given the antibiotic? (Harsanyi, 1975, p. 596)

Using the difference principle criteria, the antibiotic would go to the cancer victim, since he is the less fortunate of the two patients. By contrast however, as Harsanyi states, utilitarian ethics would suggest the opposite. The antibiotic should not be given to individual B but to individual A, since it would do individual A much more good in bringing him back to total health, and probably a long life, than it would to individual B, whose life at best would only be prolonged for a few extra months. Therefore, there can easily exist medical procedures which prolong the life of one individual for a couple of months, at huge expenses to a society; but as a result the rest of society suffers abject poverty. A maximin principle would put forth such a procedure.

Second Example

Consider a society consisting of two individuals. Both of them have their material needs properly taken care of, but society still has a surplus of resources left over. This surplus can be used either to provide education in higher mathematics for individual A, who has truly exceptional mathematical ability, and has an all-consuming interest in receiving
instruction in higher mathematics. Or, it could be used to provide remedial training for individual B, who is a severely retarded person. Such training could achieve only trivial improvements in B's condition (e.g., he could learn how to tie his shoe-laces); but presumably it would give him some minor satisfaction. Finally, suppose it is not possible to divide up the surplus resources between the two individuals. (Harsanyi, 1975, p. 596)

Here again, the difference principle would require that the surplus of resources be spent on B's remedial training, since he is the less fortunate of the two. Once again, in contrast, utilitarian theory would suggest that the surplus resources be spent on A's education, since this would accomplish "much more good" as a whole, and would also result in greater individual satisfaction.

What is possibly more disturbing, as Harsanyi points out, is the fact that the difference principle would require absolute priority to the interests of the worst-off individual, no matter what, under even the most extreme of circumstances.

Rawls is aware of these arguments, but rejects them on the ground that they will never happen in real life circumstances. What he says is that he assumes that in actuality, a society has what he terms close-knittedness:

As we raise the expectations of the more advantaged the situation of the worst-off is continuously improved. Each such increase is in the latter's interest, up to a certain point anyway. For the greater expectations of the more favored presumably cover the costs of training and encourage better performance thereby contributing to the general advantage. (Rawls, 1971, p. 158)

In oversimplifying the argument, this entire
conceptualization appears to be false; it just does not work this way. It is easy to point out changes that benefit the wealthy at the expense of the poor, the well-off against the least advantaged; for example, cutting income taxes for those in higher brackets and also cutting welfare payments for the poor. Rawls asserts that one is supposed to look at his principles in their totality, not simply offer obscure counterexamples. As an example, Rawls strongly expresses a demand for open access to all positions; but, even with perfectly fair equality of opportunity, there will remain inequalities due to various reasons, including biological and cultural inheritance. There is also the possibility of just pure chance. The result is, however, that once these inequalities exist, which they will, this "harmony of interests" will cease to exist. The result, as Arrow (1973) so rightfully points out, is that Rawls's assumption of "close-knittedness" undermines all the distinctions that he has so carefully made. For if it holds, there is no difference in policy implications between the maximin principle and the sum-of-utilities; "if all satisfactions go up together, the conflict between the individual and society disappears" (p. 251).
The Principle of Average Utility

Harsanyi and Brandt (1972), among others, have argued that the maximization of expected average utility is the only possible choice that could be made by a person using Rawls's decision procedure. Harsanyi (1975) states that people in the original position would first assign an equal probability to their occupying each particular position in society, and then select the position or arrangement that has the highest average expected utility. To determine these utility assignments, people in the original position, or Harsanyi's equivalent of it, are supposed to compare what it would be like to possess their particular distributive share in society while also possessing the subjective tastes of persons who have those shares. Two concepts which can be compared to this are Hare's universalizability criterion and also Hume's impartial spectator. Harsanyi would assume then, that the people, with the appropriate knowledge of the psychological laws and also knowledge of the appropriate information, would arrive at the same utility judgments, which means that they would be able to determine which arrangements best maximize their average expected utility.

Harsanyi supposes that the Laplacean principle would be used, and it is quite easily seen how this would result from the contractual conception of the original position. For example, if a society consisted of four standard
positions — landowners, soldiers, farmers, and rulers — each of which has a different utility value \( (u_i, u_s, u_f, u_r) \), then, with this strategy, the contractors in the original position would choose the principle which maximizes 
\[
\frac{1}{4}u_i + \frac{1}{4}u_s + \frac{1}{4}u_f + \frac{1}{4}u_r,
\]
which is the same as saying 
\[
\frac{1}{4}(u_i + u_s + u_f + u_r).
\]
In mathematical terms it describes a principle which maximizes the average utility — Harsanyi's modified utilitarian principle. Along these same lines, Buchanan and Tullock also use the contractual notion suggested by this Laplacean strategy in *The Calculus of Consent* (1962). They assume that the contractors are randomly distributed among the various positions which are possible under the voting rules they consider, like the majority and the minority.

One of the important aspects of Harsanyi's model is this:

In real life, when people express a preference for one social arrangement over another, they will often have a fairly clear idea of what their own personal position would be under both. Nevertheless, we can say that they are expressing a moral value judgment, or that they are expressing a moral preference for one of these social arrangements, if they make a serious effort to disregard this piece of information, and make their choice as if they thought they would have the same probability of taking the place of any particular individual in society. (Harsanyi, 1975, p. 598)

What is important is that under this model, there will be two sets of preferences. There will be for the contractor a set of personal preferences, which may give consideration

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
to things that are of interest to himself, and maybe the immediate family; and also a set of moral preferences, which are preferences based on the attempt to give the same weight of interests to every member of society. Thus, the principle of average utility. It is here then that we can see that Rawls bases his moral theory on the contractarian position, and Harsanyi's model yields a moral theory based on the principle of average utility, which clearly belongs in the utilitarian tradition.

Within the original position, Rawls argues that these limited conditions account for the emergence of the rules of justice as fairness. What has been argued so far is that they do not; additional assumptions about the nature of man have to be made, and even then, justice as fairness would not be regarded as the inevitable choice of rational men in the original position.

The main assumption has been that men are mutually disinterested. On the question of interest itself, Rawls knows that while he can prevent men from having particular desires or interests within the original position, he cannot take all the aspects of desire and interest away from them. Rawls (1971) states that "human actions ... spring from existing desires" (p. 568). But, the judgment on alternative rules of justice are made "solely on the basis of what seems best calculated further ...interests" (p. 584).
Men then, as a result, have interests, but not particular interests; they have a desire, but they do not realize which desires they actually possess.

The question that has been brought out in many critiques of this aspect is this: is it possible to conceive of men as having a hypothetical knowledge of what it means to have interests and desires without having particular interests and desires? Barber (1975) in his critique suggests that "mutually disinterested men might turn out to be uninterested men, men incapable of comprehending the meaning of interest" (p. 664). Rawls states this to a certain extent himself when he suggests that "some may object that the exclusion of nearly all particular information makes it difficult to grasp what is meant by the original position" (p. 138). What Barber would argue is that at the level of psychology, particularity is built into the notion of interest, and cannot be taken away without rendering all interest, on the whole, unintelligible.

It is presumably in the sense of primary goods that men in the original position weigh options in furthering their interests. But this does not answer the question at all of whether interest is intelligible at the non-existent level of particularity that Rawls offers; for the interest men take in primary goods can only be explained in terms of the potential interest they have in particular ends. One other comment on Rawls's primary good is that
he renders it so generous as to lose its practical value. It not only encompasses such things as opportunities, wealth and power; but also rights, liberty and self-respect.

As for the social choice function, can it be safely assumed that the contractors in the original position are not "influenced by different attitudes towards risk and uncertainty, or by various tendencies to dominate or to submit, and the like?" (Rawls, 1971, p.530). Nothing in the original position suggests maximin as the only rational or most rational solution to the social choice function.

The question of which strategy would be the most rational cannot be settled without further knowledge about attitudes toward risk and uncertainty which are not given in the original position. Rawls acknowledges that not merely are particular prospects uncertain but that they are also unpromising; not merely that particular statuses are applied by equal probability, but that they are assigned by enemies. If one can assume that friends would assign a favorable status, then it is no less rational to say that men would pursue a moderate-risk strategy whose aim would be to create the best possible outcome, or at least the possibility of it. One of the conditions of society is that of moderate scarcity; and in this position, man would be only too willing to gamble on a strategy. For if he were to win, he would be in a far better position than he would be in a position of austere egalitarianism. This
would be even more true if the contractors would regard his losses as relatively insignificant compared to the alternative, a minimum below that which maximin strategy guarantees he will not fall. One need not enter into a full empirical study of the gambling nature in man to argue that there are people who are willing to risk quite a bit for the chance to be rich and famous. Were one to present a historical analysis of actual developments in human society, one could probably suggest that the no-risk strategy of Rawls would be atypical of human choice in a position of uncertainty.

What Rawls has done is that in arguing for security, he is implicating a substantive special psychology, a conservative one at that. This feeling for security has been installed where rationality is supposed to be. Man in the original position is unwilling to enter a situation that promises success because it also promises failure. He is content with security and the knowledge that he will be no worse-off than anyone else because he does not want to risk freedom. This is Rawls's conception of rationality.

Let us recall the fact that the contractors supposedly choose this maximin strategy not out of altruism or benevolence or social responsibility, but solely in order to protect themselves in the pursuit of their own interests. Quoting Rawls (1971):
The theory of justice assumes a definite limit on the strength of social and altruistic motivation. It supposes that individuals and groups put forward competing claims, and while they are willing to act justly, they are not prepared to abandon their interests. (p. 281)

It can be argued that men who by nature are more spirited and optimistic, freed of the moral and altruistic constraints that Rawls does place on them, would choose to be less cautious, more willing to gamble. What is needed is for man to be motivated by a mixture of altruistic and egoistic motivation that will be spelled out in Chapter V; a man that would pursue his own interests with more vigor, and less caution. The contractors would still weigh the alternatives and accept the risks, and would be no less rational for doing so.
CHAPTER IV

RAWLS'S MORAL PSYCHOLOGY

The second major area in Rawls's theory where one can find motivational assumptions is in the section entitled "The Sense of Justice." For Rawls, it is important that a conception of justice should relate to actual human motivation in such a way that a "well-ordered society" will be stable with respect "to the justice of its institutions and the sense of justice needed to maintain this condition." He goes on to say that "inevitably, we shall have to take up some rather speculative psychological questions; but all along I have assumed that general facts about the world including basic psychological principles are known to the persons in the original position and relied upon by them in making their decisions" (Rawls, 1971, p. 456).

When one considers the importance of the concept of "the sense of justice" in Rawls's theory, it is surprising that it is also one of the least developed; in fact, Rawls provides little information about the nature of this sense of justice. The following paragraph essentially provides his explanation:

Let us assume that each person beyond a certain age and possessed of the requisite intellectual capacity develops a sense of justice under normal social
circumstances. We acquire a skill in judging things to be just and unjust, and in supporting these judgments by reasons. Moreover, we ordinarily have some desire to act in accord with these pronouncements and expect a similar desire on the part of others. Clearly this moral capacity is extraordinarily complex. To see this it suffices to note the potentially infinite number and variety of judgments that we are prepared to make. The fact that we often do not know what to say, and sometimes find our minds unsettled, does not detract from the complexity of the capacity we have. (Rawls, 1971, p. 46)

This discussion raises more questions than it answers. This feeling that some things are just and others are not, and a sense of duty to act justly is familiar to all men. But for such an important aspect of his theory, Rawls should have analyzed this a bit further. The question that will be asked in this section is how does Rawls relate the two parts of his theory — the argument for the principles of justice and the psychological theory of how human beings acquire a sense of justice. First, however, we must look at the derivation of Rawls's concept and the principles of moral psychology that he offers.
Rawls's Moral Psychology

The concept of the sense of justice is one that was first developed by J. S. Mill in Chapter III of *Utilitarianism* (1979). Mill's account of the sense of justice is based on a double origin; sentiments arise "not only from sympathy but also from the natural instinct of self-protection and the desire for security" (Rawls, 1971, p. 502). Rawls states that his contract doctrine does the same thing, but it does not do it by weighing two tendencies — egoism vs. altruism, but by a theoretical construction based on the appropriate reciprocity principles.

Rawls grounds his argument for the principles of justice on ideas about human nature and human development; in fact, one could say that it is dominated by a "strong developmental conception of human nature." Rawls emphasizes the fact that men develop plans of life to accord with their unique tastes, abilities, and opportunities. Yet he also believes that men have "natural sentiments of unity and fellow feeling." Following Mill, Rawls (1971) states that the natural end of development is a condition in which each person feels a certain unity with his fellow beings. "He desires to know that his aims and theirs are not in opposition, that he is not setting himself against their good but is furthering what they really wish for" (p. 502). Rawls, then, has a view of man based on the capacity for extensive individuality and community. However, Rawls's
theory "assumes a definite limit on the strength of social and altruistic motivation. It supposes that individuals and groups put forward competing claims, and while they are willing to act justly, they are not prepared to abandon their interests" (p. 281). Therefore, the individuality of men prevents perfect identification with their fellows.

In Rawls's theory, man is a social being. Individuals feel a bond of unity, a feeling that they share in a common endeavor, which leads to feelings of affection and devotion. Therefore, we can establish two general aspects of social nature: a "social interest" in our environment and "group" or "fellow feeling." Rawls's theory of man possesses both of these aspects.

This sense of fellow feeling is central to Rawls's moral psychology. Rawls states that a devotion to just institutions and principles arise out of our capacity to form relationships with our family, friends, and associates. As individuals begin to realize that those for whom they care benefit from just institutions and practices, a corresponding sense of justice develops. As a result, although these bonds may not form relationships between all people, in Rawls's well-ordered society these bonds are sufficiently extensive enough to where someone who would act unjustly would injure someone for whom we care. Thus this feeling of social interest supports Rawls's concept
of social stability by making individuals see that just conduct is part of their own good. (p. 490-91; 570-71)

For Rawls, social relations are not merely instrumental but are expressions of a deep psychological craving for a form of intercourse with our fellow beings. As a result, Rawls criticizes a purely market conception of the social order in which the social "network" is merely a tool for securing the private ends of individuals. (p. 520)

Developmentalism: The Rousseau Connection

Rawls's theory of human nature can be called "developmental," as it emphasizes "man's possession of innate potentialities and the conditions necessary for their realization" (Gaus, 1981, p. 63). According to a psychology of this kind, human capacities can be developed or obstructed by institutions or beliefs. Rawls goes beyond this to say that men not only will develop in certain directions, but that when natural development does not occur, certain pathologies will take place.

In many places in his theory, Rawls states that social interest, fellow feeling, and a drive for personal excellence are all natural to man. (Rawls, 1971, p. 502; 426-33) The key to his theory is his ideas on self-respect.

A plan that does not utilize our natural capacities in an interesting way, and thus flouts the Aristotelian Principle, will fail to instill in us feelings of competence and worth. But a sense of self-worth ... cannot be achieved merely by developing
Rawls goes on to argue that this sense of fellow feeling helps protect self-respect by providing support against self-doubt when personal setbacks occur.

Rousseau's Moral Psychology

Within Rawls's original position, we unanimously choose the principles of justice as to our natural capacities and our overall social advantage. From the moral point of view, these principles are undeserved and hence arbitrary; for Rawls would state that the principles would meet the criteria of both collective and individual rationality and so serve to explain the sense of justice.

The connection between Rousseau and Rawls is in their concepts of the ideal man and moral psychology. There are major differences between the theories of the two in that Rawls has not abandoned liberalism as Rousseau had done, and there are also non-egalitarian elements in Rawls which are not present in Rousseau's position.

Chapman, in his article "Rawls's Theory of Justice" (1975), views Rawls's theory as the culmination of the effort, begun by Kant and Hegel, and carried forward by Green, Bosanquet and others to adapt Rousseau's theory of the general will to the modern state. Quoting Rawls (1971): "a well-ordered society satisfies the principles of
justice which are collectively rational from the perspective of the original position; and from the standpoint of the individual, the desire to affirm the public conception of justice as regulative of one's plan of life accords with the principle of rational choice" (p. 577). Therefore, the general will is present in both the individual and society.

As Rousseau would say, man is "naturally good," and what reason requires, natural and moral sentiments will emerge to confirm and sustain. The major derivation of Rawls's position comes from Rousseau's moral psychology; given the neutralization of egoistic impulses, the morality of man can be properly developed. Chapman (1975) divides "these interdependent processes of psychological neutralization and moral transformation" through "three Hegelian phases."

First, the original position, in which the Kantian noumenal selves confront one another to cooperate against scarcity.

Second, is an institutional phase, in which their exposure to just institutions calls forth men's sense of justice; it is not a matter of "inculcation", as Barry would have it.

Third, or better dimension ... is a familial phase in which our capacity for "reciprocity", to which Rawls refers as "a deep psychological fact" (p. 494), may be initially aroused. (Chapman, 1975, p. 589)

Some critics have argued that Rawls has illegitimately enlarged his concept of human rationality. Once men are
outside of the original position their rationality is a matter of prudential deliberation and decision. Chapman asks this question of the reciprocative rationality: "Does the same conception of rationality apply to men, or Kantian selves, who set out collectively to select principles that will morally transfigure them, and which they do not know, and to these men once they have acquired their sentiment of justice?" The question of this transition from rational calculation to rational morality can be compared to the problem that Rousseau had in his explanation of the moral consequences of the social contract. Rousseau's general will was held to be both a rational standard against which to test just institutions and also a steady "inclination of men who had attained to 'natural goodness'".

Chapman (1975) states that what Rousseau and Rawls mean to say is that our natural sentiments are as rational as the capacity we have for prudential calculation.

Rawls adopts Rousseau's position in his view of human development, which includes natural response to good will, a feeling of gratitude; what will be termed reciprocity. Where Rawls's theory becomes more "elaborate" than Rousseau's is in the development of the "Aristotelian Principle" in his theory of motivation, the feeling that people find satisfaction in extending themselves, to take on the complex; something that can be found only in an advanced society.
The cognitive developmental theory of moral psychology most recently has been characterized by the works of Piaget and Kohlberg. This theory proposes that "moral development passes through invariant qualitative stages, and that the stimulation of moral development, like other forms of development, rests on the stimulation of thinking and problem solving by the child" (Kohlberg, 1980, p. 57).

The cognitive developmental approach starts from a different view of morality than either common sense or most psychology. The overall theory claims that morality represents a set of rational principles that are valid for each culture or society: the principles being human welfare and justice. In Kohlberg's theory the rules or regulations developed are more or less arbitrary, and on a whole, there are only a few ethical principles, and these are universal. One must remember, however, that a rule is not the same as a moral principle. "Thou shalt not commit adultery" is a rule for behavior in a certain situation, while Kant's categorical imperative is a principle; it is not a prescription or rule for behavior, but a guide for choosing among behaviors. It is free from "culturally defined content"; and it transcends social laws and therefore can be applied universally.

Related to Kant's categorical imperative and central
to the development of moral judgment is the principle of justice. Justice, defined by Kohlberg (1980), and accepted by Rawls as being the "primary regard for the value and equality of all human beings and for reciprocity in human relations, is a basic and universal standard" (p. 57).

In Kohlberg's theory, individuals acquire and refine a sense of justice through a "sequence of invariant developmental stages." The stages (Kohlberg, 1971, p. 164-165) are:

1. Preconventional level

At this level the child is responsive to cultural rules and labels of good and bad, right or wrong, but interprets these labels either in terms of the physical or the hedonistic consequences of action (punishment, reward, exchange of favors) or in terms of the physical power of those who enunciate the rules and labels. The level is divided into the following two stages:

Stage 1: The punishment-and-obedience orientation.

The physical consequences of action determine its goodness or badness regardless of the human meaning or value of these consequences. Avoidance of punishment and unquestioning deference to power are valued in their own right, not in terms of respect for an underlying moral order supported by punishment and authority (the latter being Stage 4).

Stage 2: The instrumental-relativist orientation.

Right action consists of that which instrumentally satisfies one's own needs and occasionally the needs of others. Human relations are viewed in terms like those of the marketplace. Elements of fairness, of reciprocity, and of equal sharing are present, but they are always interpreted in a physical, pragmatic way. Reciprocity is a matter of "you scratch my back and I'll scratch yours," not of loyalty, gratitude, or justice.

2. Conventional level
At this level, maintaining the expectations of the individual's family, group, or nation is perceived as valuable in its own right, regardless of immediate and obvious consequences. The attitude is not only one of conformity to personal expectations and social order, but of loyalty to it, of actively maintaining, supporting, and justifying the order, and of identifying with persons or groups involved in it. At this level, there are the following two stages:

Stage 3: The interpersonal concordance or "good boy-nice girl" orientation.

Good behavior is that which pleases or helps others and is approved by them. There is much conformity to stereotypical images or what is majority or "natural" behavior. Behavior is frequently judged by intention — "he means well" becomes important for the first time. One earns approval by being "nice."

Stage 4: The "law-and-order" orientation.

There is orientation toward authority, fixed rules, and the maintenance of the social order. Right behavior consists of doing one's duty, showing respect for authority, and maintaining the given social order for its own sake.

3. Post conventional, autonomous, or principled level.

At this level, there is a clear effort to define moral values and principles that have validity and application apart from the authority of the groups or persons holding these principles and apart from the individuals' own identification with these groups. This level again has two stages:

Stage 5: The social-contract legalistic orientation, generally with utilitarian overtones.

Right action tends to be defined in terms of general individual rights, and standards that have been critically examined and agreed upon by the whole society. There is a clear awareness of the relativism of personal values and opinions and a corresponding emphasis upon procedural rules for reaching consensus. Aside from what is constitutionally and democratically agreed upon, the rights are a matter of personal "values" and "opinions." The result is an emphasis upon the "legal point of view," but with an emphasis
upon the possibility of changing law in terms of rational considerations of social utility (rather than freezing it in terms of Stage 4 "law and order"). Outside the legal realm, free agreement and contract is the binding element of obligation. This is the "official" morality of the American government and constitution.

Stage 6: The universal-ethical principle orientation.

Right is defined by the decision of conscience in accord with self-chosen ethical principles appealing to logical comprehensiveness, universality, and consistency. These principles are abstract and ethical (the Golden Rule, the categorical imperative); they are not concrete moral rules like the Ten Commandments. At least, these are universal principles of justice, of the reciprocity and equality of human rights, and of respect for the dignity of human beings as individual persons.

These stages are not defined by "particular opinions or judgments, but by ways of thinking about moral matters and bases for choice." Kohlberg (1980) describes Stages 1 and 2 as preconventional, "since decisions are made largely on the basis of self-interest and material considerations" (p. 60). Stages 3 and 4 are the group-orientated stages and are the stages at which most of the adult population operates; and stages 5 and 6 are characteristic of about ten to twenty percent of the adult population, with stage 6 only having about five percent.

Kohlberg describes these stages as universal, able to be found in various cultures; tests have been completed and Kohlberg states that children in the United States, Turkey, and Taiwan for example, all have the same sequence of development.
In theory, two subjects might reach different conclusions about a certain particular moral dilemma, and can yet be located at the same stage of moral development by virtue of the "parallel modes" of moral reasoning employed, for it is the form of moral reasoning employed, for it is the form of moral reasoning, not the content of the arguments used, that Kohlberg uses as his data. This differentiation of moral reasoning from content is crucial in his theory, for it assures that the psychologist will be nonideological in status. Thus, as stated, Kohlberg finds the sequence to be universal. This idea of a universal sequence also implies that movement from one stage to another follows sequentially, and it is not possible for someone to skip a stage on the way up. Human beings, at whatever stage they are at, only are capable of comprehending the moral reasoning of others at the immediate stage above them. Except for some stage 4 subjects who temporarily have to go back to stage 2 reasoning on their way to stage 5 thinking, moral development is irreversible, that is, one can also not go back. One can speed up or slow down development, but he cannot change its order or sequence.

**Rawls's Moral Psychology**

Rawls gives an account of the development of the sense of justice in terms of a three-stage process with a
psychological principle corresponding to each of the three stages. The three stages will be stated here, and as can be seen, Rawls draws on Rousseau and Kohlberg for his theory of moral development.

First law: given that family institutions are just, and that the parents love the child and manifestly express their love by caring for his good, then the child, recognizing their evident love of him, comes to love them.

Second law: given that a person's capacity for fellow feeling has been realized by acquiring attachments in accordance with the first law, and given that a social arrangement is just and publicly known by all to be just, then this person develops ties of friendly feeling and trust toward others in the association as they with evident intention comply with their duties and obligations, and live up to the ideals of their station.

Third law: given that a person's capacity for fellow feeling has been realized by his forming attachments in accordance with the first two laws, and given that a society's institutions are just and are publicly known by all to be just, then this person acquires the corresponding sense of justice as he recognizes that he and those for whom he cares are the beneficiaries of these arrangements. (Rawls, 1971, p. 490-491)

Rawls describes these as "laws" or "tendencies." It should first be made clear that these are not "laws" of psychology, which make every aspect of development of moral reasoning explainable or predictable. They represent tendencies which are presumed to be operable, other things being equal. The way they should be looked at is the role they play in the description of the important features in the way human beings normally develop. Bates, in his article "The Motivation to be Just" (1974), points out
another feature of Rawls's theory of development. Bates points out that these psychological principles involve references to the moral concept of justice itself. It is here that Rawls differs greatly from Hobbes and his social contract theory. Throughout political philosophy, it was often assumed that when the question was asked of Why should I be moral? or Why should I act justly? it would be interpreted in terms of nonmoral motivation. This was the major premise in the attempt to identify duty and interest. What Rawls does in adopting the model of Kohlberg is imply that there may be a "natural" interest in morality itself. In using the concept of "natural", one does not mean that the moral motivation will inevitably be present in every individual. Rawls's concept of natural seems to be in line with what Aristotle suggests when he says, "Neither by nature, then, nor contrary to nature do the virtues arise in us; rather we are adopted by nature to receive them, and are made perfect by habit" (Mc Keon, 1947, p. 331). Bates suggests then, that what Rawls means by "natural" interest in morality would be this adoption by nature to receive the moral virtues.

It is with the development of the sense of justice, therefore, that one can explain the motivation for particular acts of justice in hard cases such as Hume provided, as when the interests of the individual and the interests of society conflict with the demands of justice.
For Rawls, people will act justly in these cases because they have a sense of justice.

Rawls and Kohlberg

One of the major criticisms of both the theory of Kohlberg and Rawls is that they have a liberal bias. Many instances can be found in Rawls's theory in which the liberal bias can be found to alter his framework for analysis. Kohlberg's most direct claim about Rawls is essentially that Rawls's theory represents a good example of the Stage 6 form of moral development, or to put it more precisely, Rawls's theory is a "theoretical derivative from and an attempt to justify the 'natural structure' of Stage 6" (Boyd, 1980, p. 185). Quoting Kohlberg (1979): "... the central achievement of Rawls's theory is that it represents the first clear systematic justification of the principles and methods of decision we call 'Stage 6', principles and methods of decision only partly articulated by Kant" (p. 46). Kohlberg goes on to explain that Rawls's notion of the original position combined with the concept of "reflective equilibrium" is a theoretical construct reflecting justice as a feeling of reciprocity which is present in all the six stages of development, but fully "equilibrated" at stage 6.

The most fundamental link between the two theories is the characterization of persons as self-determining.
and rule-following agents. As Rawls puts it: "moral persons are defined as persons that have a conception of the good and a capacity for a sense of justice" (Rawls, 1971, p. 10). Characterizing human beings as individual, self-determining and rule-following agents concerned with mutual respect is an assumption that differentiates Rawls and Kohlberg from many other philosophical and psychological theories. Rawls makes the assumptions of this kind in the original position as a way of elaborating and legitimizing our intuitions about the primacy of justice; and in so doing he is contrasting his theory with that of utilitarianism, which extends principles of choice appropriate for one individual's conception of the good to questions about the regulation of many persons. (Rawls, 1971, p. 28-29) As a result, we must remember that the original position is a "procedural interpretation of how persons of a certain sort seek a provisional, dynamic state of reflective equilibrium with each other about questions concerning their interactions in a world of a certain sort" (Boyd, 1980, p. 203).
Critique of Rawls's Moral Psychology

Problems with Kohlbergian Developmental Approach

Cohen in his article, "Stages and Stability: The Moral Development Approach to Political Order" (1980), suggests that a serious conceptual difficulty rests in Kohlberg's prescriptive use of his stages of development. The question that he asks is, "In what sense can one regard moral reasoning at one stage — Stage 6, for example — as being morally more adequate or superior to that of lower stages" (p. 73)? What we must look at, is how Kohlberg defines adequacy; he uses a couple of arguments.

First, Kohlberg (1973) argues that philosophers from Kant to Hare have found Stage 5 and Stage 6 types of thinking constitutive of "moral" thought. Secondly, he argues that individuals prefer the more advanced stages of moral thinking. Neither of these assertions is convincing, as Cohen correctly points out. Producing empirical evidence that shows that people prefer the higher stages does not prove that the higher stages are morally superior. In order to preserve its prescriptive use, what Kohlberg must show is that people ought to prefer to achieve Stage 6 of development; he must show why these stages are desirable. The result is that Kohlberg's psychological theory lacks an ethical theory of what is good and right, and a prescriptive theory such as he provides is in need
of this. As Cohen (1980) puts it, "to say that people are as they are — assuming they are at Stage 6 — is not to state why one should esteem them as they are" (p. 73).

For his ethical theory, Kohlberg turns to the coincidence of psychological and philosophical views of the nature of moral judgments. It is to Kohlberg's advantage that philosophers such as Kant and Hare come close to agreeing with the content of his stages, but it does not go very far in validating his claims. Even Kohlberg (1971) sees the need to supplement their views of moral reasoning in order to develop the content of his stages. This is not surprising, for without developing his own ethical theory, Kohlberg cannot show that the liberal democratic principles that formulate his Stage 6 are superior to those developed by Kant and Hare or anybody else. The result is that Kohlberg relies on his empirical evidence, and is satisfied that his views happen to coincide with those traditional philosophers. In fact, Kohlberg's argument for Stages 5 and 6 rests on the moral reasoning of Stages 3 and 4, with the moral reasoning of Stage 4 representing the closest connection to authority such as philosophers like Kant and Hare represent, while Stage 3 represents the characteristic that people prefer the higher stages, the feeling of adhering to the expectations and actions of others. The result is that Kohlberg's defense of the higher stages, 5 and 6, is left without a principled
argument, which in turn leads to the connection with Rawls.

According to Kohlberg, these stages are natural and sequential, and they are not reflective of particular moral teachings or ideology; and as stated, an equilibrium is achieved when moral dilemmas are not solvable with reference to the reasoning appropriate at that particular stage. At Stage 6, rights and duties are recognized as being correlative, and a stable moral structure is characterized as follows:

Universalizability and consistency are fully attained by the reversibility of prescriptions of actions. Reversibility of moral judgment is what is ultimately meant by the criterion of the fairness of a moral decision. Procedurally, fairness as impartiality means reversibility in the sense of a decision on which all interested parties could agree insofar as they can consider their own claim impartially, as the just decider would. If we have a reversible solution, we have one that could be reached as right starting from anyone's perspective in the situation, given each person's intent to put himself in the shoes of the other. (Kohlberg, 1973, p. 641)

It can be argued that the concepts of universalizability and reversibility do not demonstrate the moral adequacy of Kohlberg's stages either. The question of choice enters into this equation of considering whether the notions of universalizability and reversibility are useful in identifying a morally adequate solution. Cohen (1980) offers the following dilemma:

Imagine a hospital ward occupied by ten patients, each of whom is suffering from terminal kidney disease. The hospital possesses only one life-saving kidney machine, and turns to the patients to decide among themselves on a decision-making criterion. (p. 76)
In applying this dilemma to Stage 6, any course of action that all the victims would decide to take must meet the requirements of the principles of universalizability and prescriptivity. If one figures the sense of self-preservation into the equation, the requirement of universalizability and reversibility "merely contribute a vehicle for advancing a minimal risk solution." If one assumes this feeling of self-preservation and longevity, then, the only solution that is likely to meet with everyone's agreement is one that divides the risks up equally, like a lottery. Even if one applies this dilemma to the ethical theory of rights and duties as stated in Stage 6, the moral reasoning that takes place is more like Stage 2 reasoning; the stage where a right action consists of that which "instrumentally satisfies one's own needs and occasionally the needs of others." The result is that the Kohlbergian requirements of universalizability and reversibility in Stage 6 by no means insure that the resolution reached will be more morally adequate than that reached at any other stage.

Thus the connection with Rawls. When Kohlberg (1971) states that a "just solution to a moral dilemma is a solution acceptable to all parties, considering each as free and equal, and assuming some of them knew which role they would occupy in the situation" (p. 213), he is stating something that is in total agreement with Rawls. Rawls's
maximin principle is therefore a Stage 2-like argument. On the other hand, one could take Rawls's belief that individuals choose the principles of justice because they promote "our good and that of those with whom we are affiliated" as a Stage 5-like argument, the stage that undertakes a more rule-utilitarian orientation. The result is that when one applies Rawls's principles of justice to the principle of unanimous consent, it in no way guarantees that the solution will be more morally adequate than any of the lower stages of reasoning. Attempting to solve the dilemma of moral distribution of scarce human resources through the requirements of universalizibility may not generate a universal agreement. Rawls's formulation of distributive principles, even with prudential and utilitarian characteristics representative of Stage 5, and not Stage 6, has been shown previously to be less than adequate by itself.

Rawls (1971) states that "it may be objected that much social theory does well enough without using any moral ideas" (p. 492), and he is absolutely correct. Quoting Cohen (1980): "insofar as Kohlberg regards questions of distributive justice as moral conflicts that are susceptible to Stage 6 analysis, the difficulty is that those distributive principles that individuals may be willing to subscribe to as universalizibility will run the gamut of political and economic theory" (p. 77). The key to the
entire argument is this: principles that are subject to unanimous consent, rather than reflecting the Stage 6 respect for rights and duties, are most likely a result of a more basic motive, an egoistic motivation; and also, these principles might also differ quite a bit from the liberal democratic ideals that are present in the higher stages.

The Application of Rawls's Moral Psychology to the Principles of Justice

After analyzing Rawls's principles of justice and his moral psychology, we now have the opportunity to view the relationship between the two parts of his theory. It is in the original position that Rawls tells us what justice is, justice being defined by the two principles. Methodologically, the principles have to be checked against our considered moral judgments about what is just, in order to see whether the judgments we make have been derived from the principles. Rawls's moral psychology, if properly constructed, would show us how men acquire a sense of justice. In particular cases, the sense of justice can be used to explain the motivation to act justly. Bates (1974) notes that this argument by Rawls does not provide "an argument addressed to any rational self-interested being, who lacks a sense of justice, which will show him that he 'ought' to be just — in a nonmoral sense of
"ought"" (p. 13). Bates points out that this is exactly the argument that Foot (1967) demanded in "Moral Beliefs," but she reconsidered this demand when she stated:

I was in this difficulty because I had supposed — with my opponents — that the thought of a good action must be related to the choices of each individual in a very special way. It had not occurred to me to question the often repeated dictum that moral judgments give reasons for acting to each and every man. This now seems to me to be a mistake. Quite generally the reason why someone choosing on A may "be expected" to choose good A's rather than bad A's is that our criteria of goodness for any class of things are related to certain interests that someone or other has in those things. When someone shares these interests he will have reason to choose the good A's; otherwise not. Since, in the case of actions, we distinguish good and bad on account of the interest we take in the common good, someone who does not care a damn what happens to anyone but himself may truly say that he has no reason to be just. The rest of us, so long as we continue as we are, will try to impose good conduct upon such a man, saying "you ought to be just," and there is this much truth in the idea that there are categorical imperatives in morals. (Rawls, 1971, p. 9)

As Bates points out, although such an argument as suggested cannot be provided, it still does not hold against the argument for the principles as such, for no such argument is needed.

What questions are left unanswered? We know what justice is, and we know why men will be just. But the major question is still left unanswered. Why ought one to be just? The principles of justice, as presented, are checked by the considered moral judgments; these judgments however, are made by those with a sense of justice, the
sense of justice defined by reference to a set of psychological principles which themselves involve a reference to justice as defined by the principles of justice. What this concoction shows is that the whole system is inter-defined, everything relates to everything else; but if one looks closely, the whole Rawlsian system is never really justified. What would be a proper justification? Bates (1974) suggests that a justification of the whole system consists "of an argument which could be presented to any individual on any occasion, which would not appeal to any moral sentiment, but would show him that from his own point of view justice would be good to him" (p. 13).

**Rawls's Concept of Reciprocity**

Rawls advances a theory of moral development consisting of three stages, and at each of these stages the psychological principle of reciprocity applies:

> The tendencies assert that the active sentiments of love and friendship, and even the sense of justice, arise from the manifest intention of other persons to act for our good. Because we recognize that they wish us well, we care for their well-being in return. Thus we acquire attachments to persons and institutions according to how we perceive our good to be affected by them. The basic idea is one of reciprocity, a tendency to answer in kind. (Rawls; 1971, p. 494)

Pritchard, in his article "Rawls's Moral Psychology" (1977), suggests that "it is important to notice that this tendency to reciprocate is not a merely self-interested response. One does not reciprocate only in order to receive
benefits. One reciprocates because of receiving benefits from others, but the response is one of genuine other-regard" (p. 63). We must remember that this sentiment toward others can only take place by receiving some kind of benefit, which results in a feeling of gratitude; and as Pritchard suggests, this seems to fall short of generating the notion of respect for all. Even if the psychological principle of reciprocity is sufficient, it falls short of what Rawls requires in his conception of justice as fairness:

But for one who understands and accepts the contract doctrine, the sentiment of justice is not a different desire from that to act on principles that rational individuals would consent to in an initial situation which gives everyone equal representation as a moral person. Nor is it different from wanting to act in accordance with principles that express men's nature as free and equal rational beings. (Rawls, 1971, p. 478)

The problem however, is that the principles adopted in the original position cover only those who have the potential to reciprocate. This works out well with the principle of equal-respect, but it seems to "transcend" the principle of reciprocity, which is a principle that is concerned only with returning in kind; and as Pritchard points out, Rawls at times seems to suggest that the sense of justice extends beyond simple reciprocity:

We desire to act on the natural duty to advance just arrangements. And this arrangement goes beyond the support of those particular schemes that have affirmed our good. It seeks to extend the conception they embody to further situations for the good of the larger community. (Rawls, 1971, p. 474)
If we stay within the confines of the principle of reciprocity, it is not at all clear what would account for people adopting this "wider concern." Pritchard goes on to analyze the difficulties Rawls's psychological account presents for the adoption of the principle for equal respect; but for the purposes here, it will suffice to note that Rawls extends beyond the principle of reciprocity in establishing his sense of justice, a principle of reciprocity that will be used as an example of altruistic motivation in the next chapter of the paper.
Hume's method of political analysis was grounded in the technique known as "analytic dissection." Hume used this technique with devastating effect against the state of nature hypothesis present in some of the social contract theories. He suggested that it was "slippery logic" which sought to justify rebellion against authority on the basis of a "mere philosophical fiction" like the state of nature; and Hume denied that society was a result of a direct agreement aimed at eliminating certain "inconveniences" present at the pre-social condition. This arrangement that was called society, for Hume represented a set of responses to human needs and desires: between altruism, which diminishes in its power as it is extended through circles such as the family, friends, neighbors, and strangers; and selfishness — which increases as societal demands on the self become more abstract and remote. As a result of this, as man found himself in this condition of moderate scarcity, his insecurity deepened, leading him to demand from others, and leading him to invade the material possessions and economic goods of others. In his Treatise, Hume suggested that this "natural" condition could be overcome by artificial arrangements which would restrain what Hume termed man's "heedless and impetuous passion," and it would also serve to gratify man's instinct for self-interest.
It was in this aspect that Hume shifted positions in his theory from the other philosophers. Although Hume was aware of the long, painful process by which society was shaped, and was aware of the existing institutions; he accomplished what Wolin (1954) calls "a minor revolution" in political thought (p. 253). Political analysis to Hume was to be conceived in psychological rather than juridical categories. Writers such as Hobbes and Locke, while they were aware of the importance of the concept of human nature in political analysis, had nevertheless approached it merely as a preliminary discussion to the more important argument for the social contract. Conceived in this way, concepts such as obligation, duty, and justice were conceived of as "logical derivations" from the basic concept of sovereignty, or as Locke called it, the "supreme power," which was established by the social contract. Hume on the other hand turned this analysis around. He suggested that concepts such as obligation, duty, and justice were consequences of human attitudes and expectations. These concepts were to be analyzed on psychological, not juridical grounds.

Where does Rawls fit into this scheme? Rawls's argument justifying moral rights takes the form of a social contract: purely self-interested creatures, meeting behind a veil of ignorance, agreeing to his two principles of justice over the alternatives. One would think that
this sort of argument does not leave much room for a theory of human nature. These creatures of "rational self-interest" seem far removed from a democratic personality. One could also argue that such self-interested creatures would not select principles suitable to men infused with fellow feelings and social interest. As a result, more than one critique has correctly pointed out that Rawls's "egoistic" argument is inconsistent with his theory of man.

Rawls's theory of human nature, however, plays a major role in his argument from the original position. One of the primary aims of the parties in the original position is to arrive at principles that promote healthy development. Rawls does not speak directly in these terms, but instead talks of the desire of the parties to secure self-respect. Quoting Rawls (1971), they "would wish to avoid at almost any costs the social conditions that would undermine self-respect" (p. 440). In his theory, self-respect is associated with healthy moral development, while a loss of self-respect dictates some form of pathological growth.

Along with ascribing to the parties the goal of securing self-respect, Rawls also gives them a sufficient knowledge of human nature to pursue their goal. Rawls provides them with what we have determined as the Aristotelian Principle and the facts of moral psychology. The
contractors in the original position do not know much, but they do know that their self-respect will suffer if they do not develop a plan that will suffice for their development. This development, according to Rawls, consists of the individuals wanting to develop liberty in order to secure their individuality. Rawls (1971) also states that parties in the original position know that their self-respect depends upon their plans being encouraged by others (p. 178-179). Thus the key point: by their plans having to be encouraged by others, man in the original position realizes that he must avoid a primatized social order and instead must seek a society that encourages social interest.

The other instance in which human nature plays a role in Rawls's argument is through the parties' evaluation of the stability of rival conceptions. For Rawls (1971), to check a conception of stability means to determine whether "in view of the laws of moral psychology men will acquire a desire to act on it" (p. 138). Parties in the original position, then, are aware of their capacities for fellow feeling and a sense of justice and evaluate the options of the various proposals with what they know of the nature and limits of these capacities. Having already shown the limitations of Rawls's strategy, let us substitute for Rawls's moral psychology and sense of justice a position closer to the concept of human nature in Hume. Let us assume that man is basically egoistic, capable of
only certain feelings of altruism, devoid of Rawls's concept of fellow feeling and his sense of justice. Given this conception of human nature, man would decide that a stable political order is possible only under a "sovereign" powerful enough to make "free-riders" do their part.

The basic issue can be posed as follows. Let us assume that in acting in society, an individual can be conceived of as affirming three distinct aims: (a) achieving social justice; (b) meeting moral obligations other than social justice; and (c) attaining personal ends of a nonethical character.

We may assume that a person of a "normal" human nature, a nature consisting of a mixture of egoistic and altruistic motivation, would "trade-off" among these three concerns in choosing a course of action. Rawls's argument, however, requires the individual to give strict priority to the goal of social justice over all other concerns, when it comes to choosing social policies.

Rawls's principle of equal liberty requires a form of democracy in which the decisions made in some sense represent the choice of citizens. Therefore, in a just economy, citizens, as voters and legislators, support policies that conform to the Difference Principle, with full knowledge of the effect that these policies will have on their social position.
The result is this; citizens are required to legislate social justice even when it conflicts with other moral obligations or personal concerns. These conflicts are what Rawls refers to as the "strains of commitment," and the frequency and severity of the conflicts depends on the distributive principle that happens to be taking place in the economy in question. In a capitalistic society, even with the egalitarian leanings of Rawls's theory, the strains of commitment would be quite severe, severe enough to a degree of instability.

Methodologically, there seems to be no system that could satisfy the Difference Principle, however successful it would be in reducing the strains of commitment. It appears that there is no acceptable theory of moral behavior that could claim the lexical priority of social justice over other moral obligations and personal concerns.

To summarize, Rawls's moral psychology enters into the choice situation in the original position in two areas. First, the contractors use what they know about the Aristotelian Principle and their social nature to help secure their self-respect; and secondly, moral psychology enters into the choice situation in comparing the alternative conceptions of justice. We have seen how Rawls's moral psychology, based on Kohlberg's developmental approach, has led to results that are less than satisfactory;
where man's social nature has led to a sense of justice that oversteps the boundaries of normal altruistic behavior; and where man's social nature leads to a choice of the Difference Principle, a principle flawed in its very construction.
CHAPTER V

THE CONCEPT OF STABILITY

When Rawls speaks of the concept of stability, he speaks of it only within the context of its psychological orientation. When Rawls discusses the acquisition of the sense of justice, he discusses it in terms of the members of a well-ordered society. In this section however, the concept of stability will not only be looked at in its psychological aspect, but also in its relation to political obligation, and also in the context of the economic implications to which Rawls's theory would lead.

The hypothesis states that in conditions of just or nearly just institutions, Rawls's sense of justice would lead to a level of instability that would be unacceptable. Therefore, before analyzing Rawls's concept of relative stability, we should first look at the role that institutions play in this analysis.

The general benefits that institutions provide are often what in economics is called public or collective goods, a notion defined by Olson (1965) in this way:

A common, collective, or public good is here defined as any good such that, if any person x, in a group x₁...xₙ consumes it, it cannot feasibly be withheld from the others in the group. (p. 14)
This relates to our problem in that, in large cooperative schemes, it is often possible for people to enjoy the benefits of consuming collective goods without sharing in any of the work in the maintaining of this particular institution. Hume talks about this in his Treatise when he gives the example of two persons agreeing to drain a meadow which they both possess; one thousand people agree to participate in the draining. Human nature states that "each seeks a pretext to free himself of the trouble and expense and would lay the whole burden on others" (Hume, 1964, p. 538). Basically, Hume's argument is that in such large cooperative schemes, people tend not to want to bear any of the work, since they will be able to receive the benefits anyway; Hume asserts, as does this thesis, that this leads to a basic instability in an institution of this sort. This instability must be taken care of by a strong centralized government, which uses some form of coercion in order to make people share in the burdens. This is the same view taken by a number of political philosophers and economists, most notably Baumol (1965), an economist who has extended this economic concept of public goods to political institutions by stating that government intervention with the threat of sanctions is justified in order to bring some form of stabilization to these cooperative institutions; institutions which insure such things
as education of the population, investment in the future, free trade and so on. Olson (1965) also deals with this subject; his book applies this conception to the theory of groups, arguing for example that labor unions should be allowed to make membership compulsory, in that in this way unions can maintain a stable form of association, since those who benefit from the policies must bear the burdens of supporting it.

This type of institutional instability to which the phenomenon of collective goods leads was first noted by Hobbes, whose work has many analogies with the approach of economists like Baumol and Olson, the basic one being that his theory works on the basis of egoistic assumptions alone. Hobbes's conception of the state of nature attempts to clarify a situation which would arise where men, having only their own desires in mind, try to enter this cooperative scheme in order to receive the benefits without accepting the burden. Richards (1971) proposes two sorts of instability in these cooperative schemes. First, "if they are sufficiently large, they will produce benefits to all persons irrespective of whether or not they have contributed to the support of the scheme" (p.149). As a result the Hobbesian rational egoist will not contribute to the scheme, choosing instead to reap the benefits without bearing the burden. Baumol proposes the second sort; "if each beneficiary of the scheme knows that the other
beneficiaries are strongly tempted not to contribute, or only pretend to; then he will wish to be first, or among the first, who does not contribute since he will gain more, and lose less, by being among the first" (Richards, 1971, p. 149). What Hobbes is stating is that cooperative schemes will always break down, since people will have no security concerning the obedience of others, if they obey or disobey. Hobbes's remedy for such a situation is the creation of a person with absolute coercive power; one whose power is publicly known and acknowledged, one who commands obedience. In this way, the stability of cooperative institutions is assured, people will know that others will obey, if they obey; and there is an overriding prudential reason, since the certain threat of the sovereigns' coercion will outweigh any probable benefits from disobedience. In this way, both sources of institutional instability are removed.

The relevance of these remarks is this: given that institutions often provide public goods and are subject to both the instabilities of the Hobbesian situation, and given the assumption of strong egoistic motivation, one of the main concerns of the contractors in the position of the social choice function will be to agree on an ultimate standard which, when it is accepted and acted on, will best stabilize such institutions. As is evident from the discussion in Chapter III of this paper, Harsanyi's
principle of average utility will fulfill this criteria; and Rawls's maximin principle, combined with his sense of justice, will lead to the sorts of instability that have been spelled out by Hobbes. What Chapter III has shown is that the maximin principle has wholly unacceptable moral implications, and that one does not obtain a satisfactory concept of justice if one imagines that the criterion of justice is chosen by people in the original position in accordance with the maximin principle. The difference principle itself would require us to give absolute priority to the least well-off individual, even under the most extreme conditions. Even if the individual's interest were affected only in a very minor way, and all the other individuals in society had opposite interests of the greatest importance, the single individual's interest would always override anyone else's.

As a result, it is better to analyze the concept of stability not only in the context of its psychological implications, but also in regards to the concept of political obligation, and the economic implications that would result from Rawls's theory.
Stability and the Sense of Justice

Rawls treats stability in two stages. The first stage is in the discussion of the acquisition of the sense of justice by the members of a well-ordered society; the second is when he examines the question of congruence, that is, "whether the sense of justice coheres with the conception of our good so that both work together to uphold a just scheme" (Rawls, 1971, p. 453)

First we must look at Rawls's conception of a well-ordered society, a society Rawls states "designed to advance the good of its members and effectively regulated by a public conception of justice.... A society in which everyone accepts and knows that the others accept the same principles, and the basic social institutions satisfy and are known to satisfy these principles.... (and where) Justice as Fairness is framed to accord with this idea of society" (Rawls, 1971, p. 453-454).

For Rawls, a stable society is conceived of in this way:

a well-ordered society endures over time, its conception of justice is presumably stable; that is, when institutions are just (as defined by this conception), those taking part in these arrangements acquire the corresponding sense of justice and desire to do their part in maintaining them. One conception of justice is more stable than another if the sense of justice that it tends to generate is stronger and more likely to override disruptive
inclinations and if the institutions it allows foster weaker impulses and temptations to act unjustly. The stability of a conception depends upon a balance of motives: the sense of justice that it cultivates and the aims that it encourages must normally win out against propensities toward injustice. To estimate the stability of a conception of justice (and the well-ordered society that it defines), one must examine the relative strength of these opposing tendencies. (Rawls, 1971, p. 454-455)

The assumption is that stability is a desirable feature of a conception of justice, and that other things being equal, as Rawls (1971) suggests, "the persons in the original position will adopt the more stable scheme of principles" (p. 455). As a result, this leads us, as it does Rawls, to a comparison of the two major conceptions of justice offered. The argument that Rawls makes for his conception of justice as fairness has already been analyzed, and the sum-of-utilities position of Arrow and Harsanyi has also been discussed.

Rawls's argument is that at least to some degree, human nature is such that persons acquire a desire to act justly when these persons have lived under and benefited from just institutions. For Rawls (1971), "a conception of justice is psychologically suited to human inclinations ... (and) acting justly is part of our good" (p. 455). Therefore Rawls (1971) states that there is congruence between the conception of justice and the conception of the good, and as a result, his conception is much "more in line with the principles of moral psychology" (p. 455).
The question of stability arises when a supposedly just scheme of cooperation is not in equilibrium. From the standpoint of the original position, Rawls's principles of justice are collectively rational; and "everyone may expect to improve his situation if all comply with these principles, at least in comparison with what his prospects would be in the absence of any agreement" (Rawls, 1971, p. 497). Rawls admits that in this situation, from the perspective of the individual, "both first-person and free-rider egoism would be still better." The problem arises in his specifications in the original position; it is here that Rawls does not allow for these alternatives. Rawls admits that if an individual is so inclined, he "can sometimes win even greater benefits for himself by taking advantage of the cooperative efforts of others. Sufficiently many persons may be doing their share so that when special circumstances allow him not to contribute (perhaps his omission will not be found out), he gets the best of both worlds: on these occasions anyway things proceed much as if free-rider egoism had been acknowledged" (Rawls, 1971, p. 497).

It is at this point that the problem of instability arises. The position of this paper and the criterion set by Rawls does not differ; the difference lies in the fact that to insure stability, Rawls states that men must have a specific sense of justice, or a concern for those who
would be left out in the breakdown of this cooperative scheme, with Rawls preferring both. When man has this sufficiently strong sense of justice, his temptation to violate the rules is overruled, and as a result just schemes are stable. It is man's rational plan of life regulated by this sense of justice that leads to Rawls's conclusion. This is the alternative to the positions set forth by Hobbes and Hume as stated in the introduction.

The argument is this: Rawls's alternative does not fit the criterion; man is inherently egoistic in his motivation, and is not subject to the development of a sense of justice that would lead to a basic instability itself, something not foreseen by Rawls. This misconception of man's human nature has led Rawls to develop his theory of moral development and his sense of justice that does not remove the two types of instability as proposed by Hobbes. Rawls fails to recognize the extent to which the burden of proof rests with him in this matter of social stability.

Rawls's argument proceeds this way: the more the three elements or psychological laws are realized, the stronger the sense of justice.

The first enlivens the sense of our worth strengthening the tendency to answer in kind, the second presents the moral conception so that it can be readily understood, and the third displays the adherence to it as attractive. The most stable conception of justice, therefore, is presumably one that is perspicuous to our reason, congruent with our good, and rooted not in abnegation but in affirmation of the self. (Rawls, 1971, p. 499)
The argument follows that Rawls's sense of justice is too strong, not one that would be rationally contracted in the original position. In comparing the contract view with the utilitarian position, there is no question that the unconditional concern of other persons and institutions for our good is far stronger in the contract view. It is not the position of this paper to dispute this claim, in fact it is the exact opposite; Rawls's argument for a strong sense of justice, with the emphasis on the reciprocity principle, correctly asserts a strong feeling for the fellow man, a strong feeling for justice, and this is the problem itself.

There are several areas in which the sense of justice is stronger than the sentiment offered by the utilitarian conception. First, as suggested by Rawls, the "unconditional concern" for other people and institutions for the good is stronger in the contract view. The criterion Rawls sets up insures that the principles of justice guarantee everyone an equal liberty, and also assures that individual claims will not be neglected for the sake of the larger sum of benefits, or for the whole society. The priority rules and the difference principle, buttressed by its Kantian interpretation and the reciprocity principle, insure this. This caring for others and the strengthening of our self-esteem is much more intense in his contract interpretation than it is in the utility principle.
For Rawls, mutual ties exist that bind an entire society, ties that do not exist in the utilitarian framework. Utilitarianism stresses the capacity for sympathy, an alternative altruistic assumption that is not as strong as when one adopts Rawls's two principles of justice. And as projected by Mall in the discussion of Hume, utilitarianism also stresses a strong central authority. Rawls asserts that the social system regulated by justice as fairness would lead to the identification with the good of others, and an appreciation of what they do would be an element in our good. This feeling is possible because of this sense of mutuality that is implicit in the principles of justice themselves. Quoting Rawls (1971): "With the constant assurance expressed by these principles, persons will develop a secure sense of their own worth that forms the basis for the love of humankind" (p. 501). The principle of utility, on the other hand, appeals to the capacity for sympathy as the foundation of just conduct in the absence of reciprocity.

Rawls argues for the greater stability of his principles of justice by considering only those psychological laws which he thinks are true. In this paper however, the concept of stability is considered important enough to look at in greater detail.
Rawls's Conception of Political Obligation and Civil Disobedience

In looking at the question of instability as applied to institutions, one must consider Rawls's concept of "civil disobedience" which he sets forth in his chapter on "Duty and Obligation." First, it should be said that Rawls does not attempt to justify his conception by an analysis of the choice function in the original position. Instead, he simply asserts that "when they adopt the majority principle the parties agree to put up with unjust laws only on certain conditions," and that it "could not have been within the meaning of the duty of justice in the original position" that the parties be "required to acquiesce in the denial of their own and others' basic liberties" (Rawls, 1971, p. 355). Rawls does not offer a clear definition of what the "basic liberties" are, which limitations on them constitute a "denial" of liberty as opposed to a necessary restraint, or how the different liberties are to be balanced against one another. Perhaps more importantly, he does not demonstrate that, assuming his view of justice could be made consistent, anything approximating it could be or is politically actualizable. Nonetheless, Rawls (1971) states that "we submit our conduct to democratic authority only to the extent necessary to share equitably in the inevitable imperfections of a constitutional system" and that "in the long run the
burden of injustice should be more or less evenly distributed over different groups in society" (p. 355).

The result of this abstract view of political obligation is that it is very subjectivistic. It would seem that if a view of this sort were accepted, the stability of any regime would be in question; and therefore, if one were to apply Rawls's view to the original position, it would seem that it would not likely be chosen by any contractors who were in any sense rational or aware of political realities. Rawls (1971) states "that there is a limit on the extent to which civil disobedience can be engaged in without leading to a breakdown in respect for law" (p. 374); or at least provides no operationally adequate restraints on exceeding the constitutionally responsible limits on civil disobedience. One could read into his writing that one is encouraged to disobey the law whenever he feels that a "basic liberty" has been violated (p. 372), or when his government's aims in a war are "sufficiently dubious" (p. 381), or even when "the public decision concerning the level of savings' is wrong. (p. 296)

Although Rawls's man possesses a sense of justice, he seems not to possess a certain loyalty to a government, that would lead him to make sacrifices for his country; Rawls instead suggests that, "Given the often predatory aims of state power, and the tendency of men to defer
to their government's decision to wage war . . . a general willingness to resist the state's claim is all the more necessary" (Rawls, 1971, p. 382).

There are two key questions that should be asked, questions to which Rawls does not provide any substantive answers. Why is it that an individual would be more likely to be right on an issue than a properly constitutional government? And by what standard should the individual citizen go to determine whether the government's policies are "sufficiently" unjust to render civil disobedience. The only answer that Rawls (1971) provides is his statement that "it is often clear that these freedoms are not being honored" (p. 372). Even when this is true, it is not the answer to the question of when disobedience is legitimate; the question is whether the deviation from Rawls's principles of justice is or is not necessary. Rawls's man is left to his "intuition" to make this decision.

Quoting Rawls (1971): "A law or policy is sufficiently just, or at least not unjust, if when we try to imagine how the ideal procedure would work out, we conclude that most persons taking part in the procedure and carrying out its stipulations would favor that law or policy" (p. 357). Also, Rawls denies that even "if many rational persons were to try to simulate the conditions of the ideal procedure and conducted their reasoning and discussion accordingly, a
large majority anyway would be almost certainly right" (p. 358). The result is that different individuals who try to "imagine" this outcome of an "ideal procedure" would most likely draw different conclusions about it. Therefore, it seems that Rawls himself is unable to provide some substantive criterion for judging what is a "just" policy in given circumstances.

The problem with Rawls's concept of civil disobedience is the same problem that underlies the rest of his theory. It is based upon a political psychology which is not based on any serious observation of human nature. For Rawls, (1971) civil disobedience is simply "one of the stabilizing devices of a constitutional system, although by definition an illegal one" (p. 383). For those who "may object to this theory of civil disobedience that it is unrealistic," Rawls states that he has "assumed throughout that we have to do with a nearly just society" (p. 386). Rawls does not consider at all what the effect of widespread disobedience to law would do for the stability of a society. In fact, quoting Rawls '(1971): "The theory of justice has nothing specific to say about these specific practical considerations" (p. 376). At times Rawls mounts a virtual disdain for authority and the problem of order: "if justified civil disobedience seems to threaten civic concord, the responsibility falls not upon those who protest but upon
those whose abuse of authority and power justifies such opposition" (Rawls, 1971, p. 390-391).

Unlike most of the political philosophers mentioned in Chapter I, Rawls has totally ignored the psychological factors that are necessary to the support of any government, and he has also underestimated the difficulty of maintaining the stability of one that is reasonably just. All Rawls does, is assume that the main reason why people will support a government is that they correspond to his model of justice. He does not show that the sense of justice that he describes actually exists in men. If he did, there would be no need for government at all. The main point is this: Rawls does not take into account men's tendencies to be biased judges in their own cases, and to want more for themselves than what is their due; nor does he show that men possess this innate sense of equality, as opposed to the desire for men to acquire as much as possible at the expense of others. And most importantly, as far as civil disobedience is concerned, Rawls does not consider how this encouragement of disobedience to law may undermine that "reverence" for law that was present in the writings of Aristotle and The Federalist Papers (Madison, 1941), in their argument for the stability of any government.

The main defect of Rawls's theory of political obligation, as it is in other parts of his theory of justice,
lies in the fact that he fails to derive his principles from a realistic examination of political life. Had Rawls considered, as Hobbes did, the alternative to the maintenance of civil society, he would have realized that the citizens of a constitutional government have much for which they can be obligated. The alternative to the maintenance of a stable society is not something that resembles Rawls's original position, where parties are mutually disinterested, but something more like the state of nature in Hobbes or Hume, a condition that would discourage citizens from disobeying laws with which they disagree. Hobbes's harsh reading of man's potential for evil may be exaggerated, but it cannot be ignored.

Stability: The Economic Implications of Rawls's Theory

It is obvious that there must be to some degree a feeling of other-regardingness in the function of a political system. Rawls uses this fact to argue that the use of the market system analysis when applied to the political process is imperfect. As a result, Rawls assumes two types of behavior: in the market, he agrees that selfish behavior is socially correct; but he holds that in the political process, this emphasis on the self would never lead to perfect justice. There is something to Rawls's position; however, it should be argued that political competition does serve some of the same functions as
competition does in the market process. The position taken by many political philosophers is that the expression of voting of one's own interests seems to be an essential part of the information process. If voters do not express their own interests, then how is anyone to know that justice is being carried out? Arrow (1973) states it this way:

I would hold that the notion of voting according to one's own beliefs and then submitting to the will of the majority represents a recognition of the essential autonomy and freedom of others. It recognizes that justice is a pooling of irreducibly different individuals, not the carrying out of policies already known in advance. (p. 258)

Application of Rawls's "difference principle" to our present society would involve a redistribution of wealth and income, or at least sufficient to achieve and to maintain fair equality of opportunity and to guarantee men's self-respect. Rawls does not say how large these transfers of wealth and income would be. The sum-of-utilities rule could also lead towards equalization of income, if, as Arrow (1973) points out, "it is assumed that all individuals have the same utility function which displays decreasing marginal satisfaction from additional increments of income" (p. 258). Sen suggests that how egalitarian Rawls's principle would prove to be depends on how its distribution affects total income. Rawls holds that the close-knittedness of people in society means that perfect equality of income is not to the advantage of the least well-off, but that they will benefit by an increase in
income to those higher up in the income scale. As a result, in Rawls's theory, considerations of incentive and productivity are presumably brought into "equilibrium" with the rights and needs of the least advantaged.

Chapman suggests that the difference principle is not only a way of acknowledging moral equality, but it is also a way to deprive men of "threat advantages" against one another. The difference principle "provides each individual with a maximin fall-back position, beyond which no enemy can drive him."

Chapman (1975) goes on to suggest that "the rational and the attitudinal components of Rawls's theory are even more closely and functionally related than he perhaps realizes" (p. 390). Rawls (1971) does not say that "... embedded in the principles of justice is an ideal of the person that provides an Archimedean point for judging the basic structure of society" (p. 584). The question that Barry asks is how does one logically get to an "ideal of the person" from decisions based on rational calculations about wants and goods? Chapman (1975) suggests that the path is psychological.

May it not be that Rawls's distinctive interpretation of the implication of collective rationality depends upon his Rousseauen reading of human nature and dynamics? That is to say, in a society of equals, free of domination and its corrupting consequences, we would and do undergo a change of moral purpose. Rawls's ideal person, like Rousseau's, has a sense of, and a will for, justice. His moral attitudes arise naturally in a rationally organized society.
This society has its origin in rational calculation; its justification lies in its capacity morally to transform its members. In this sense, the members of the good society embody the moral consequences of their collective rationality. (p. 390)

Rawls's misconception therefore, can be explained in part by his adoption of some of the premises. Unlike Rousseau, however, Rawls states that with the adoption of his principles of justice, men will acquire a sense of justice that is enduring. In order for this to endure, Rawls's man must not only be brought up as was Rousseau's Emile, but he also must sustain a high level of political activity. If men are to regard their laws as just, they have to take part directly in forming these laws, otherwise their level of political consciousness will decrease. As has been shown in this paper, Rawls (1971) shares with Rousseau a view that human development includes "a natural appreciation of demonstrated good will, which arouses the appropriate response of gratitude" (p. 499). But Rawls goes on to postulate the "Aristotelian Principle" of motivation on top of the Rousseau-Kohlberg conception of human development. For Rawls (1971), the principle "accounts for many of our major desires, and explains why we prefer to do some things and not others by constantly exerting an influence over the flow of our activity. Moreover, it expresses a psychological law governing changes in the pattern of our desires" (p. 427). This principle is invented to show
that men want to use the capacities required and encour-
egaged by Rawls's society, and that therefore we should
rationally choose that society and it's form of justice.
The Aristotelian Principle enables us to finally reach
Rawls's promised society, and it is basically the elabo-
ration of the sense of justice. In Rawls's society, man
is a social being. He should try not to be self-suffi-
cient, but accept his own weaknesses, and recognize that
everyone has to play a role and make an equal contribution
to society. For Rawls, the man who is not socially and
politically active is mentally deficient, and leads a
deficient life. For Aristotle, the man who does not be-
long to civil society is considered either a beast or a
god. Rousseau, however, considered the solitary man the
only good man; for Rawls, he is the only bad man. Further
explanation and elaboration on the meaning of the Aris-
totelian Principle is never really given, and conclusive
evidence of its existence is never offered; however, it
is safe to say that in this way Rawls actually transcends
Rousseau's vision of the general will.
An Alternative Model of Psychological Assumptions

Rawls is unusual in maintaining opposing views of the psychological assumptions for the analysis of politics and economics. On the whole, most of those who deal with both subjects tend to form one position: those who think that the economic system must be designed on the basis of self-interest also think that the political system must be designed on the same basis; while those who view politics on the basis of a cooperative scheme tend also to analyze economic institutions on the basis that men work for the common good. In reality, economics and politics are only separated by analysis, and the argument for a single psychology leads both to a more realistic view of man's human nature and a more practical way to analyze. The change here, however, is that although this paper argues for a single psychology, it does not argue for man being totally self-interested, or totally unselfish; instead, an alternative model will be given in which man will be assumed to be a self-interested creature with circumstances of altruism that also enter into the motivational assumptions. This model therefore, will be based upon the theory of human nature that was provided by Hume, and discussed in Chapter I.

Hume's theory consists of man being motivated by self-interest combined with a feeling of sympathy, and
it is this basically utilitarian conception that will be used as the basis for the rest of the model. The concept of self-interest is a tricky one, and will be used in the following way.

An egoist might be able to argue that a person's only real interest lies in simply advancing his own happiness. Everything else is a means to this interest. The question that arises is this: is self-interest the same as being interested in oneself?

We will start by defining "self-interest" as "looking out for one's own happiness." Now, as Barnhart (1976) points out in his article "Egoism and Altruism," it is easy to see that an individual's personal happiness may not entirely be tied up with finding out the details of his own life that make him happy. In other words, he may have to be concerned with the details of somebody else's life in order to find personal happiness. Just being concerned with yourself can lead to a very boring life.

If this is true, says Barnhart, then it follows that self-interest does not necessarily entail egotism. An egotist is one who would never tire of focusing on the details of his own life. Therefore, it is important to distinguish egoism and egotism. The egotist is someone who is so conceited as to think that the lives of others are totally inconsequential to his own life. An egoist,
on the other hand, is someone whose foremost concern is his own personal happiness. Unlike the egotist, the egoist may find that his life is in no way more interesting than someone else's, and that in order to find personal happiness, he may have to give a great deal of attention to the details in the lives of others. Unlike the egotist, the egoist may be uncomfortable talking about himself and may or may not spend time and energy thinking that others find his interesting.

As stated, the egoist is concerned about his own happiness above all else. But, this does not mean that he is concerned to the point of excluding everything else. The egoist is also capable of being unselfish, or to a certain extent being altruistic. A person is unselfish if he can contribute to the happiness and freedom of another person without having any self-interest beyond simply feeling pleasure in this other person's happiness. The argument here is that people do not help another person without expecting to get something in return. This argument however, states that a person can still combine unselfishness with self-interest, or to put it as Barnhart (1976) does: "his self-interest or self-love has expanded to the point that it can include as a part of itself an objective interest in the happiness and freedom of another person.... An objective interest in the happiness and
freedom of another person cannot be accommodated by narrow egoism. Self-interest is broader than egoism" (p. 105).

The model is based on self-interest, and under this concept we will have the concepts of limited egoism and limited altruism. Limited egoism will be defined as a man who is concerned about his own happiness above all else, but not to the point of excluding all else. The limited egoist is capable of unselfishness, a willingness to contribute to the happiness of others if he can receive something in return. It is this aspect of reciprocity that makes up one component of limited altruism, a concept to be detailed below. The result is that we have man's human nature based upon self-interest, with the two main components being limited egoism and limited altruism. This position of both economic and political man coincides with Hume as he states that man is made up of both egoistic and altruistic motivations.

This model was derived from the writings of Nagel (1975) and Arrow (1972), and basically is a combination of their efforts. Since the concepts of self-interest and limited egoism have already been defined and are readily understandable due to their position in political and economic thought, it is only the concept of limited altruism which will be detailed.
The Three Cases of Limited Altruism:

(1) The contribution of support to an institution or practice from which the contributor benefits, even though his benefit is not contingent on his contribution.

(2) The attempt to pursue and avoid certain causal relations between one's own welfare and that of others.

(3) The inclusion of altruistic motives within the scope of service offered for sale.

1. When a person does such things as donating money to his old college, or giving blood for example, he may explain this type of behavior by saying that he has benefited or may benefit in the future from similar behavior by others. This has the look of a straight trade, but Nagel points out that it is not: "he benefits from like actions by others, but neither those actions nor the benefit are contingent on what he himself is doing now. And if you point out that his likelihood of receiving blood in the future if he should need it is not significantly increased by his giving today, that will rightly be dismissed as irrelevant. He is not under the illusion that he is engaged in a trade" (Nagel, 1975, p. 63).

How would one classify this type of behavior? Nagel rightfully points out that it is not simple self-interest, nor simple altruism, for the explanation does refer to benefits received. What the person is doing is making
a contribution to an institution, hoping that it will benefit him at some time, and is dependent upon actions like this from himself and others like him in order to maintain an institution. He is not willing to be a free-rider but wants to do his share.

Rawls would classify this under a moral category itself, the sense of fairness; but it would be better to explain this as a part of limited altruism. If someone benefits from a practice to which others contribute, he is aware not only of his benefit but its relation to their prior actions, and to the general prevalence of the practice. He is in a particularly good position to realize how their failure to behave in this way would have affected him, and to be grateful for their participation in the practice. When he finds himself in a position to make a similar contribution, he can understand in terms of his own care how a failure to contribute would affect others, and can apply to himself the resentment or gratitude he would feel if the tables were turned. (Nagel, 1975, p. 64)

This capacity to put oneself in someone else's shoes is behind most altruistic behavior. It gives the impression of an exchange of services, but it is actually an experience that shows dependence upon other people; it shows an interchangeability of roles.

2. The second case, pointed out by Arrow (1972), is the desire to be the cause of benefits to others, as opposed to the mere desire that they benefit. People care about their causal relations in various ways; most people would rather benefit others than harm them. They do not want to be the cause of misery to others, nor do they want to
be harmed as a result of the benefit to others. The things matter to a person are not only his own welfare and other people's welfare, but also certain causal relations between his own welfare and that of others.

Arrow in his article "Gifts and Exchanges" (1972), offers three different hypothesis that could be used to explain this circumstance of altruistic behavior. They are as follows:

(1) The welfare of each individual will depend both on his own satisfaction and on the satisfactions obtained by others. We here have in mind a positive relation, one of altruism rather than envy.

(2) The welfare of each individual depends not only on the utilities of himself and others but also on his contribution to the utilities of others.

(3) Each individual is, in some ultimate sense, motivated by purely egoistic satisfaction derived from the goods accruing to him, but there is an implicit social contract such that each performs duties for the other in a way calculated to enhance the satisfaction of all. (p. 348)

This third hypothesis is in the spirit of Kant's categorical imperative and Rawls's theory. The first two distinguish between two levels of utility: each individual, according to Arrow (1972), "may be regarded as deriving satisfaction from the goods he receives, but his overall aim is to maximize welfare, a function of the satisfaction of all; he derives a utility from seeing someone else's satisfaction increased" (p. 348). The second differs from the first "in that welfare is, derived not merely from an
increase in someone else's satisfaction but from the fact that the individual himself has contributed to that satisfaction" (Arrow, 1972, p. 348). This first hypothesis has been used by economists to show why people give to others or vote for a redistribution of income. Arrow (1972) defines it this way: "under the first hypothesis I would prefer that you rather than I give to a third individual, but in the second case I might well prefer to give myself, because I would have the satisfaction of personal participation in social welfare" (p. 349).

These hypotheses do have a systematic bearing on stability as a whole; one can try to rationalize this type of behavior in terms of one of the first two hypothesis or in terms of the social contract hypothesis. The social contract theory as carried out by Rawls would lead it has been argued, to a basic instability; these types of causal relations are best explained by the utilitarian criterion, and this degree of fellow feeling is best explained by the doctrine of sympathy.

3. The third case for limited altruism was the inclusion of altruistic motives within the scope of service offered for sale. When applied specifically to economic life, one could doubt that altruism plays a significant role by arguing that there is a contradiction in the idea of paying someone to be altruistic. This contradiction does
not exist however; altruistic motives do exist in economic exchanges. Quoting Nagel (1975):

"Basically, it is an error to assume that what we pay for is always something within the control of the seller. If it is entirely in his control, and if he makes it available to us in exchange for money, his motive is gain and not altruism. But in fact, though a person may control the availability of something he produces or does in exchange for money, this does not mean that he controls the thing itself. (p. 66)

For example, when one hires a baby-sitter he pays them not only to perform certain tasks and exercise certain skills, one also pays them to act for certain motives which they cannot control at will. One hires a baby-sitter who likes children and cares about them, and who will treat them properly not just because she is going to be paid for it. In this way altruism or benevolence can be paid for. As a result, we cannot base human motives on self-interest alone, for there is something other than mere self-interest in what is being paid for. Therefore, altruism is involved in economic exchanges.

These three cases of limited altruism combined with the doctrine of self-interest represent a proper balance between egoistic and altruistic motivation, not only for economic man, but for political man as well. It is a much more realistic picture of how society actually exists. It is a society in which the citizens and legislators are motivated by their own self-interest, but not to the point of excluding all else; for these same citizens and
legislators do fight for altruistic objectives also. These cases explain why under some conditions rich people will support legislation benefiting the poor, and it also is able to show how many politicians and elected officials have followed their own moral convictions and have made unpopular decisions which may hurt their chances for re-election. In Rawls's just society, citizens and legislators are never motivated by their own self-interests or in the case of the legislator, by the selfish interests of their constituents. They are motivated by their strong sense of justice, a motivation that is not only unrealistic, but would also result in instability.
CONCLUSION

In developing a theory of justice, it is essential to acknowledge a certain imputed psychology and it is preferable to explain the choice as well as outline the character of the psychology that has been developed. The psychology suggested here is one which takes as given the demonstrated ability of human beings to cooperate with one another and which bases this capacity in the tradition of Hume, to a mechanism of sympathy which brings together individual egos to a social situation of a variety of social and political actors. With a focus of this kind, one can consider concepts such as freedom, equality, or political obligation in a society without having to explain it away, or conjure it up in mid air.

On the one hand, an unrestricted egoist philosophical psychology would lead to grave problems in the area of social and political obligation; despite the efforts of Hobbes, it is hard to conceive how it can generate any valid obligation at all. On the other hand would be a psychology of unrestricted altruism, a psychology in which universal values are the focus of all obligations and where the actions of individuals are only very rarely related to the self. Whatever its rhetorical worth, a
philosophical psychology of this sort is outweighed by the problems that would result with the findings of an empirical study.

What is needed is a theory based largely upon well-understood self-interest. A society in which everybody neglects his own interests, and is looking out for everybody else's interests would conceivably generate chaos. The same is true of a society that is dominated by a "sense of justice." The person who is best suited to be concerned for an individual's happiness is the individual himself. Altruism that is cut loose from self-love would render us helpless in trying to determine when someone is doing something good for anyone.

A society made up of men dominated by Rawls's sense of justice might conceivably exist and survive, but it has already been proven that there is no reason to suppose that more happiness would result in such a society than we already possess. People can make serious blunders in helping others and can be greatly unskilled in deeds to see that justice is provided; these deeds done with ignorance and lack of skill could lead to disaster. One could argue that it is presently such deficiencies that make our own society less enjoyable than it might be, and it is highly doubtful that incorporating us with Rawls's sense of justice would improve the situation in the slightest. Indeed, it has been shown to do the exact opposite.
In this long and involved critique of Rawls's theory, an airtight case for utilitarianism has not been provided. There is no way to justify the utilitarian ideal as an ideal for everyone. But neither can one justify any ideal that comes short of utilitarianism. Rawls's theory of justice is an attempt to develop a viable alternative to utilitarianism. By examining the psychological orientation of his theory, and by looking specifically at the maximin principle as his answer to the social choice function, we have found that Rawls does not succeed in offering a viable alternative to the utilitarian position. Instead, he offers a position in which public levels of motivation would be so high as to result in instability. Rawls goes beyond the moral psychology of Rousseau and transcends a sense of justice, a sense where the level of political activity would be so high as to render the operation of society as a whole chaotic; for historical experience has shown that motivational states of this kind only present themselves in rather short periods, and it is at these times of revolution or war in which the stability of a society is most likely to disintegrate. The burden of proof had been placed on Rawls to show that his altruistic assumptions, with the high level of political activity as depicted by the sense of justice, would not result in high levels of social instability. He fails
to adequately deal with this problem. As a result, the problems with Rawls's theory of justice can be traced to a most primary of misconceptions: his theory of man and the psychological principles of which this man is characterized.
BIBLIOGRAPHY


Bates, S. The motivation to be just. Ethics, 1974, 85(1), 1-17.


