

The Season of Fiction Is Over: A Study of the "Original Position" in John Rawls' A Theory of Justice

Philip A. Boynton

Follow this and additional works at: <http://digitalcommons.osgoode.yorku.ca/ohlj>
Article

Citation Information

Boynton, Philip A.. "The Season of Fiction Is Over: A Study of the "Original Position" in John Rawls' A Theory of Justice." *Osgoode Hall Law Journal* 15.1 (1977) : 215-249.
<http://digitalcommons.osgoode.yorku.ca/ohlj/vol15/iss1/7>

This Article is brought to you for free and open access by the Journals at Osgoode Digital Commons. It has been accepted for inclusion in Osgoode Hall Law Journal by an authorized editor of Osgoode Digital Commons.

THE SEASON OF FICTION IS OVER:¹ A STUDY OF THE "ORIGINAL POSITION" IN JOHN RAWLS' *A THEORY OF JUSTICE*

By PHILIP A. BOYNTON*

A. INTRODUCTION

*A Theory of Justice*² by John Rawls has proved to be a very major contribution to moral, political, and legal philosophy. That it was likely to do so was evidenced by Rawls' earlier works on justice and by the enthusiastic response of scholars to these.³ A clue to the theory's true significance, however, is perhaps provided by the breadth and diversity of the readership which the publication of the full work has attracted.⁴ No doubt this has been due, to some extent, to the unusually polyphonic view that Rawls takes of his subject.⁵ But one feels that, more than this, the perceived importance of the book lies in the fact that it seems to mark something of a return from moral philosophy's 'academic' preoccupation with semantic and analytical matters to the normative issue of the right and wrong way to behave.⁶

By subordinating questions of definition and meaning to a bold endeavour to develop "a substantive theory of justice,"⁷ Rawls appears to offer hope of basic moral guidance for our everyday activities. That he should do

© Copyright 1977, Philip A. Boynton.

* The author is a Sessional Lecturer at Osgoode Hall Law School. An earlier draft of this article constituted the written requirement for the Master of Laws Degree at the Harvard Law School, 1976.

¹ Jeremy Bentham, *The Theory of Fictions*, ed. C. K. Ogden (London: Kegan Paul, 1932) at 32.

² John Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1973). Hereinafter referred to as Rawls.

³ J. Rawls, *Justice as Fairness* (1958), 67 *Phil. Rev.* 164 and "Distributive Justice" in P. Laslett and W. G. Runciman, eds., *Philosophy, Politics, and Society* (Third Series, Oxford: Basil Blackwell, 1967).

⁴ See N. Daniels, ed., *Reading Rawls: Critical Studies of a Theory of Justice* (New York: Basic Books, 1974) at XI. Hereinafter referred to as *Reading Rawls*. My assessment of the popular impact and very general significance of Rawls' book owes much to Daniels.

⁵ Thus he examines in detail such questions as civil disobedience, conscientious objection, the relevance of social groupings such as the family to political institutions, and the nature of loyalty.

⁶ This distinction corresponds approximately to the contrasting concerns of meta- and normative ethics. The preoccupation of the former, it is suggested, with questions about the meaning of ethical concepts and the nature of judgements in which these concepts are used, has been perceived as unfortunate because of its related omission to provide guidance in the practical matters of moral obligation and goals of action.

⁷ Rawls at 579.

so at this point in the history of western democracy possibly represents, in much the same way as a number of the great ethical theories of the past have done,⁸ an ideological response to the particular exigencies of the period. This may be one of the factors which makes Rawls' work seem so significant. Certainly, taken together with the general proposition that the perpetuation of social order requires some sort of moral unity, the idea that contemporary secular society is itself suffering from a deficiency of moral unity which in the past has often been furnished by religion, suggests the possibility that Rawls' book may go some way towards meeting a serious social *desideratum*.⁹

Unhappily, the teachings of modern ethical theory do not augur well for Rawls' enterprise. Thus, it is not by chance that, subsequent to the Enlightenment and the empiricist¹⁰ rejection of the lingering belief both in intelligible essences and the corresponding absolutes of universal moral rules and standards of value, moral philosophers have become increasingly disinclined to undertake major projects of substantive moral inquiry.¹¹ Rather, this cautious attitude has been the direct result of a pervasive skepticism concerning the concept of objective value. In fact, as some writers have observed with alarm,¹² since the seventeenth century the strength of the skeptical movement has grown to great proportions. This has been so much the case, that even such disciplines as the natural sciences that are less obviously vulnerable than ethics to a debunking relativism¹³ have not been free from challenges to their complacent claims to objectivity. Given this powerful subjectivist skepticism, Rawls' bid to resurrect a quasi-ontological status of justice and thereby, as he admits, to make the same possible for morality in general, is

⁸ The moral and political philosophies of Hobbes, Rousseau, and Marx would be examples.

⁹ There are really two points being made here: firstly, that there exists in the West a general sense of moral disenchantment, possibly corresponding to a Durkheimian "anomic" state conditional on the emergent division of labour; and, secondly, that the American phenomena, particularly in the 1960's, of radical student militarism, race riots, law enforcement excesses, Viet Nam, Watergate, and the CIA scandals, threatened a serious disintegration of the liberal-democratic context in which *A Theory of Justice* was written.

¹⁰ The reference here is to the empiricist tradition, associated with the idea of the relativity of value and the elusiveness of a concept of the good, of Hobbes, Locke, Berkeley, and Hume.

¹¹ This disinclination only represents a general trend towards a fairly comprehensive skepticism which seems to be culminating in the twentieth century. Since the Enlightenment there have been, in fact, a number of fundamental attempts to salvage the necessary conditions of the good. As Professor Unger points out, the epistemology of Kant was the first in a series of major efforts, exemplified by the theories of Freud, Levi-Strauss, and Chomsky, "to build a science of man on the conceptions of a universal mental apparatus with inherent predispositions." Alternatively, there have been "those like Vico, Hegel, and Marx" who sought the humanly objective or universal in the teleological dynamic of history. R. M. Unger, *Knowledge and Politics* (New York: The Free Press, 1975) at 114. Hereinafter referred to as Unger.

¹² See I. Scheffler, *Science and Subjectivity* (New York: Bobbs-Merrill, 1967).

¹³ There are, of course, different versions of relativism. In this context, the term is employed to mean that no one of the innumerable moral judgements that different people make can in any way be objectively justified as against any other.

singular for its optimism in an area of epistemology whose history suggests little justification for optimism.

The significance for legal theory of the dichotomy between ethical objectivism on the one hand and relativism on the other, and of *A Theory of Justice* in relation to this dichotomy, is plainly fundamental and far-reaching. Thus, if one adopts an extreme, relativistic standpoint, it becomes very difficult to talk meaningfully about a judicial decision being just or unjust or even morally right or wrong. Of course, relativism does not necessarily preclude the concept of a right or wrong decision in a purely legal sense and, indeed, there are some who would identify justice with nothing more than the 'correct' application of legal rules.¹⁴ Yet even a thoroughgoing legal positivist who affirms the analytic distinction between law and morals would no doubt cherish the opportunity of assessing the value of laws, their goodness or badness, according to an external standard of objective morality.¹⁵ This is the advantage that objectivist theories like Rawls' enjoy.

In the 1930's, the emergence in the United States of the school of jurisprudence loosely known as "legal realism" seriously undermined the formalist notion that the legal process implements justice by the scientific and certain application of legal rules to situations of social conflict. All the same, the logical possibility of justice was well able to survive the realization that judicial decisions frequently have many determinants above and beyond, and sometimes quite unrelated to, legal rules.¹⁶ Theoretical justice may even survive the recognition that, because of the shifting and ambiguous meaning of language in general,¹⁷ and because, also, of the inexhaustible fund of human and social contingencies that constantly change the job that the law has to do, legal rules can never be exclusively determinative of all judicial decision-making.¹⁸ What the idea of justice cannot survive, however, (except in that

¹⁴ From the legal perspective, a partial meaning of "correct application" will often be found in the explicit language of the positive law addressed primarily to the general public. However, a full understanding of this procedural concept can only be achieved by reference to the secondary rules of law, sometimes of a constitutional status, addressed to legal officials. The more difficult question of the status of secondary rules is unfortunately beyond the scope of this paper, but for an authoritative elucidation of the matter see H. L. A. Hart, *The Concept of Law* (Oxford: Clarendon Press, 1961).

¹⁵ There is much confused and careless use of the term "positivism." For present purposes, all one need note is that legal positivists are logically free to adhere either to moral objectivism or to relativism.

¹⁶ The American "realists" were much more concerned to point out the existence of these determinants than to elucidate their intractable nature and, even now, relatively little is known about the antecedents of judicial decisions. For an early and somewhat primitive attempt to consider the problem from a psychoanalytic standpoint, see J. Frank, *Law and the Modern Mind* (New York: Tudor Publishing Co., 1935). More recently, the methods of the social sciences have been employed in studies of the judicial process. See G. Schubert, *Judicial Decision-Making* (New York: Free Press of Glencoe, 1963), and P. L. Rosen, *The Supreme Court and Social Science* (Urbana: University of Illinois Press, 1972). On the socialization of judges, see M. D. A. Freeman, *The Legal Structure* (London: Longman, 1974).

¹⁷ See Wittgenstein, *The Philosophical Investigations* (New York: Macmillan, 1968).

¹⁸ The more skeptical view would be that since no two cases are ever factually identical, and because the language of the positive law always requires some sort of interpretation, then there can be no instances at all of legal rules rendering the outcome of litigation completely certain.

virtue's most narrow and legalistic formulation) is an epistemology which denies the very existence of moral facts. From the relativistic perspective, a judge can no more make an unjust decision than he can make a just one; he can no more be wrong in his holding than he can be right.¹⁹

Thus far, Rawls' theory has been considered from the highly abstract perspective of its significance to the very important issue of justifying moral judgements. In seeking to establish a substantive theory of justice deriving justification or, to employ political terminology, legitimacy, from a basis in objective moral reasoning, Rawls obviously throws much light on this question. Needless to say, the failure of *A Theory of Justice* to achieve its avowed aim would not, of itself, conclusively resolve the interminable controversy surrounding the true nature of substantive moral theory. Nonetheless, following 'hard on the heels' of a protracted period of thoughtful moral skepticism, the failure of a grand objectivist effort like Rawls' would go some way towards confirming the skeptics' doubts. At any rate, an analysis of the particular issue of Rawls' attempt to establish an objective theory of justice and the relationship of this attempt to the general issue of justifying moral judgements of all kinds, will comprise the primary subject matter of the second half of this paper.²⁰ Before delineating the overall pattern of our inquiry, however, it may be instructive to provide a basic outline of the theory of justice. As will become more evident presently, the main perspective to be employed here and throughout the paper will be a methodological one.

A Theory of Justice is most obviously characterized as belonging to the tradition of the social contract. Accordingly, in order to formulate the basic principles of justice, Rawls asks what primary rules for the governance of society would be arrived at by free, equal, rational, and self-interested persons seeking to engage together in a social contract. The procedures and criteria adopted in answering this question are essential to the "original position" which constitutes the methodological foundation of the theory, and herein are contained all the elements, including Rawls' two serially ordered principles of justice,²¹ upon which the substantive details of the theory are based.

Given this method, one is quickly prompted to ponder the purpose of Rawls' atavism: why did Rawls resurrect the fiction of an original social

¹⁹ For a jurisprudential version of this sort of relativism, see the Scandinavian "realists": A. Hägerström, A. V. Lundstedt, K. Olivecrona, and A. Ross.

²⁰ Clearly, the second part of the inquiry is being undertaken at a high level of ethical abstraction. Since our concern is with Rawls' objectivism as a generic concept, no attempt will be made to engage in comparative analysis of rival objectivist theories such as utilitarianism.

²¹ Rawls' two principles of justice are, first, that each person has an equal right to the most extensive equal liberties compatible with similar liberty for all, and, second, that all inequalities in liberty can be justified only as being an advantage to the least advantaged. An enactment of a less extensive liberty must, if it is to be permissible, strengthen the total system of liberty shared by all; and an unequal liberty must be rationally acceptable to those citizens with the lesser liberty. The two principles apply to all primary social goods — that is, the things that rational people want most — and not just to liberty. Since, moreover, the principles are "serially ordered," the first takes absolute priority over the second.

contract which, despite its enormous political impact two hundred years ago, has long since been cast aside and ridiculed as a product of unbridled and useless idealism?²² And what was Rawls' object in introducing the concept of a hypothetical "original position"; did he intend that a theory of justice conceived therein should provide moral guidance for our actual behaviour? In attempting to answer these threshold questions, one would do well to remember at the outset that Rawls' avowed concern is solely with "justice as fairness."²³ Consequently, his formulation of the original position is calculated to entail fair procedures which will ensure that the social principles contractually agreed to are, in fact, principles of justice according to the "fair" conception. In particular, Rawls employs to this end a device he calls the "veil of ignorance." The fictional "parties in the original position" (P.O.P.s)²⁴ are located behind a "veil of ignorance" so that they are unaware of many of the facts relevant to their existence. As the author says,

. . . 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. . . . 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 . . . [furthermore,] the parties do not know the particular circumstances of their own society. . . .²⁵

Rawls argues that as a result of this ignorance the parties choosing rules and principles in the original position will be effectively deprived of the partiality usually associated with a known position in society and will therefore decide upon a distribution of goods that will be fair to everybody and universally recognized as such. They will do so, of course, not out of any sense of altruism, but out of self-interest and fear of the unknown. For while they remain unaware of their own position in society they dare not propose anything but an egalitarian social order lest they later find themselves in the role of the socially and economically downtrodden.

By this method, it is claimed, the veil of ignorance remedies the in-

²² Both the French and American Revolutions took, as much of their ideological basis, the material of the social contract theories, particularly Locke's concept of the natural rights of man. Since that time, however, the notion of natural rights flowing from a body of Natural Law has fared very badly in the constant war that has been waged between it and utilitarianism.

²³ Rawls contrasts his formulation of "justice as fairness" and the elimination of morally arbitrary distinctions made by the state in its distribution of social goods, with traditional utilitarian conceptions of justice. According to Rawls, the notion of fairness is missing from utilitarianism, which subordinates institutionally-imposed fair-dealing among society's members to the prior goal of maximizing the general welfare or happiness. Justice as fairness, on the other hand, has a special weight such that it is not to be justified by resort to some higher-order principle. In Rawls' view, the social contract expresses the element of fairness in justice, by allowing the parties mutually to acknowledge principles which have been established under conditions of identical freedom for everyone and to which, it is recognized, all are to be equally subject, in times both of advantage and disadvantage.

²⁴ Henceforth, the parties in the original position will be termed "P.O.P.s" after R. M. Hare, "Rawls' Theory of Justice," in *Reading Rawls* at 87. Hereinafter referred to as Hare.

²⁵ Rawls at 137.

justice of nature's unequal distribution of goods which, from a moral standpoint, is quite arbitrary. Factors such as heredity and the social standing of one's family need no longer be relevant to one's starting place in society. Hence, one's prospects in life are not blighted at birth. Furthermore, according to Rawls, the principles emerging from the original position will occasion the establishment of social institutions of such a type that, combined with the ideal notion of "pure procedural justice," will ensure fair treatment for everyone throughout his lifetime.

From this analysis, it appears that a major objective of Rawls' method is the establishment of conditions guaranteeing the morality and, more precisely, the justness of the social principles agreed upon. Put somewhat differently, the conditions of the original position are meant to ensure that the P.O.P.s adopt the moral point of view in their decision-making. For the sake of clarity, we may call this aspect of Rawls' method the justificatory function. Clearly, it is concerned with the radical issue, adumbrated earlier and to be discussed more fully in the latter half of the paper, of the nature and meaning of moral judgments.

Certain writers²⁶ have suggested that there is, in fact, something of a paradox in the "original position's" justificatory role. Their point is that if the justification of Rawls' two principles is to be inferred from the mere act of agreement, then the original position and the idea of self-interest will be superfluous as justificatory arguments. On the other hand, if principles other than Rawls' cannot be discounted at the outset as being self-evidently unfair, then it is difficult to see the justification for excluding information the unfairness or moral arbitrariness of which has not been objectively established.²⁷

Ronald Dworkin tries to extricate Rawls from this difficulty by arguing that the conditions of ignorance which the original position imposes are not arbitrary or intuitive, but have a justificatory power drawn from an objective "deeper theory." Such deeper theory, one should not be surprised to learn, is the 'right-based' theory of the great tradition of the social contract, particularly as espoused by Locke. Predictably also, these rights are natural rights springing from Natural Law, a concept which has long been closely associated with social contract theories. While it will be beyond the terms of reference of this work to debate the time-worn issues of the different theories of Natural Law, one is again reminded of how an assessment of the justification of Rawls' theory inevitably involves a conceptual entanglement with broader ethical questions.

If *A Theory of Justice* is most conspicuously significant for the support it attempts to give to the debilitated tenets of objectivist ethics, and if a general estimation of Rawls' argument must take serious account of this matter, there still remains the question of the theory's internal coherence.

²⁶ R. Dworkin, "The Original Position," in *Reading Rawls* at 21-2. Hereinafter referred to as Dworkin. Also, T. Nagel, "Rawls on Justice," in *Reading Rawls* at 6. Hereinafter referred to as Nagel.

²⁷ For a detailed analysis of this problem, see *infra* at pp 234-39.

Our particular concern at this point is with the analytic function of Rawls' method. Rawls, however, seems to emphasize the social contract's and the original position's justificatory role, and is thus preoccupied with demonstrating that the two principles which the P.O.P.s are said to choose are, indeed, just. He omits to consider the prior, analytic question of whether his method is capable of generating any sort of organizing social principles, regardless of their justness. In view of this omission, it will be necessary to examine the theory's analytic functionality closely.

The first task of this paper, then, will be to determine by careful analysis, whether Rawls is correct in supposing that it is possible to establish principles, adequately geared to the governance of modern society, by means of the relatively simple mechanism of hypothetical, de-individuated, rational, and self-interested persons selecting principles in a situation of ignorance as to their own particular interests in society. This exercise will, of necessity, entail a broader examination of the psychological condition of the P.O.P.s and of their capacity for making rational decisions generally.

It can be seen that for the purpose of thematic clarity, our inquiry may be conveniently divided into two sections. In the first, Rawls' theory may be criticized in its analytic capacity: that is, given the author's basic premises, and notwithstanding what is said about the superiority of this particular theory of justice over other accounts of that virtue, can it be demonstrated that the variegated and intractable problems of contemporary social administration are reducible to the simpler issues identified in the original position? Irrespective of the answer to this question, the aim of the second section will be to determine whether Rawls' particular brand of 'contractarianism' is capable of justifying the alleged principles of justice. Since the perspective of analysis to be employed in this paper is clearly methodological rather than substantive, and in view of the fact that the distinctive feature of the Rawlsian method resides in an adaptation of social contract theory to the particular form of the "original position," our critique may appropriately comprise a study, first, of the analytic role of the original position and, second, of the justificatory.

B. THE ANALYTIC ROLE OF THE ORIGINAL POSITION

What relatively minor attention Rawls pays the analytic role of the original position has already been mentioned, and nowhere is this neglect more dramatically evidenced than in his employment of the veil of ignorance. As will soon be made apparent, what the veil entails above all is a substantial interference with the personalities of the parties to the contract. Predicting the consequences, therefore, of the operation of the veil, which is precisely what Rawls is concerned to do, surely ought to involve a prolonged consideration of the concept of the human personality. Yet, as a specific subject for analysis, this is something Rawls eschews completely. Even the incidental treatment it receives is carefully limited to that necessary to support the argument for Rawls' sought-after two principles of justice. However, because in our examination of the original position we, too, are anxious to forecast the consequences of the imposition of the veil of ignorance, it will be incumbent upon us to consider seriously the nature of the human personality and the

relevance this has to the ability of the purblind 'contractors' to do the job Rawls assigns them. Since Rawls does not provide any overt or developed theory of personality, it will be helpful to approach this issue indirectly.

The first task will be to try to discover in a general way, and with no preconceptions as to the outcome of the inquiry, whether or not Rawls seems able to demonstrate that certain principles²⁸ for regulating social institutions are derivable from the original position. Such an investigation will undoubtedly lead to a consideration of the concept of personality and the question of the consequence of Rawls' very scant treatment thereof.

Once the discussion has reached this stage it will have become obvious that the problem of personality represents, in embryonic form, the more general issue of human nature. Hopefully, by reflecting on the notion of personality only insofar as it is pertinent to a thorough study of the analytic role of the original position, it should be possible to lure into the open the pre-supposed view of human nature which presumably lurks behind Rawls' theory. As a matter of convenience, then, this part of the critique may be subdivided, in an approximate way, into two sections: the analytic role of the original position assessed in the light, first, of personality theory and, second, of the theory of human nature.

1. *The Analytic Role of the Original Position and the Concept of Personality*

Rawls' main dilemma in employing the original position is that while he needs to strip the P.O.P.s of a sufficiency of their knowledge as to their individual and collective circumstances to ensure the formulation of those principles which he deems just, he must, at the same time, entrust them with enough information to facilitate normal processes of decision-making. The question then, is: how much knowledge-stripping should take place? As will soon become apparent, this is a problem with both psychological and logical components. For the moment, we shall be primarily concerned with the former. The logical problem of calculating what to put into the original position, bearing in mind that this will determine what ultimately can be extracted therefrom, will become of more critical significance upon a consideration of Rawls' attempt to justify his principles.

It has been suggested²⁹ that when writers employ in their theories men of a merely hypothetical character, those theories become immediately secure against criticism that takes as its basis a psychological assessment of real persons. *Prima facie*, this is an argument that might apply to the P.O.P.s, and

²⁸ For present purposes it need not matter whether or not these principles are just.

²⁹ See A. Stone, *Psychoanalysis and Jurisprudence Revisited* (1972), 10 Am. Crim. L. Rev. 357 at 374. In painting a generally gloomy picture of the prospects of a "major rapprochement between psychoanalysis and law" and in referring to the efforts of one legal philosopher in particular, Ronald Dworkin, to treat a basically psychological issue in a merely superficially or even non-psychological way, Dr. Stone suggests that philosophers may immunize themselves against criticism directed toward their theories from the psychological standpoint, by postulating in those theories men that are only hypothetical in character.

certainly it would go some way towards explaining Rawls' relative neglect, mentioned earlier, of the analytic function of the original position. Rawls has made it plain that "the conception of the original position is not intended to explain human conduct except insofar as it tries to account for our moral judgments . . ." ³⁰ and he may, for this reason, hold psychological arguments irrelevant. ³¹ However, because the generation of social principles out of the original position depends so fundamentally on the P.O.P.s' psychologies functioning in a manner which, as will be seen later, is archetypally or essentially human, although under abnormal conditions, this defence would seem to fail. In point of fact, the "great weight" which Rawls places on "the facts of human psychology" ³² is apparent at numerous points throughout the text, especially in those areas, largely in Part III, which are devoted to arguing for the enduring quality of the two principles of justice.

First and foremost, of course, the P.O.P.s are assumed, despite their pared personalities, to be capable of making normal judgments of self-interest. ³³ At any rate, they have enough knowledge of general moral psychology to be able to perceive the overall significance of their particular selection of principles, ³⁴ especially as regards the latter's prospects for inspiring adequate loyalty when "actually adopted" ³⁵ and put into application. In other words, in addition to assuming an approximately normal functioning of the P.O.P.s' psychologies, Rawls infuses into the original position a conscious commitment to the laws of moral psychology, and since the P.O.P.s are obliged to take this into account in protecting their individual interests in the contract, it cannot avoid finding expression in the principles themselves.

Finally, the most cogent evidence of Rawls' reliance on the facts of human psychology occurs when he argues that "the most stable conceptions of justice are presumably those" that rest on certain human tendencies which are particular expressions of "a deep psychological fact" ³⁶ of the human condition. It does not matter what these tendencies are or whether they do, indeed, express "a deep psychological fact" of the human condition. What is important about Rawls' argument is the logical implication that the creative capacity of his hypothetical men rests far too solidly on the psychological dispositions of real persons for it to be arguable that the original position, due to its imaginary form, is impervious to psychological criticism. Consistency demands that Rawls cannot have 'the best of both worlds.'

This leads, then, to an initial estimation of the plausibility of the original position as a means of arriving at principles for establishing just social institu-

³⁰ Rawls at 120.

³¹ See B. Barber, "Justifying Justice: Problems of Psychology, Politics and Measurement," in *Reading Rawls* at 300. Hereinafter referred to as Barber.

³² B. Barry, *The Liberal Theory of Justice* (Oxford: Clarendon Press, 1973) at 17. Hereinafter referred to as Barry.

³³ Dworkin at 47. Dworkin accepts Rawls' assumption in this matter. It will be considered shortly whether he is correct in doing so.

³⁴ See Barry at 17.

³⁵ Nagel at 7.

³⁶ Rawls at 494-95.

tions. It will be appropriate to begin by examining Rawls' artifice, the veil of ignorance. Certain critics have suggested that the veil is basically a device for establishing, much as "ideal observer" theories do, fundamental conditions of impartial judgement.³⁷ However, as will become clear later, what is controversial here is the means by which Rawls claims to be able to achieve impartiality.³⁸ Before contemplating, though, the justificatory power that the veil of ignorance is believed to confer upon the original position, we must see how the veil affects the latter's analytic function. As intimated earlier, one feels impelled to question the somewhat glib assumption that the P.O.P.s would be possessed of the usual powers of ratiocination and decision-making despite their ignorance of their personal situations. We know that the rationality of the P.O.P.s is essential to eliciting the two principles of justice from the original position, but it may transpire that the notion of rationality is incompatible with the knowledge-stripping method. Since the aim of depriving the P.O.P.s of so much knowledge about themselves is to render them universal men, thereby purging them of their usual partiality, the basic function of the veil of ignorance is to deindividuate. This is absolutely in accord with Rawls' foremost intention, of course. What has to be asked, however, is whether or not there are repercussions of deindividuation that go beyond Rawls' design. For the moment, there is no need to question that the veil accomplishes the requisite impartiality. What seems a more immediate threat is the possibility that the effects of knowledge-stripping are not restricted to those intended; that in respect of the original position's analytic role, at least, Rawls has been too strict in the exclusion of information.

The chief respect in which the veil of ignorance threatens to achieve too much centers upon the mutilation, which the deindividuating process of knowledge-stripping seems bound to effect, of the P.O.P.'s' personalities. Since the development by Freud of the theory of the unconscious and the associated structural concepts of the id, ego, and superego, it has been generally accepted that the human personality is a phenomenon which evolves to some extent during the individual's lifetime. Most of the contention that has surrounded this matter has sought to know not whether the personality can be moulded by the individual's experience over and beyond what is determined innately, but rather just how much it can be moulded thus. These common-places of psychology are adverted to here simply to emphasize that because the concept of personality is to a certain degree dependent upon a congeries of social factors, it cannot help but assume a highly differentiated, compound form. Consequently, one should expect that the effects of interference will be compound in nature. This, it is suggested, is liable to render the task of comprehensively forecasting what they will be extremely difficult and, in instances of drastic interference, probably impossible. For this reason, R.M. Hare concludes "that it is a wide open question how the P.O.P.s would choose; he [Rawls] has reduced the information available to them and about them so much that it is hard to say what they would choose. . . ."³⁹ With

³⁷ Barry at 12 and Hare at 93.

³⁸ The impartiality of the P.O.P.s in the original position is a question that will be considered in the section *infra* on the justificatory role of the original position.

³⁹ Hare at 106-07.

regard to the veil of ignorance, Rawls is silent on these difficulties. It must be supposed, therefore, that he takes for granted that the process of knowledge-stripping accomplishes as much and no more than is intended. It is our task now to determine whether he is justified in this assumption.

Whilst the possibility may be conceded of there being certain types and degrees of knowledge, relating to personal and social circumstances, which someone could be deprived of without any significant distortion of his personality, and whilst the effects of such a deprivation might, to some extent, be predictable, the case in hand is plainly not of that class. The operation of the veil of ignorance is so vigorous that it is unnecessary to decide at this point, where, if some sort of distinction between essential and inessential knowledge were thought feasible, a line of demarcation might be drawn. Since it seems certain, then, that the enforced ignorance of the P.O.P.s will have repercussions not overtly contemplated by Rawls, it will be imperative to identify these and determine their effect, if any, on the functioning of the original position. In particular, one must ask whether the deindividuated P.O.P.s would be capable of rational thought and self-interested decision-making.

Clearly, any effort to gauge the impact of deindividuation on the creative role of the P.O.P.s must take account of the concept of personality. The trend in modern personality theory has been away from the Cartesian dualist view which perceives a logical separation in the personality of the mind from the body, towards a physical monist conception. However, there are certain variations on the theme of physical monism and some of these will have to be examined here. The dualist interpretation will also have to be given consideration since, at first sight at any rate, it may appear more favourable to Rawls' theory.

A feature of much physical monist thought on the concept of personality is the idea of the essentiality of an individual's situation in society to his personal identity. Indeed, it has been asserted that "the conception of being a person always includes an account of relations among persons."⁴⁰ Professor Stuart Hampshire, also an adherent to this view, sees human action, being the chief determinant of one's social circumstances, as the key to the problem of personal identity.⁴¹ Thus, he says,

[w]e find our intelligence and our will working and expressing themselves in action, and at a particular time, and just these movements, or this voluntary stillness, are unmistakably mine, if they are my actions, animated by my intentions . . . I can only be said to have lost a sense of my own identity if I have lost a sense of where I am and what I am doing.⁴²

Now, is not this situation manifestly analogous to that in which the P.O.P.s find themselves? The 'heavy-handed' method of the veil of ignorance suggests that it is.

⁴⁰ Unger at 193.

⁴¹ S. Hampshire, *Thought and Action* (London: Chatto and Windus, 1959). Hereinafter referred to as Hampshire.

⁴² *Id.* at 75.

How, then, might the P.O.P.s' loss or absence of a sense of identity affect their functioning in the original position? To start with, it must be realized that only by way of a recognition of his discrete physical self in its relation to other objects around him can an individual have any conception of the society he inhabits. In the same way that "[t]he experience of standing apart from others is the most basic condition of selfhood,"⁴³ so, also, is it an essential ingredient of the individual's objectivity. If this were not so there could be no distinction between the individual *qua* individual and the individual as a component of society. His subjective and objective roles in society would be conflated. Similarly, A.J. Ayer interprets Hampshire as arguing that man is able to form the idea of an objective world only because he is himself a physical object, with a situation in time and space, and with powers of action and movement which cause him to come into contact with other physical objects, and particularly other beings with whom he can communicate.⁴⁴

The implications of this idea for the veil of ignorance and the original position are immediately apparent. Bereft of all knowledge of their own position in society the P.O.P.s may not be able to conceive of society at all and, at best, can have only the most impoverished and distorted conception. Consequently, it would seem that they are ill-equipped to contemplate society's problems. If this is so, then the suggestion that they should actually formulate principles for the foundation and administration of a just social union is unwarranted. Against this, it might be said that because the principles alleged to arise out of the original position are, according to Rawls, simply expressions of the P.O.P.s' natural psychological tendencies, then they are not dependent upon a conception of society at all. But this will not do. Since the two principles of justice are only developed out of the P.O.P.s' desire to protect their interests located in actual society, it may hardly be contended that they would have any motive for imposing safeguards if they had no notion of society in the first place. Also, it is doubtful whether, effectively lacking any experience of the world external to their own minds, the P.O.P.s would even have a capacity for simple thought. However, this is an issue which may be more appropriately resolved at a later juncture. For the moment, we may conclude that, insofar as Professor Hampshire's concept of a person is acceptable, it does appear seriously to undermine the analytic role of the original position. Before making any final judgement, however, it is important to consider whether other interpretations of 'persons' may more successfully accommodate Rawls' argument.

P. F. Strawson, who may be classed, with A. J. Ayer, as a modified physical monist, argues that the concept of a person is primitive and therefore not analysable in terms of the smaller elements of bodies or mental substances.⁴⁵ On the contrary, the subject to which the characteristics of consciousness are ascribed must be equated with that to which is also ascribed a

⁴³ Unger at 56.

⁴⁴ A. J. Ayer, *The Concept of a Person* (London: Macmillan, 1968) at 93. Hereinafter referred to as Ayer.

⁴⁵ P. F. Strawson, *Individuals* (London: Methuen, 1959) at 101-04. Hereinafter referred to as Strawson.

set of physical characteristics — and this is what may be called a person.⁴⁶ It follows from this that the physical properties which are treated as delineating a person remain essential to the concept. Indeed, Strawson asserts that when we define a person by ascribing experiences to a subject and when we also infer those experiences from his physical condition or outward behaviour, then the ground of that inference, that is, the existence of a subject's physical characteristics, does, in fact, constitute a "logically adequate" criterion for the attribution of personality. Strawson draws an analogy which is illustrative of this idea. He says:

[i]f one is playing a game of cards, the distinctive markings of a certain card constitute a logically adequate criterion for calling it, say, the Queen of Hearts; but in calling it this, in the context of the game, one is ascribing to it properties over and above the possession of these markings. The predicate gets its meaning from the whole structure of the game.⁴⁷

Thus one cannot fully conceive of the Queen of Hearts without conjuring up ideas about its position in the pack and its role in games. Stated somewhat differently, this means that the identification of the card as the Queen of Hearts implies much more than a mere pictorial design on the surface. It implies that this is a card that ranks as a certain power in relation to the rest of the pack, that occupies a particular position in the series of cards that constitutes a suit, and that has various other characteristics besides. Moreover, the identity of the Queen of Hearts will alter according to whatever game is being played and the role that the card has therein.

Strawson employs this analogy to throw light upon the connection between a person's inner experience and its outer perceptible expression. It may be usefully extended, however, to elucidate the relationship between personal identity and social association. Whereas the Queen of Hearts is identified by its position in the pack and its role in games, the human personality is identified by its role in society. How else may someone be identified than by the 'impact' he makes on his environment? How, for instance, might one identify Winston Churchill other than as 'Winnie,' son of Randolph, member of the nobility, cynosure of British politics and defender of liberty and the Empire? What would it mean to know Churchill if not as War Prime Minister, the ebullient, cigar-smoking, Tory and popularizer of V-signs? Without this and other similar information one could only have a very restricted idea of who Winston Churchill was, and if one were completely deprived of an acquaintance with the physical impressions that Churchill made on the world, it seems unreasonable to suggest that one could identify a person in the name at all. Yet the lack of such an acquaintance would not affect only other people's identification of Churchill. It would also affect Churchill's identification of himself. Under these circumstances he would be bound to lose all sense of his own identity.

In certain respects Strawson's approach is similar to Hampshire's. Both stress the importance of social interaction to the idea of personality. Hamp-

⁴⁶ *Id.* at 102.

⁴⁷ *Id.* at 110.

shire claims that even young children do more than simply receive sense data from the external world. Taking his cue from Freud and the notion of the opposed "pleasure" and "reality" principles, Hampshire suggests that infants acquire a sense of self by perceiving the constraints imposed upon the fulfillment of their wishes by other selves and objects generally. Without this social association, an individual will lack a conception of an objective world and, as we have seen, such a conception must be in fact the correlative of a sense of self. An effectively similar conclusion may be drawn from A. J. Ayer's moderated physical monism, even though it places less emphasis on action. Ayer holds to the "no-ownership doctrine of the self" first coined, but then rejected, by Strawson. This doctrine maintains that "a person's ownership of states of consciousness consists in their standing in a special causal relation to the body by which he is identified."⁴⁸ Put slightly differently, the fact that one's experiences "are the experiences of the person that one is, depends on their being connected with this particular body."⁴⁹ Ayer differs from the strict physical monists, therefore, in supposing that while one's identity is causally linked to a body, such body need not be one's own, though as a matter of contingent fact it always will be.

Prima facie, the insistence on embodiment need not, of itself, embarrass the P.O.P.s since Rawls does not demand their disembodiment. From the foregoing discussion, however, it should be apparent that because the self depends so crucially on a perception of separation from others, then embodiment such as that of the P.O.P.s, in an abstract, pre-social void, with no criteria of measurement, could furnish only the most diluted or adumbrated sense of personal identity. Embodiment, wherein the body represented nothing more than a useless husk or appendage to the mind, would no more conduce to proper personality development than would a totally disembodied consciousness. The conclusion to be drawn from this analysis seems to be that there is no variant of the physical monist theory of personality that does not strongly suggest that the veil of ignorance would operate substantially or even completely to erode the P.O.P.s' sense of personal identity. Since the notion of self-interest which is said to motivate the P.O.P.s in their choice of principles is quite inconceivable in the absence of the logically prior concept of self, the analytic role of the original position looks to be critically jeopardized.

On its face, the dualist interpretation of personality is better able to accommodate the P.O.P.s' analytic function. Rawls' assumption that the application of the knowledge-stripping, deindividuating process of the veil of ignorance would not impair the P.O.P.s' capacity for decision-making perhaps suggests a reliance on a radical divorce of the mind from the body, such that the physical abstraction of the P.O.P.s from society would not automatically render their minds practically inoperative. This view would certainly show some consistency with the essentialist conception of human nature which will later be attributed to Rawls.⁵⁰ Unfortunately, not even resort to an abandoned

⁴⁸ Ayer at 116.

⁴⁹ *Id.*

⁵⁰ *Infra* at p. 230.

Cartesian dualism can salvage the original position in its analytic capacity. For although it may be logically possible that the mind could continue to function in a disembodied state, albeit in a totally academic fashion, it is not the case that the self would be simultaneously preserved by it doing so. The P.O.P.s' effective disembodiment is as damaging to their sense of self, and thence to their self-interest, according to the dualist theory of personality as it is according to the physical monist. Thus, it seems the whole gamut of personality theory fails to support the original position within which the indispensable concept of self-interest cannot be sustained.

Benjamin Barber refers briefly to the same problem of psychology when he states that

. . . there is a considerable question . . . whether it is possible to conceive of men as having a hypothetical knowledge of what it means to have interests and desires without having particular interests and particular desires. . . . At the level of psychology it seems possible that particularity is built into the notion of interest and that it cannot be cut away without rendering interest unintelligible.⁵¹

Barber goes on to suggest⁵² that Rawls is conscious of this difficulty and responds to it with the idea of "primary goods." Rawls' argument is that, although the P.O.P.s are not allowed to know their particular wants, they do realize that they will have wants, and that therefore they should do all they can at present to facilitate the promotion of those wants later. Primary social goods are intended to be seen, basically, as the means by which the later advancement of particular interests is made possible. As Rawls puts it, the P.O.P.s rationally attempt to "win for themselves the highest index of primary social goods [mainly liberties, opportunities, powers, and wealth], since this enables them to promote their conception of the good most effectively whatever it turns out to be."⁵³ It is in this sense that the P.O.P.s are thought to be prompted by self-interest to formulate the two principles of justice. This 'instrumentality' argument, however, only works on the pre-supposition that it is possible to experience interest of some sort, in the absence of particular interests stemming from an awareness of the total relationship between the self and society. Yet, as has already been noted, such a notion of interest is quite untenable. Locating the P.O.P.s one step removed from their particular wants is not sufficient to render the analytic function of the original position practicable. Whether "primary goods," the distribution of which the two principles of justice are intended to regulate, constitute means or general wants on the one hand, or ends or particular wants on the other, is irrelevant to Rawls' difficulty, since in either case the notion of interest remains indispensable, yet impossible in the absence of particularity that characterizes the original position.

Even if the P.O.P.s' self-interest were maintainable, there must still be a serious doubt concerning their ability to perform their 'legislative' function under the conditions of the original position. For as noted earlier, the lack of a conception of society, which appears to be an ineluctable entailment of

⁵¹ Barber at 295.

⁵² *Id.*

⁵³ Rawls at 144.

the veil of ignorance, may preclude the possibility of the P.O.P.s engaging in any form of rational activity. Since their existence has been reduced to the vacuous and academic, it is difficult to imagine what source their thought processes might have. Indeed, it is problematic whether reason could operate at all under conditions that are tantamount to chaos. If this is so, then the original position must fail in its analytic capacity on two counts: first, because the P.O.P.s would probably be incapable of thought that was to any degree rational, and second, because any thought that was engendered in the original position would not be self-interested.

2. *The Analytic Role of the Original Position and Human Nature*

It was suggested earlier, and has been implicit throughout the preceding discussion of personality, that the veil of ignorance depends for its operation on an essentialist conception of humanity. Such a conception makes a distinction between what is essential to humans and what is inessential, so that man is viewed as comprising an essential human core with varying levels of contingent accretions largely, if not exclusively, arising out of the process of socialization. Insofar as this picture accurately represents the human condition, it is theoretically possible to strip away the inessential layer of man and still be left with a distillate which is not only recognizably human, but the very essence of humanity.

It is not difficult to understand the attractions of such a conception of humanity. Two centuries ago, social contract theorists employed the distinction between the humanly essential and the humanly contingent as a means of access to what they took to be human nature or the essence of man. The utility of the distinction, it was thought, lay in the fact that whatever decisions were made in the context of the contract would have a special moral force because they were expressive of human nature. The manner in which Rawls uses the idea of the contract as a justificatory device for the two principles and whether or not he is successful in so doing, are matters that will be discussed in the next section. At the moment, it is appropriate to consider the more basic issue of the validity of the essentialist conception of humanity and the way in which this bears upon the capacity of the original position to perform its analytic function.

“Essentialist theory,” as Professor Unger points out, “forces one to accept the classical doctrine of intelligible essences and all the untenable ideas about mind and society that doctrine implies.”⁵⁴ From the perspective of determinism, upon which all modern science is based,⁵⁵ there can be no justification for suggesting that the attributes a person acquires through his social experience are any less necessary than those with which he is born.⁵⁶

⁵⁴ Unger at 194.

⁵⁵ B. F. Skinner, *Science and Human Behavior* (New York: Macmillan, 1964) at 6. “If we are to use the methods of science in the field of human affairs, we must assume that behavior is lawful and determined.” Hereinafter referred to as Skinner.

⁵⁶ Rawls, of course, does not make an explicit distinction between essential and inessential human characteristics. In assessing his implicit reliance on this sort of distinction, however, we may, for the sake of argument, infer an approximate identification of the essential with the innate and of the inessential with the socially conditioned.

Of course, it is conceivable that all people have certain basic traits in common. But even if this is so, obviously it does not, by itself, justify an assumption that the job of separating out essential and inessential features is to be done most plausibly or validly, at the level of the unsocialized individual in the original position. Yet, as the second section of this paper will make clear, the justificatory role of the original position involves Rawls in just such an assumption.

An awareness of any limits to our predetermined conduct is naturally very important to theories of justice, associated as these are with the elusive concept of desert. This is made clear by Rawls' invocation of the veil of ignorance, the primary role of which is to remedy the injustice of nature's morally arbitrary distribution of goods. As Rawls states the idea, "no one deserves his place in the distribution of native endowments, any more than one deserves one's initial starting place in society . . . [C]haracter depends in large part upon fortunate family and social circumstances for which he can claim no credit."⁵⁷ Yet is it not a logical extension of this argument that, since everything a man does is ultimately a product of the forces that formed him and for which he is not responsible, namely his innate attributes and his environment, then there is nothing for which he can rightly claim credit? Ultimately, 'nature' must be acknowledged as responsible for all of man's achievements and, *a fortiori*, for all of his attributes. In this respect, therefore, the distinction which Rawls draws between the essential rational being and his contingent qualities, including his predispositions and abilities as well as his wants and interests, is fundamentally arbitrary.

Another difficulty with the essentialist conception of human nature relied on by Rawls, is the fact that the supra socio-historicism that it involves rests on the assumption that human behaviour consists only in those elements which are shared by all other people, irrespective of their socio-historical situation. Thus, once it is established that the further social development, through time, of an individual's personality is a phenomenon that is not only contingent but also logically independent of his essential human nature, then the social and historical evolution of such nature is precluded at the outset.

The true import of essentialism is perhaps most effectively to be illustrated from this evolutionary perspective. On a strict interpretation of the essentialist thesis, human nature must be identical now not just with what it was, say, in prehistoric times, but even with the nature of ancestors remoter still. Even if we restrict ourselves to *homo sapiens* we should be quite in accord with essentialist theory in extending the period of identification to many thousands of years. Moreover, insofar as human evolution represents a continuation of more primitive primate and other forms of animal life, there seems to be no objective reason why we should not go back further still. If the idea of an essential human nature is still tenable at this point, it is evident how reduced and primitive a concept it would have to be, to have survived intact such an epic development. In fact, it would have to be so inchoate that there must be serious doubt whether it would be able to

⁵⁷ Rawls at 104.

generate even the most basic principles for the administration of a modern social system. Besides, can one seriously believe that Rawls would deem his two principles of justice as appropriately applicable to both Neanderthal Man and twentieth century liberal Americans? According to Rawls' criterion for admittance to the original position, Neanderthal Man would almost certainly be excluded.⁵⁸ Yet it seems to be Rawls' foremost intention to produce a model of justice that conforms as closely as possible to the essential nature of this rude forbear of contemporary humanity.

In view of the objections outlined above, one might wonder why the essentialist conception of human nature ever attracted support. Perhaps a major, if negative, reason is that it avoids certain 'problems' that beset relativism, the other main theory of human nature.⁵⁹ Relativism, in this context, basically avers that man's nature consists in the specific patterns of his social circumstances. Traditionally this view is associated with the socio-historicist belief, nurtured by modern social theory, that because the nature of man corresponds to the form of society, it must vary in accordance with social changes occurring through time and space.⁶⁰ The primary drawback of relativism is that the simple identification of human nature with societies and individuals at any given time and place leaves us devoid of objective standards by which one society or individual may be compared critically with another. It may be argued, however, that relativism is disapproved mainly because, more than essentialism, it is associated with determinist theory and all the positivistic objectionability which that entails,⁶¹ notably, the want of human freedom and the ultimate meaninglessness of the notion of moral responsibility. Once it is realized how vulnerable many of the arguments against determinism are, once it is accepted that the basic idea of determinism — that every event has a cause — is consistently acted upon by almost all ordinary people almost all of the time, as well as by researchers in every realm of science, and once, finally, it is admitted that the argument for indeterminism is primarily an emotional plea, then the case against relativism loses much of its cogency.

⁵⁸ Rawls at 505. The condition of entry to the original position is a sense generally of morality, and particularly of justice.

⁵⁹ Relativistic theories of human nature are not to be identified with ethical relativism. Although ethical relativists generally accept relativist analyses of human nature, the principles of ethical relativism do not represent necessary entailments of a relativistic view of human nature.

⁶⁰ Unger at 194. My discussion of the supra socio-historicist and socio-historicist conceptions of human nature owes much to Professor Unger's analysis of these matters.

⁶¹ Skinner at 6-7 and 10.

[The] possibility [of determinism] is offensive to many people. It is opposed to a tradition of long standing which regards man as a free agent . . . [N]o one who is a product of Western civilization can accept [the deterministic view of human behaviour] without a struggle.

. . . The conception of a free, responsible individual is embedded in our language and pervades our practices, codes, and beliefs. Given an example of human behavior most people can describe it immediately in terms of such a conception. The practice is so natural that it is seldom examined. A scientific [or deterministic] formulation, on the other hand, is new and strange.

Professor Unger, however, urges that we reject traditional relativism and embrace an alternative thesis which might be viewed as a variant of essentialism or a hybrid of the essentialist and socio-historicist conceptions of human nature. Thus, in divorcing essentialism from its customary partner, supra socio-historicism, Unger suggests that "the kernel of humanity can be conceived to evolve in history, and the relations that define the individual or the species may be regarded as timeless."⁶² He goes on to say that,

[I]f one takes human nature as the universal and its manifestations in forms of social life and individual personalities as particulars, one may say that human nature is neither an ideal entity that subsists in its own right nor a mere collection of persons and cultures. Instead, it is a universal that exists through its particular embodiment. . . . Each person and each form of social life represents a novel interpretation of humanity, and each new interpretation transforms what humanity is . . . [Finally, this view allows us] to affirm the existence of a unitary human nature though we deny that this nature stands above history.⁶³

For present purposes it is unnecessary to cast judgement on this thesis, though it may be worth pointing out that it seems metaphorical rather than descriptive and therefore provides little guidance for the factual identification and differentiation of universals and particulars. But that is by the way. Of far more significance is the fact that, despite embracing the possibility of "a unitary human nature," Unger's thesis departs sufficiently from traditional essentialism to be fatal to Rawls' use of the veil of ignorance. This will become apparent from a brief consideration of the way in which the original position and its underlying essentialism are meant to justify Rawls' two principles.

Perhaps the most important reason why Rawls uses an essentialist conception of human nature as the foundation of his theory, is that by so doing he can claim to be regarding the principles of justice from "the perspective of eternity."⁶⁴ Apparently, therefore, Rawls equates essentialism with supra socio-historicism. Whilst a timeless conception of justice, even if theoretically possible, might not, in the long term, prove either appropriate or practically helpful to the vicissitudes of social arrangements and the accompanying modifications of the human condition, it is nonetheless easy to appreciate the attractions that such a conception has. By setting up a clearly defined, ultimate, or perhaps penultimate goal for the endeavours of humanity, much would certainly be done for the cause of social efficiency. The ultimacy of the goal, moreover, might inject into social and individual existence a meaningfulness which a transient conception of justice would be less qualified to provide. More obviously still, a supra socio-historical notion of justice would more closely correspond to the customary conception of justice as a constant moral prescription, rather than something resembling a social expedient.

Unquestionably, much can be said for the justificatory and other values of absolute, timeless justice. It is evident, however, that the claims of rela-

⁶² Unger at 194.

⁶³ *Id.* at 195.

⁶⁴ Rawls at 587.

tivism are too powerful to be ignored altogether. Professor Unger thus feels compelled to accommodate within his theory the idea of the essentiality of man's social associations to the concept of humanity,⁶⁵ and he can only achieve this, and simultaneously preserve the notion of a unitary human nature, by rejecting the supra socio-historicist contention. In view of the aforementioned implications of supra socio-historicism, it is not surprising that he adopts this course. In any event, even Unger's compromising theory allows no room for Rawls' brand of essentialism.

To conclude the first half of our inquiry, then, it must be said that the original position fails to fulfill the analytic role assigned to it. Rawls is unable to demonstrate that the complex task of formulating principles of justice for the governance of modern, heterogeneous society, is reducible to the relatively simple matter of self-interested individual choice. This is so for two reasons: first, Rawls takes no account of the compound effects on the individual psychology of the attempt to pare the P.O.P.s down to their 'essential,' and therefore morally relevant, human characteristics; and, second (and this is clearly related and even prior to his first oversight), he simply takes for granted a conception of the human condition that makes an illegitimate distinction between essential features of humanity on the one hand, and contingent features on the other. It remains to be seen whether this view of human nature is as damaging to the justificatory role of the original position as it is to the analytic.

C. THE JUSTIFICATORY ROLE OF THE ORIGINAL POSITION

1. *The Reliance on Human Nature*

A superficial examination of the justificatory role of the original position might suggest that Rawls manages to avoid any appeal to human nature.⁶⁶ His concern, rather, is with the identification of criteria relevant to the determination of justice, and it is not immediately apparent that such criteria need be related to the conditions of human nature. The veil of ignorance, after all, ensures the P.O.P.s' ignorance of their separate notions of the good, so that those whose predilection is one ideal will be prevented from favouring it over that of others. Similarly, the P.O.P.s are deprived of all knowledge relating to their innate talents, psychological dispositions, and social situations, since this, Rawls believes, is quite arbitrary from a moral

⁶⁵ Marx seems to have a similar view of human nature:

[t]he *human* essence of nature exists only for *social* man; for only here does nature exist for him as a *bond* with other *men*, as his existence for others and their existence for him, as the vital elements of human reality; only here does it exist as the *basis* of his own *human* existence. Only here has his *natural* existence become his *human* existence and nature become man for him. *Society* is therefore the perfected unity in essence of man with nature. . . .
Karl Marx, *The Economic and Philosophical Manuscripts in Early Writings*, ed. Q. Hoare, trans. R. Livingstone and G. Benton (New York: Vintage Books, 1975) at 349.

⁶⁶ M. Fisk, "History and Reason in Rawls' Moral Theory," in *Reading Rawls* at 54. Hereinafter referred to as Fisk.

perspective. In sum, the original position is meant to rule out "those principles that it would be rational to propose for acceptance, however little the chance of success, only if one knew certain things that are irrelevant from the standpoint of justice."⁶⁷ The question which must be asked now is: what determines the relevance of the conditions of the original position to the concept of justice? What, in other words, makes the principles allegedly elicited principles of justice?

The answer which Rawls supplies is that of reflective reason: in the final analysis, the pertinence of the conditions of the original position depends upon the extent to which those conditions generate a conception of justice which harmonizes most with our firmest common-sense convictions regarding that virtue. When a moment of perfect harmony is achieved (and a moment is all it is likely to be) that will be a moment of "reflective equilibrium."⁶⁸ Later, we shall study the implications which this justificatory mechanism has for Rawls' purported repudiation of intuitionism. For now, it is enough to notice that because, in Milton Fisk's words, "our reflective sense of justice is the ultimate arbiter on admittance to the original position,"⁶⁹ it becomes crucial to the position's power of justification. The question to be considered, therefore, is whether or not reflective reason is able to impart to the original position the justificatory authority it requires. But first it is necessary to ask why the original position requires an external source of justification at all.

It was noted earlier that contract theory was customarily employed as a means of access to a rational element in humanity, this in turn being thought to generate principles of peculiar moral force. This method, however, confronted a recurring puzzle: which constituents of the human situation were to be omitted in the quest for the kernel of humanity? In our assessment of the merits of Rawls' attempt to reduce the problems of social justice to a matter of individual choice of principles, it was found that a serious danger, when distinguishing the necessary attributes of humanity from the contingent, was the possibility that omitting all those features regarded as chance accretions resulting from social experience would leave insufficient material, either to yield the desired formula of justice, or, as our investigation showed to be far more probable, to yield anything at all. On the other hand, admitting more information would involve courting another sort of danger: if historically specific information were included, the original position would no longer be defined from the standpoint of eternity and would therefore lack the capacity to confer an enduring moral weight upon the principles agreed to.

Such is the perennial dilemma of the social contract theorist and also, it would appear, of Rawls.⁷⁰ The latter does not explicitly admit, of course, that he is appealing, in much the same way as Locke, to a state of nature. Nonetheless, whatever the original position is meant to represent, the same

⁶⁷ Rawls at 18.

⁶⁸ *Id.* at 20.

⁶⁹ Fisk at 54.

⁷⁰ For a succinct analysis of this dilemma, see *id.* at 53-57.

problem arises, and it does not seem that Rawls can circumnavigate it by means of reflective reason. For even if it is assumed that Rawls manages to wriggle off the first horn of the contractualist dilemma and is able to extract principles from the original position, there are grounds to believe that reflective reason cannot achieve the neutrality required for justification other than by confronting the same intractable problem of human nature that social contract theory has faced all along. An inspection of the manner in which Rawls interprets the justificatory function of the original position reveals why this is so.

Rawls makes light of his ultimate dependence on human nature partly by postponing the question of what human nature actually comprises. It does not take very long to discern, however, that for Rawls the original position generates principles of justice because its conditions are also those of human nature. Thus, one is told that it is "[b]y acting from these principles [of justice flowing from the original position] that persons express their nature as free and equal rational beings. . . ." ⁷¹ Again, Rawls explains that "when we knowingly act on the principles of justice in the ordinary course of events, we deliberately assume the limitations of the original position. One reason for doing this is to give expression to one's nature."⁷² Because reflective reason is the judge of what is allowed to be introduced into the original position it, too, must be seen as no more than a conduit from the concept of human nature and can therefore have no justificatory power independent of it. This point should be well noted, for apart from anything else, it foreshadows a question that will be relevant to the discussion later; namely, whether or not the role of reflective reason is rendered largely idle by the fact that the justificatory force of the original position is said to be pre-established by the latter's identification with human nature, so that reflective reason only comes into effect in doubtful cases. At this juncture, though, we may be content to record that in the end Rawls' efforts to justify his two principles are inextricably tied to the source of morality thought to dwell in human nature.

Because the difficulties of utilizing the concept of an essential human nature as a means of justification have now been made plain, and since, moreover, we have been unable to substantiate any distinction between the humanly essential and the humanly contingent, we need not debate here Fisk's argument that freedom and equality, the two features of human nature which Rawls identifies, can actually be shown to reflect certain ideological biases. Nor is it necessary to consider the more subjective assertion that Rawls' alleged liberal biases are less desirable than those of certain alternative societies. It is sufficient to note, rather, that the peculiar relevance which Rawls' principles have to current American society⁷³ suggests

⁷¹ Rawls at 252.

⁷² *Id.* at 253.

⁷³ Unfortunately, we cannot argue this point here since a discussion of the two principles themselves is beyond the scope of this paper. However, it is certain that Rawls would want to affirm the particular and current social relevance of his principles and it has been our contention all along that a comprehensive relevance to one place and point in time can only be achieved at the expense of the universality and timelessness of justice.

that, if the original position can generate social principles at all, it can only do so by accepting the taint of socio-historical bias. Thus, it seems, if Rawls is to escape impalement on the first horn of the contractualist's dilemma, he must inevitably sacrifice himself to the second. To determine whether he does so, it will be instructive at this point to examine one respect in which there is an apparent failure to achieve neutrality in the original position.

2. *Bias in the Original Position*

An objection that may be raised against Rawls' theory relates to the claim that, given the P.O.P.s' state of ignorance, advocating principles leading to a fair social system would be the only policy that they could adopt rationally. According to Rawls, rational expression of the P.O.P.s' self-interest will lead to the selection of principles the same as "those a person would choose for the design of a society in which his enemy is to assign him his place."⁷⁴ This is Rawls' "maximin" principle. It "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 outcome of the others."⁷⁵ Thus, as was noted earlier,⁷⁶ under Rawls' formulation, the P.O.P.s would choose an egalitarian social order out of self-interest and fear of the unknown. Yet would they necessarily do so? Is there not the possibility that certain people, daunted by the prospect of a life of undistinguished anonymity amongst a race of standardized egalitarians, might choose to gamble on their futures? Might they not be willing to run the risk of suffering a lifetime of relative oppression in order to indulge the hope that power, glory, and, most of all, affluence might come their way? This idea is well illustrated by a consideration of the way in which people want things.

Notwithstanding the difficult problem of distinguishing wants of absolute and instrumental value, it can be argued that, although there exist certain items which everyone wants and would prefer to have more of rather than less, this will often only hold true so long as some people are able to have more in relation to others. Thus, although we should all probably like to win the pools, it does not follow that we should like everyone to do so. The desire for more, at least in terms of wants superfluous to the requirements of subsistence, very often springs from this sort of relativism and, in view of the ultimate futility of an eternal accumulation of goods, would be meaningless without it.⁷⁷ This idea is especially obvious in Western societies where the popular impulse is to 'keep up with the Joneses,' and better still, to 'get ahead.'⁷⁸ For this reason it is certainly possible that some people will be willing to risk receiving less in order to have the chance of acquiring more. While Rawls recognizes the theoretical possibility of this 'all or nothing'

⁷⁴ Rawls at 152.

⁷⁵ *Id.* at 153.

⁷⁶ *Supra* at p. 219.

⁷⁷ Barry at 119. Much the same point was made by Durkheim almost one hundred years ago: E. Durkheim, *Suicide et natalité, étude de statistique morale*, (1888), 26 *Revue Philosophique* 446.

⁷⁸ Barry at 119-25. Barry goes on to argue that his relativist principle is readily applicable to Rawls' primary goods of wealth and power. He also makes out a similar case for liberty which, he concedes, is less easily accommodated within his principle.

approach, he dismisses it as irrational. The question then, is whether or not the 'take no risks' policy which is adopted by the P.O.P.s and which leads to the equalizing principles of justice, should be regarded as the inevitable outcome of rational choice exercised under the conditions of the veil of ignorance.

To act rationally, it might be said, is to act in a way geared to the most efficient achievement of certain goals. What is rational for a person, therefore, is contingent upon the ends he sets himself. The nature of those ends will, in turn, depend on two factors: the person's psychological constitution and the totality of his circumstances as he perceives them. This exposition makes it apparent that Rawls is faced with a problem — due to the veil of ignorance the P.O.P.s can have no idea what their particular goals in life are. The preceding analysis suggests that all they really know is that they are motivated by self-interest. In a society of finite resources and differentiated personal abilities the normal thrust of human impulses is away from egalitarianism. In the original position the results produced by self-interest, uncomplicated by consideration of people's external circumstances, are exclusively determined by whatever form that self-interest happens to take; the form itself is determined by the P.O.P.s' psychological propensities. The principles for the government of society that it is rational for a P.O.P. to select, therefore, are a function simply of the P.O.P.'s willingness to gamble on his future want-satisfaction. So far as the original position is concerned, this means that the risk-minimizing adoption of egalitarian principles may only be designated rational if a risk-minimizing mentality is presupposed in the beginning.

Rawls certainly does not make explicit any importation into the original position of a general psychological conservatism. The purported neutrality of the determinants of entrance to the original position prevents him from doing so. Nonetheless, it is difficult to escape the conclusion that he does furnish the P.O.P.s with conservative personalities. His omission to guard against charges of this type of bias presumably stems from his taking for granted the neutralizing effect of the original position, as it allegedly conforms to human nature. The unfeasibility of a justificatory appeal to nature was made plain enough earlier. It has been suggested, moreover, that if a natural human psychology were identifiable, knowledge of human behaviour makes it very unlikely that it would assume the form Rawls identifies in the original position. In Benjamin Barber's view, "a consideration of actual historical developments and concrete institutions as they manifest special psychologies may in fact suggest that the no-risk predilection for security may be atypical of human choice in the face of uncertainty."⁷⁹ Barber then goes on to enumerate some psychological reasons for this.⁸⁰

⁷⁹ Barber at 298.

⁸⁰ *Id.* at 298-99. "The 'satisfactory minimum' afforded by the maximin rule may not be 'satisfying' at all by the criteria of maximin satisfaction. Avoiding pain or penury or powerlessness may not be measurable on the same scale as achieving (and enjoying) pleasure, wealth, or omnipotence. This possible asymmetry may in turn reflect fundamental psychological disparities between the need for security expressed in the fear of pain, in anxiety, in the longing for serenity, and perhaps even in the death drive, and the need for self-expression manifested in the quest for freedom, for spontaneity, for domination and for self-fulfilment."

Finally, he proposes that, far from ensuring a risk-minimizing strategy amongst the P.O.P.s, the original position may actually encourage a vigorous and unrestrained pursuit of self-interest because it dispenses with the constraints that some sense of morality and altruism normally imposes.⁸¹ To return to the basic point, however, it has to be concluded that, given the usual latitude of human psychologies, the carefree approach of the gambler need be no less rational for certain people than a cautious, security-conscious policy is for others.

Not surprisingly, the problem of temperamental bias can be traced back to the contract theorists permanent conundrum of deciding what may be included in the situation of making the contract and what may not. Rawls' particular dilemma may be stated very simply as follows: the principle of egalitarianism can only be arrived at in the original position by introducing into the latter a specially weighted psychology characterizing the contracting parties; but by doing this the autonomy of rational choice, upon which Rawls' whole idea of justice as fairness is based, is necessarily eroded. Several other critics of *A Theory of Justice* have called particularly into question Rawls' claim that the policy of maximin is the necessary product of reason uncolored by subjective influence. It has been argued, for example, that Rawls' exclusion from the original position of knowledge of probabilities is morally arbitrary in its design to eschew utilitarianism in favor of equality.⁸² Again, it has been suggested that the choice in the original position is irreducibly one among different types of societies with advantages and disadvantages,⁸³ while the proposition has also been put forward that Rawls opts for maximin as though fairness constituted some kind of "fixed point in our value system."⁸⁴

3. *Naturalistic Justification and General Objectivism*

Hitherto, in asking why purposive ratiocination in the original position would produce principles perceived as just, we have emphasized the difficulties Rawls encounters in attempting to afford the contract situation justificatory power, generally by way of the P.O.P.s' impartiality, and particularly, though only implicitly, by way of the conformity of their circumstances to human nature. As yet, we have not related Rawls' difficulties to the general problems always involved in making naturalist or other objectivist forms of ethical argument.

In his concluding "remarks on justification"⁸⁵ Rawls denies using the "naturalistic" method of justifying ethical propositions. Yet, he clearly means to be presenting some sort of very substantial objectivist ethical theory. His unashamed neglect of "the meaning of moral words . . . and the nature and logical properties of moral concepts"⁸⁶ thus allows him "to get on with the

⁸¹ *Id.* at 300.

⁸² Hare at 106-07 and D. Lyons, "Nature and Soundness of the Contract and Coherence Argument," in *Reading Rawls* at 162. Hereinafter referred to as Lyons.

⁸³ Barry at 123.

⁸⁴ Lyons at 159.

⁸⁵ Rawls at 578.

⁸⁶ Hare at 81.

task of developing a substantive theory of justice⁸⁷ while simultaneously exhibiting a confident disregard for noncognitivist skepticism.⁸⁸ Although Rawls' optimism in this project is refreshing, his carefree attitude towards "questions of definition and meaning"⁸⁹ often leads him to assume and speculate. How seriously these lapses undermine his anti-intuitionism and objectivist aspirations will be considered presently. Certainly, the overall nature of his endeavor in *A Theory of Justice* necessitates that he at least adduce evidence for the two principles' objective justification. It has already been demonstrated that the attempt to meet this requirement by way of the combined operation of the original position and reflective reason is in certain respects vulnerable. Whether, though, this two-way or symbiotic mode of reasoning is open to charges of intuitionism is again a matter that must be postponed for later discussion. It is sufficient at this point to note that Rawls' objectivism, at least on its face, is of a fairly diluted kind.

Rawls intends his objectivism to appear very much like impartiality or rational consensus. Ostensibly, the P.O.P.s do not apprehend the principles of justice as having any objective foundation external to their minds. Rather, the principles are reached on the basis of what is agreed to as being reasonable. They are "the principles that we would want everyone (including ourselves) to follow were we to take up together the appropriate general point of view."⁹⁰ The conditions of the original position ensure that "we do not look at the social order from our situation but take up a point of view that everyone can adopt on an equal footing. In this sense we look at society and our place in it objectively: we share a common standpoint along with others and do not make our judgements from a personal slant."⁹¹

Professedly, then, Rawls' objectivism is non-definist. The claim to have rejected both naturalist and metaphysicalist definism is presumably an attempt to preserve his avowed objectivism from the more telling criticisms of noncognitivism. Most importantly, Rawls wants his theory to avoid being interpreted as contending that *ought* can be defined in terms of *is*, and *value* in terms of *fact*. What he says instead is that his principles of justice pass the test of objectivity, insofar as an objective morality can be established, as opposed to deduced, by way of consensus achieved through the neutrality of pure reason.⁹² Unfortunately, there are immense difficulties in purging

⁸⁷ Rawls at 579.

⁸⁸ By "noncognitivist skepticism" is meant the view, contrasted with "definist" thinking in ethics, that the language of value and ethical judgements cannot be translated into factual descriptions of the world and of persons and their actions in it. While noncognitivist theories do vary in the degree to which they allow that value and ethical judgements are capable of justification, even the most moderate noncognitivism holds that reasons put forward as justifying judgements of this sort have, ultimately, to be relative to individuals or their social groups.

⁸⁹ Rawls at 579.

⁹⁰ *Id.* at 516. Rawls' emphasis on universalization reveals the Kantian tenor of his theory.

⁹¹ *Id.* at 516-17.

⁹² Again, in stark contrast to the utilitarians' moral emphasis on want-satisfaction, Rawls evinces a Kantian willingness to let reason identify the normative standards of right and wrong behaviour. For a penetrating and summary analysis of the "morality of reason" see Unger at 49-51.

reason of all subjective bias. To begin with, it is clear that different peoples and societies do make different judgements of morality and value. In order to solve this problem, those who argue for the possibility of absolute impartiality would have to prove that there are conceivable conditions under which such differences would no longer exist. These conditions would be those that represented what was left of persons' situations when certain inessential, prejudicing influences had been removed. This, of course, is exactly what Rawls sought to accomplish in the original position. Yet, as we saw earlier, his attempt to achieve neutrality in the contract situation ended in failure. We also discovered that Rawls perceived the impossibility of the task and was obliged to make compromises with definistic forms of objectivism. It was for this reason that he attempted to smuggle into the original position the bolstering effects of an appeal to the objective realm of human nature. In reality, then, Rawls' objectivism, what there is of it, seems to be of a traditional definist kind. What I wish to do now is to look briefly at the more serious objections to definist objectivism, particularly as they relate to the original position; to notice how the case against subjectivism, or relativism, appears to have partially the same 'emotional' basis as the argument, described earlier,⁹³ against determinism; and, finally, to point out how the acute vulnerability of definist objectivism eventually forces Rawls to rely, at numerous crucial stages of his theory, on precisely that doctrine, the doctrine of intuitionism, which it was one of his foremost intentions in *A Theory of Justice* to repudiate.

The Defects of Definism

Ever since G. E. Moore's *Principia Ethica*, published in 1903, sought to demonstrate the "naturalistic fallacy" entailed in any attempt to translate the language of moral or value concepts into descriptive terms, great emphasis has been placed on the irreducible prescriptiveness that is a feature of value language.⁹⁴ R. M. Hare, for example, states that "what is wrong with naturalist theories is that they leave out the prescriptive or commendatory element in value judgements, by trying to make them derivable from statements of facts."⁹⁵ What is important about this point, of course, is that the prescriptions of moral language are not verifiable in the way that descriptions are. For example, when we teach our children that 'murder is wrong,' we may be saying that murder is prohibited by the state we live in. But in the same words we shall also be saying 'do not murder' and, in addition, we shall be expressing our own disapproval of murder. Yet it is nonsensical to ask whether the sentence 'do not murder' is true or false and the same applies to the truth or falsehood of our feelings about murder. Because of this, many modern philosophers have convincingly argued that all moral judgements are unavoidably subjective. This is said to be so despite the belief of most laymen that words like 'good,' 'bad,' 'right,' and 'wrong' have much the same mean-

⁹³ See *supra* at p. 232.

⁹⁴ By value language is meant terms such as 'good,' 'right,' and 'ought,' all of which are used as value terms and all of which are characteristic of moral language.

⁹⁵ R. M. Hare, *The Language of Morals* (New York: Oxford University Press, 1964) at 82.

ing for all of us. In fact, it is this belief that I want to contend motivates the objectivist argument in the first place. For as with determinism so with subjectivism: if an apparent consensus of feelings cannot be shown to have any objective basis, then the actions that spring from those feelings can have no real legitimacy. The subjectivist thesis has, therefore, what is for most people the undesirable consequence of making the determination of moral disputes impossible. Thus, general agreement on the fact that torture is wrong will mean nothing more than that most people feel it is wrong, and this is something that we should probably not want to say, especially in a democracy where the execution of the policies of the majority is felt to be a legitimate way to govern society. On the subjectivist view there can, at best, be only a pragmatic legitimacy in ignoring or suppressing one man's wishes when most people believe that what he wants is wrong.

The problem which the objectivists discern in subjectivism, then, is that it reveals the structure of society to be completely without legitimacy. Such structure represents nothing more than the domination of certain classes and groups over others. The fact, however, that subjectivism can⁹⁶ be construed to cast society in such an unfavourable light may make it an unattractive theory, but it is no less tenable for that.⁹⁷ It is in the sense, therefore, of claiming the verifiability of moral judgements in order to avoid the undesirable consequences of subjectivism that objectivism has at least partially an emotional basis.

Thus far, in outlining the defects of definist objectivism, we have taken Moore's basic criticism and developed it into a thoroughgoing subjectivist skepticism. Apart from highlighting the problems of the definist position which we have argued Rawls covertly takes up, this approach has served to illustrate the philosophy to which Rawls' project is necessarily opposed. The theoretical shortcomings of definism, of themselves, need not have driven us this far however. We might have stopped short with the intuitionism of Moore. What I wish to show now is that ultimately this is exactly the course adopted by Rawls. Motivated initially by a desire to demonstrate the possibility and worth of substantive moral theory in the face of subjectivist skepticism, he tries first to establish the objectivity of his principles in consensus reached by way of impartial reason. When the difficulties of this course become evident, he tries to reinforce his claim to objectivity by a furtive resort to definism, even though his disavowal and the stubborn problems of definist theories prevent him from embracing this position other

⁹⁶ Commonly, of course, the more extreme subjectivists such as A. J. Ayer are able to deny the harmful potential of their views by reminding us that meta-ethical theory, as such, is not meant to be a direct guide to human conduct.

⁹⁷ A. J. Ayer, "On the Analysis of Moral Judgments" in M. Munitz, ed., *A Modern Introduction to Ethics* (Glencoe, Ill.: Free Press, 1958) at 546.

If it could be shown . . . that the general acceptance of the sort of moral judgements that I have been putting forward (that is, the doctrine of extreme subjectivism or emotivism) would have unhappy social consequences, the conclusion drawn . . . might be that the doctrine ought to be kept secret . . . [B]ut this is a question I am not now concerned to argue. What I have tried to show is not that the theory I am defending is expedient, but that it is true.

than very gingerly. Caught adrift between the inhospitable poles of defnivist objectivism on the one hand, and subjectivist skepticism on the other, Rawls takes compromising refuge in the haven of intuitionism. Yet he does so, as we know, in direct contradiction of his express intention.

5. *Rawls, Intuitionism and Subjectivism*

Rawls' main criticism of intuitionism is that it is inferior to rational modes of ordering moral principles. He complains, first, that intuitionist theories comprise a number of primary principles that are often in conflict in their practical application and, second, that they include no scheme of priority rules for determining such conflicts. Since Rawls regards a dependence on intuition to resolve these issues as something less than adequate, he is obliged in presenting his own theory to establish standards that enable our conflicting views of morality to be compared and ordered objectively. With this in mind he explains:

[t]he only way . . . to dispute intuitionism is to set forth the recognizably ethical criteria that account for the weights which, in our considered judgments, we think appropriate to give to the plurality of principles. A refutation of intuitionism consists in presenting the sort of constructive criteria that are said not to exist.⁹⁸

In qualification of this assertion and presumably to forestall premature criticism, Rawls admits not only that intuitionism cannot actually be refuted,⁹⁹ but also that "any ethical view is bound to rely on intuition to some degree at many points."¹⁰⁰ What he does not admit is that such points in his theory are often of crucial significance.

What seems unreasonable about Rawls' appeal to intuitionism is that most of the places at which he relies on it are those where his theory purports to diverge from conventional treatments of justice. To illustrate, it is obvious that one's common-sense conceptions of justice usually entail vague notions of a 'fairly' equal distribution of the things that people want and probably also the idea that, when there is a discrepancy in the expectations of different social groups, that discrepancy should not continue to grow at the expense of the least advantaged group. The novel aspect of Rawls' theory lies in his attempt to clarify and systematize these woolly ideas. Rawls is quite candid about this. In the preface to his book, he modestly disclaims "any originality for the views . . . put forward." He admits that "[t]he leading ideas are classical and well known" and that his "idea has been to organize them into a general framework by using certain simplifying devices so that their full force can be appreciated."¹⁰¹ Unfortunately, it is in his attempt to "organize" the classical ideas of justice that Rawls ultimately falls back upon intuitionism. This fact is apparent in his utilization of the important concepts of "primary social goods" and "the least advantaged member."

⁹⁸ Rawls at 39.

⁹⁹ *Id.*

¹⁰⁰ *Id.* at 40.

¹⁰¹ *Id.* at viii.

Rawls' definition of a person's good¹⁰² is ultimately reducible to the statement that something is good for somebody with a rational plan of life if it can be shown to have the properties which it is rational for him to want. This definition, if it is not simply tautologous, suffers in that, by equating man's good with the satisfaction of his wants, it renders the good objectively indeterminate. Identification of someone's good, therefore, eventually requires a resort to intuitionism. Thus, in asking what combination of primary social goods it would be rational for society's representative least advantaged member to prefer, Rawls concedes that here we must "admittedly rely on our intuitive capacities."¹⁰³

Rawls has a similar but more serious difficulty with his concept of the representative least advantaged member of society.¹⁰⁴ The concept's single function is to serve the operation of the "difference principle," the critical component of Rawls' second principle of justice. Subordinate only to the universal primacy of liberty, the second principle of justice stipulates, in its 'bare bones,' that all inequalities in liberties must be justified only as being an advantage to the least advantaged. Thus, it aims at mitigating the morally arbitrary inequalities of natural distribution. In doing this, it employs the difference principle, a device which, according to Rawls, removes the indeterminateness of natural distribution "by singling out a particular position from which the social and economic inequalities of the basic structure are to be judged."¹⁰⁵ This is the position of the least advantaged member. In this way the difference principle insists that, given the framework of social institutions required by justice as fairness, the greater benefits enjoyed by those better situated are just, only provided that they are necessary to the improvement of "the expectations of the least advantaged members of society."¹⁰⁶ In other words, the principle can never be satisfied by any degree of improvement in one man's situation unless there is a concomitant improvement in the situation of the least advantaged member.

Analysis of the basic idea of the second principle of justice thus underlines the importance of the concept of the least advantaged member and the implications of any difficulties it entails. A major problem is the determination of who, in fact, the least advantaged member is. Rawls proposes two possible models,¹⁰⁷ but is obliged to concede that any definition is bound to be largely arbitrary. Again, then, we must rely on our intuitions. Somewhat

¹⁰² Unfortunately, a full discussion of Rawls' concept of "goodness as rationality" is beyond the scope of this paper. Here it should be noted only that our summarized interpretation is based on the application of Rawls' "thin theory of the good" (at 399-404) to persons' plans of life (at 407-16).

Rawls himself says that "a person's good is determined by what is for him the most rational . . . plan of life . . . [T]he good is the satisfaction of rational desire." Rawls at 92-93.

¹⁰³ Rawls at 94.

¹⁰⁴ *Id.* at 97-98.

¹⁰⁵ *Id.* at 75.

¹⁰⁶ *Id.*

¹⁰⁷ *Id.* at 98.

unreasonably, however, Rawls implies that this does not really matter.¹⁰⁸ On the contrary, one would have thought an objective definition was essential to the validity of the difference principle. The least advantaged member is, after all, the standard against which all social inequalities are to be measured. The failure to identify him objectively will mean that whatever the definition finally adopted, it will inevitably contain a subjective bias to be reflected in the particular distribution of goods which justice as fairness is said to demand. The overall significance of identifying the least advantaged member intuitively may be gauged by understanding the conceptual importance of the difference principle which depends upon it. Thus, it should be remembered that it is the difference principle that sets Rawls' theory of justice distinctly apart from utilitarian conceptions of that virtue.¹⁰⁹ It was made plain in the preface that this was one of Rawls' chief aims, and if there are still any doubts about how critically he regards the role of the difference principle in achieving it and in the establishment of his theory generally, reference need only be made to his apocalyptic comment that the general conception of justice¹¹⁰ "is simply the difference principle applied to all primary goods including liberty and opportunity."¹¹¹

Rawls' project to systematize common sense notions of justice may be undermined by other problems than the use made of intuitive judgements in calculating a person's good and in operating the difference principle. Joel Feinburg has argued that intuitionism is also admitted into the substantive theory in the ordering of principles. This is said to be especially evident in Rawls' treatment of civil disobedience.¹¹² Thus, for an individual to know when he is justified in engaging in a "public, non-violent, conscientious act contrary to law"¹¹³ under conditions where there is a presumption in favour of obedience, requires criteria above and beyond those for determining the justice of social institutions. Nonetheless, the same contract method is employed to provide guidelines for individual conduct, and subsumed under these there emerges a collection of "natural duties" calculated to be in overall furtherance of justice as fairness.¹¹⁴ Because of the diversity of these duties, however, complications and conflicts can arise in the determination of specific issues and this seems to be the case with civil disobedience.

For Rawls, justifying civil disobedience is a matter of individual discretion¹¹⁵ and "the difficulty is one of a conflict of [natural] duties."¹¹⁶ Since

¹⁰⁸ *Id.*

¹⁰⁹ Lyons at 142-43.

¹¹⁰ For the purposes of this inquiry the distinction between the "special" and "general" conceptions of justice is of no particular significance. The distinction relates simply to the determination of when the first principle of justice is to have an inviolable priority over the otherwise universal applicability of the second principle.

¹¹¹ Rawls at 83.

¹¹² Joel Feinburg, "Rawls and Intuitionism," in *Reading Rawls* at 118. Hereinafter referred to as Feinburg.

¹¹³ Rawls at 364.

¹¹⁴ *Id.* at 114-17.

¹¹⁵ See Feinburg at 120.

¹¹⁶ Rawls at 363.

“precise principles that straightway decide actual cases are clearly out of the question,”¹¹⁷ resolving the conflict of duties has, in the last analysis, to be an intuitive process. Finally, it should be noted that the problem which Feinburg draws attention to is not simply that of the individual lacking sufficient competence or information to be able to assess what would be a just course of action in any given complex of circumstances. The difficulty, rather, is that in a society which, institutionally, is only imperfectly just,¹¹⁸ there will be occasions when an adherence to laws designed to further justice will produce unjust results. On such occasions, the decision as to whether to do one’s duty by obeying the law, or to accede to the demands of more particular justice, will be a matter for the individual’s own conscience or intuitions.

Potentially far more ruinous of Rawls’ theory than any of the preceding charges of intuitionism is the possibility that the social contract method turned to account by Rawls is itself radically dependent upon intuitive judgments. Certainly, there is much on the surface of the theory to support this interpretation. As noted earlier, the very idea of moral principles being generated by popular agreement suggests at least a departure from the ‘strong’ objectivity one would think Rawls is striving for. Furthermore, as R. M. Hare has indicated, the entire book is shot through with the speculative language of intuitionism.¹¹⁹ If the original position facilitates nothing more than the clarification, ordering, and general systematization of what we optimally think reasonable, may Rawls’ method be truly proclaimed fundamentally non-intuitive? In asserting the intuitive character of principles proceeding from consensus, Hare states quite unequivocally that it may not.¹²⁰ What seems unreasonable in Rawls’ approach, he suggests, is not the contention that a theory must be acceptable to its proponent, but the implicit claim that the popular acceptability of a theory is somehow commensurate with its truth content. Rawls, of course, denies that his particular contract method is basically intuitive. Yet the recurring resort to such phrases as, “what it seems reasonable to suppose,” “I assume there is a broad measure of agreement that,” “commonly shared presumptions,” “we are confident,” and more besides, reinforces the suspicion that the true source of the principles of justice is Rawls’ intuitions. Is there anything in *A Theory of Justice* to divert us from this conclusion?

A feature of Rawls’ ‘contractarianism’ that might salvage the credibility of its claim to objectivity is the use made of the ‘coherence’ approach to moral theory. The ‘coherence’ argument used by Rawls is that the principles of justice are not simply arrived at by collective agreement. They are established, rather, by a working harmonization of what is initially agreed to in the

¹¹⁷ *Id.* at 364.

¹¹⁸ Rawls’ book is basically a composition in “ideal theory.” As Feinburg says, Rawls “presumes that his original choosers are to select principles that will regulate a ‘well-ordered society,’ that is, a society in which everyone acts justly, all laws are just, and all citizens always comply with them.” (Feinburg at 116-17). When, as will often be the case, our actual institutions diverge from the ideal conception, then at least we have indicators for policies of amelioration.

¹¹⁹ Hare at 84.

¹²⁰ *Id.* at 83.

original position and what, retrospectively, the considered judgements of ordinary people suggest is reasonable. When these two perspectives produce an acceptable focus then the ideal state of "reflective equilibrium" has been reached.

According to Ronald Dworkin, the worse excesses of intuitionism that Hare attributes to Rawls can only be derived from a misinterpretation of the particular coherence argument that Rawls makes.¹²¹ Dworkin describes two models of 'coherence' theory and suggests that the 'natural' model which assumes an "objective moral reality"¹²² is the one which Hare wrongly infers in his 'assault' on Rawls. Alleged to be more consistent with Rawls' method is the "constructive" model which "treats intuitions of justice not as due to the existence of independent principles, but rather, as stipulated features of a general theory to be constructed."¹²³ In this way, the constructive model assumes the general body of our sincerely held moral convictions, and then works to constrain, within the bounds of justice, the acts which these intuitive convictions "might be said to warrant."¹²⁴ What Dworkin really seems to be saying, then, is that Rawls' 'coherence' theory gives scope to persons' intuitions but only insofar as they do not violate popular interests.

But even if we accept Dworkin's interpretation, and it does seem consistent with our earlier comments on the concept of "reflective equilibrium," how far does it take us in defending Rawls against accusations of a basic reliance on intuitionism? Why should the procedure of working "from both ends"¹²⁵ produce objective results when each end represents nothing more than our several common sense opinions on the one hand, and the compromising agreements we have worked out amongst ourselves on the other? In this respect, coherence arguments appear incurably circular. As one critic has put it, they seem to move us back and forth "between our current attitudes and the principles they supposedly manifest."¹²⁶ Because the latter can be 'checked' against the 'facts' of our impartially-arrived-at and considered judgements they may be said to mirror our fundamental moral beliefs; but even if they do, may they then be regarded as resting on anything more than arbitrary and transient values which by chance are held in common at one particular moment? For this sort of reasoning to amount to a justification of moral principles would seem to require an assumption of "either a complacent moral conventionalism or else a mysterious 'intuitionism' about basic moral 'data'."¹²⁷ Dworkin has pointed out that Rawls has frequently been thought 'guilty' of the former attitude. If he is, he seeks to justify his "conventionalism" by locating it in the "basic moral data" which Lyons perceives as his only alternative. Since, as demonstrated earlier, Rawls' attempt to

¹²¹ Dworkin at 27-34.

¹²² *Id.* at 27.

¹²³ *Id.* at 28.

¹²⁴ *Id.* at 31.

¹²⁵ Rawls at 20.

¹²⁶ Lyons at 146. For my appraisal of Rawls' use of coherence arguments I am indebted to Lyons' analysis.

¹²⁷ *Id.* at 147.

objectify this 'data' by identifying it with an essential human nature is doomed to failure, it can only be concluded that his apprehension of a common morality among men is intuitive in origin. If this is so, and Rawls is really driven back to the position of the intuitionists, what are the implications of this for the justificatory role of the original position?

In the first place, of course, such a retreat would mean that the original position could have no more justificatory power than intuitionist theories are ever capable of conferring. Now while intuitionists such as Moore have claimed that intuitive judgements can be justified by reference to a realm of self-evident, non-natural properties in respect of which ethical statements are said to be made, such a claim is, in fact, very hard to defend rationally. This is particularly so in view of the indefinable nature of the properties which are supposed to be objectively intuited. The more reasonable inference to be drawn from theories like Moore's is that intuitionism is ultimately reducible to subjectivism. At any rate, Rawls is certainly aware of the difficulties that beset intuitionism and it is surely such an awareness that prompts him to seek ostensible justification of his principles elsewhere. That he makes this endeavour and that he fails in it, is, I believe, a sufficient final comment on the capacity of the original position to justify the social principles contractually agreed to.

D. CONCLUSION

In concluding I do not intend to rehearse or summarize the preceding arguments in criticism of the original position. It should be clear that the great weight of evidence suggests that Rawls' contractual fiction works in neither its analytic nor its justificatory capacity. There remains only the task of assessing the broader implications of Rawls' ultimate failure, particularly his inability to justify the alleged products of the original position. First, one must infer the fruitlessness of Rawls' attempt to demonstrate the possibility of objective moral theory. Indeed, his abortive endeavours in this regard serve merely to reaffirm the moral theorist's ineliminable predicament, a predicament, moreover, that the skeptics have long since resigned themselves to. Yet, if the skeptical movement in ethics, particularly in its more extreme relativistic forms, has been largely negative in emphasis, perhaps, in one sense, the same may be said of Rawls. For if, in our efforts to avoid the practical consequences of a total anarchy of values, we seek to impose some sort of moral order on society internally, then it is essential that we be quite conscious of the nature of the act we are perpetrating. If the final lesson to be learned from *A Theory of Justice* is that the 'fairest' way of establishing a moral order for society is to appeal to a consensus approximating as nearly as possible to an impartial popular rationality,¹²⁸ then, in order that such a method be shielded from the abuses of an unanswerable claim to legitimacy by the conservative holders of power, the 'representatives of consensus,' it is

¹²⁸ This conclusion makes clear the appropriateness of Rawls' contract method in revealing the nature of morality to reside in the rules by which people, under special conditions, can agree to be mutually bound.

imperative that the moral order be seen for what it really is, namely, an expression of pragmatism, a social expedient. The detriment of Rawls' theory is that by purporting to confer an impossible legitimacy on a chosen moral order, it threatens to subvert this sort of 'realism.' It acts, in other words, to perpetuate the fiction of the possibility of an objective morality. Whether or not Rawls compounds this fault, as many have suggested he does, by objectifying the principles of justice in the idealized condition of contemporary American society, is a question that need not be addressed here. At this, the terminus of our inquiry, we need only conclude that the most appropriate response, both to Rawls' employment of the artifice of the original position and the false kind of ethical theory that it generates, is that, surely, "the season of fiction is over."¹²⁹

¹²⁹ *Supra*, note 1.

