November 21, 2002: Bells tolled, tears flowed, and mourners mourned as they are wont to do. The philosophy community had lost one of its own. John Rawles, noted author of *A Theory of Justice*, died at home of heart failure. According to the Harvard Gazette, Harvard University President Lawrence H. Summers stated,

> I am deeply saddened by the death of John Rawles. He combined profound wisdom with equally profound humanity. Few if any modern philosophers have had as decisive an impact on how we think about justice. Scholars in many different fields will continue to learn from him for generations to come.

What kind of man elicits such sentiments from the President of Harvard University? What impact did this combination of “profound wisdom with equally profound humanity” have on society? Why was this combination so unique and exceptional in the discipline of philosophy? To discover the answers to these questions, we must look at the word in which John Rawles lived and examine how his view of humanity shaped his belief in justice.

### 0.1 Utilitarianism

The dominant philosophical paradigm during much of Rawles’s life was utilitarianism, *à la* Bentham and Mills. In utilitarianism, actions are evaluated in terms of how they affect the total amount of ‘utility’ in the world, utility being
the "sum of pleasures over pains." This philosophy is often reduced to the adage, 'the greatest good for the greatest number.' This adage finds itself in many places. L. Ron Hubbard's Scientology bases its ethics on the greatest good for the greatest number. Many arguments in favor of globalization utilize it. It is even used in science fiction movies. Yet, such a view allows many things most would find repulsive. For instance, the repression of a minority would be acceptable as long as the greater number received a greater good. In other words, slavery would be acceptable if a small number of persons were enslaved and the remainder of society gained through cheaper goods and services. Infanticide of a disabled child would be allowed as long as another child were conceived to take its place. Rawles did not approve of these results. Just as Locke's philosophy was written in direct opposition to that of Thomas Hobbes, John Rawles found his inspiration for his theory of justice from the distasteful conclusions of utilitarianism.

To fight against utilitarianism, Rawles returned to an earlier age, an earlier version of philosophy. Rawles reintroduced the social contract to the world, essentially reincarnating Hobbes, Locke, and contractarianism. To place his philosophy on terra firma, Rawles needed to return to the beginning, to an environment without government, to a state of nature.

0.2 The State of Nature

While the state of nature is most associated with Thomas Hobbes, it is a generic term for man before the institution of government. Different contractarians had different views of this pre-societal environment. In each case, the milieu in which the philosopher lived shaped his or her view on this state of nature; however, two archetypes of this state of nature persevere. The first was created by Thomas Hobbes, the second, by John Locke.

Hobbes, experiencing the English Civil War of the mid-seventeenth century, saw the state of nature as being "a war of every man against every man" in a life that was "solitary, nasty, poore, brutish, and short." A little over three centuries later, another Englishman fleshed-out this vision in his celebrated novel, The Lord of the Flies. In this book, Sir William Golding tells a tale of...
a group of young men who are deposited in the state of nature (i.e. marooned on a tropical island with no adult guidance). The story describes the descent of man to a war of every man against every man, to a battle for survival, and to a life that is definitely nasty, brutish, and far too short.

In direct contrast to Hobbes, John Locke, a product of the ‘Glorious Revolution,’ posited a much more benign state of nature. It was still dangerous and liable to reduce to a state of war at any moment, but it was definitely more collegial than that of Hobbes.\(^9\) In the Lockeian state of nature, the inhabitants give up their liberties to each other to become more secure in their lives. Approximately three centuries after Locke’s writings, an Englishman told a story set in a Lockeian state of nature. Robert M. Ballantyne wrote the book, *The Coral Island*, in which three young men found themselves deposited in the state of nature (i.e. marooned on a tropical island with no adult guidance). In this case, however, the three did not descend into a Hobbesian state of nature. Life was not brutish, nasty, or short. There was no war of every man against every man. It was a life of the three working together to create a more stable, more protected, environment. It was, in short, the quintessential Lockeian view of man without government.\(^{10}\)

1 The Original Position

To bring back contractarianism, Rawls again explores the state of nature and how it affects man. However, instead of using the term ‘state of nature,’ Rawls uses ‘original position.’ As with many other philosophers, Rawls did not believe that man ever actually existed in this state. It was, for him, a useful hypothetical situation in which the rules for society could be chosen. There are two features that make this state of nature unique: the characteristics of its denizens and the veil of ignorance. Before delving into these two features, I must mention the relevant background in which Rawls found himself during his life.

1.1 A Short, Relevant Biography

John Rawls was born in Baltimore, MD in 1921. In Kent, Connecticut, he attended a private preparatory school. Later he attended Princeton University, earning a B.A. degree in 1943 and a Ph.D. degree in 1950. He spent his entire academic life at such luminous universities as Princeton, M.I.T., Cornell, and Harvard, eventually earning the position of Conant University Professor at Harvard University in 1979. He held that position until his death in 2002.

\(^9\)John Locke, *Two Treatises on Government*, Chapters 2 and 3.

\(^{10}\)Some, this author included, would argue that Sherwood Schwartz created a Lockeian system for the television-viewing audience in the late 1960s in America. How else can one explain Gilligan’s Island and all of Ginger’s dresses?
Why is this important? The environment in which a person finds himself or herself will affect how he or she sees the world. Hobbes experienced the English Civil War and had a decidedly negative view on human nature. Locke lived through the ‘Glorious Revolution’ and had a positive view on human nature. John Rawles spent his entire life in and around Ivy League universities. His family was extremely wealthy, and Rawles was an obvious supporter of both the New Deal and the Great Society. As such, one can expect Rawles’s state of nature to be an intellectual exercise with its inhabitants coexisting peacefully. In actuality, this is exactly what we find.

1.2 Reflective Equilibrium

If the purpose of the original position is to serve as the setting for the creation of a set of principles of justice, then the mode of creation is a dialectic method. In his first published article, Rawles lays out a method for reasoning in ethics, which attempts to

Find reasonable principles which, when we are given a proposed line of conduct and the situation in which it is to be carried out and the relevant interests which it affects, will enable us to determine whether or not we ought to carry it out and hold it to be just and right.\textsuperscript{11}

In this method, one starts by examining the logical consequences of a community-supported concept of justice. These implications are then compared to concrete and specific cases in the community. If there is any discrepancy between the two, then either the concept of justice or the interpretation of the concrete examples is modified. The process is then repeated until the divergence is eradicated as much as possible. Once this point is reached, the concept has reached what Rawles terms reflective equilibrium: reflective, because the process requires the parties to examine in depth their views on justice, equilibrium, because both sides are in balance, at least temporarily.\textsuperscript{12} This may seem to be a method of dispassionate people. However, the original position requires passion.

1.3 Absolute Egoism and Rational Actions

The workers of the dialectic method described above have two features. First, they are absolute egoists.\textsuperscript{13} Second, they are rational in the economic sense of...
These two together ensure that whatever principles decided upon will be best for those involved in the creation process. An egoist is someone who makes decisions based on what is best for them, and absolute egoist does so with absolutely no regard for others. In the economic sense of the word, a rational actor is someone who performs efficient actions in the achievement of his or her goals. However, Rawles goes further than this.

Instead of merely stating that the actors perform efficient operations in trying to reach their goal, Rawles states their logic of action will be governed by the maximin principle. This principle states the actor will choose the action that will maximize the minimum outcome. The following example and Table 1 will illustrate this point. Let us say that you and an accomplice (aide) are arrested for breaking and entering. The district attorney has four options with which to charge you: Felony A carries a penalty of 10 years in prison; Felony B, 8 years; Felony C, 5 years; and Felony D, 2 years. The district attorney separates you and your accomplice and explains to you the following scenario: as it stands, if neither of you confesses, you will both serve 5 years in prison. If you both confess, you will both serve 8 years in prison. However, if you talk and your fellow conspirator does not, then you only get 2 years, while he gets 10 years in prison. Of course, if the opposite happens, he gets 2 years, and you get 10 years. What should you do? This is the classic ‘Prisoner's Dilemma’ from game theory (see Table 1). The solution depends upon whether you use a conservative (either maximin or minimax) or a chancy and speculative (either maximax or minimin) procedure. Rawles conjectured that those in the original position would use the more conservative of the two methods, specifically the maximin.\(^\text{15}\) As a result, the worst possible outcome will always be avoided.

### 1.4 The Veil of Ignorance

The original position is a theoretical situation in which people choose principles of justice to govern society. These people are absolute egoists, and they are rational in the economic sense. More to the point, the rational decision-making method they utilize is the maximin method. What keeps those in the original position from exploiting these features and turning the world into an oligarchy? It is the third and most interesting element of the original position: the veil of ignorance. No one in the original position knows anything about themselves or about others. They are completely ignorant about their position in relation

\[\text{Table 1: The Prisoner's Dilemma}\]

<table>
<thead>
<tr>
<th></th>
<th>Fink</th>
<th>Quiet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fink</td>
<td>8, 8</td>
<td>2, 10</td>
</tr>
<tr>
<td>Quiet</td>
<td>10, 2</td>
<td>5, 5</td>
</tr>
</tbody>
</table>

\(^{14}\text{Rawles, A Theory of Justice, §26.}\)

\(^{15}\text{Rawles, A Theory of Justice, 130.}\)
Rawles introduced this requirement for three reasons. First, it removes from consideration all morally irrelevant features of the actors. Second, it reintroduces the *ideal spectator* framework. Third, it essentially reduces the number of actors to one.

The first will go without explanation. The others will not. The ideal spectator is a device used to ensure that the decisions are made from an “impartial perspective,”17 that those making the decision do not make the decisions to specifically benefit themselves. This may seem to contradict the environment of the original position, but the third requirement demonstrates this is not quite the case. If everyone in the original position knows nothing about his or her abilities, either absolutely or relatively, then everyone becomes the same generic person. Make this person an egoist, and you will create someone who will look out for themselves. Add to the mix a maximin scheme for creating the principles of justice, and those principles will protect the lowest amongst us.

This is exactly what Rawles wanted. He did not accept the utilitarian view of the ‘greatest good for the greatest number’ simply because it allowed people to lose all rights in the name of the ‘greatest good.’ It allowed for killing and slavery as long as the amount of utility in the world was increased. These consequences of utilitarianism were anathema to Rawles. To create a theory of justice that required a ‘safety net’ for the worst-off among us, he created the original position and its three characteristics: egoist actors, rational maximin decision-making processes, and a veil of ignorance to level it all. To repeat, the three features of the original position guarantee the safety net Rawles so desperately desired in a just society.

### 1.5 Justice as Fairness

Unwritten in all this, but a definite motive force, is the idea of justice as fairness. In his 1957 article, Rawles clearly states,

> The fundamental idea in the concept of justice is that of fairness. It is this aspect of justice for which utilitarianism, in its classical form, is unable to account, but which is represented, even if misleadingly so, in the idea of the social contract.18

Fairness is ephemeral. No universal definition for fairness exists. However, the concept of fair is well understood. It bespeaks of equality for all people. For Rawles, this equality is an equality of opportunity and of treatment before the law. It is not an absolute equality of all mankind. For Rawles, the members

---

in the original position would choose this sort of equality as a result of their imposed ignorance. This equality would be enforced in the two principles of justice.

## 2 The Two Principles of Justice

What principles of justice would result from the original position? What would they entail? From that hypothetical original position, Rawls established two principles of justice. In accord with his dialectic principle, there is an early version and a later version of each. I will take each of the two principles, provide the logic leading up to them, state them, and explain their effects.

### 2.1 The First Principle: Equal Basic Liberties

The egoist seeks what is best for himself. Using the maximin schema, a certain guaranteed minimum, a protected level is created for these interests. Rex Martin states that every individual has two fundamental interests: "an interest in being able to formulate and live according to some particular conception of the good and an interest in exercising one’s ‘sense of justice’ and being motivated by it."

That is, each individual has an interest in living in terms of his or her own conception of right and wrong. Quite clearly, Rawls was speaking from an individualist, liberal democratic standpoint.

And so, he states the first principle of justice as,

\[
\text{Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all.}
\]

What this means is that, as long as there is no conflict between liberties, everyone is to have the greatest number possible, as guaranteed by the constitution. The key word in this principle is *compatible*. As long as freedoms do not interfere with one another, there are no complications. However, once there is a conflict, there must be some method to decide between the conflicting liberties. This is where the second principle of liberty enters. This is a result of these two principles being lexicographically ordered. In a lexicographical (or lexical) ordering, the first principle must be completely satisfied before the second is used.

---


2.2 The Second Principle: Distributive Economic Justice

In the original position, Rawles required that the people were completely ignorant of their abilities, both absolutely and relatively. The real world is not like this. Rawles recognized that there is a vast array of abilities and ability levels in society. These *de facto* inequalities must be taken into consideration in a full theory of justice. The rules fashioned in the original position create a protected minimum in society. Thus, the first formulation of the second principle is,

Social and economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone’s advantage, and (b) attached to positions and offices open to all.\(^{22}\)

Further reflection by Rawles led him to change the first part. Specifically, he gave more import to the safety net created in the original position. Instead of merely requiring that the inequalities be to everyone’s advantage, he required that they grant the greatest benefit to the least advantaged. This redesign always gives the least advantaged an increased benefit. Not only are they to gain from the inequalities, they are to gain a greater share. This greater share reduces overall inequalities and is in accord with what those in the original position would desire. Remember that those in the original position are behind the veil of ignorance, and as such have no knowledge of their position in society. They would create principle of justice that would benefit the least amongst them, as they would realize each of them could be that least member.

John Rawles rewrote the second principle as,

Social and economic inequalities are to be arranged so that they are both: (a) to the greatest benefit of the least advantaged, consistent with the just savings principle, and (b) attached to offices and positions open to all under conditions of fair equal opportunity.\(^{23}\)

It is interesting to note here that the two parts are reversed in *Political Liberalism*.\(^{24}\) What does this mean? In *A Theory of Justice*, Rawles states these two principles and their parts are lexically ordered, thus part a must be satisfied before part b is used.\(^{25}\) In *Political Liberalism*, Rawles only states that the two principles are lexically ordered, not their parts. Thus, in the intervening years, Rawles may have decided that the two parts of the second principle were equal in value.

What are the effects of this principle? The consequences of the first part, otherwise known as the *difference principle*, were already discussed. The effects

---

\(^{22}\)Ibid., 53.

\(^{23}\)Rawles, *A Theory of Justice*, 266.


of the second part may be minor in the US. It states that the acceptable requirements for attainment of office are only those that are morally relevant. To see why this is true, let us return to the veil of ignorance. In removing all knowledge of a person’s abilities, the only rules created are those based upon relevant features, that is, the concept of justice represented in this veil of ignorance is that all distributions should be based solely on morally pertinent issues. Thus such factors as family history, wealth, and beauty are ignored when deciding which person should hold which office. The only factors considered are those factors affecting office performance, like intelligence, ability, etc.

Finally, one interesting consequence of this principle is taxation. The idea of redistributing wealth is quite evident in his writings. Rawls believes that it is perfectly just to tax the rich to the point where they begin to reduce their investment in society. At that point, since the poor suffer, the taxation becomes unjust. As this cusp point varies throughout history, it is up to the legislature to determine the just rates of taxation.

3 The Lexical Results

This lexical ordering of priorities is very important to Rawls. In A Theory of Justice, he clearly specifies this ordering. Why? Is it merely to clarify the order of rule application? Or is it something deeper? Rawls himself affirms it is something deeper. The stated lexical ordering exhibits a deeper hierarchy in his theory, viz., there is a definite ordering of priorities in his theory of justice. The first priority is liberty; the second, justice.

3.1 The Priority of Liberty

The first principle stated “each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all.” As this is the primary principle, the only thing allowed to restrict a liberty is a conflicting liberty. Thus Rawls asserts the primacy of liberty over all other concepts. He states this priority as,

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

---

28 Ibid., 266.
29 Ibid.
The first part details the requirements for a narrower liberty to supersede a broader liberty. It is allowed, but only if it strengthens the current system of liberties. The result of this requirement is that overall liberty is enhanced. The second part echoes the ultimate motivations of those in the original position: the least can never be made more disadvantaged, for it would go against their interests as egoists. As a consequence, the only way a lesser liberty can be substituted for a broader liberty is if the least member of society allows it.\textsuperscript{30}

\section*{3.2 The Priority of Justice over Efficiency and Welfare}

The second principle of justice also gives hint to an underlying hierarchy for Rawls. While the first principle establishes a type of equality, the second principle deals with inequalities and lays out the requirements for those inequalities to be just. As it is the second principle, it holds primacy over all other principles except for the first. Thus, those principles espoused by the utilitarians are subordinate to it. Here, Rawls acknowledges that utilitarianism has some valuable points to make. Here also, he resolves the fundamental problem he had with utilitarianism, the subjugation of “one’s individual interests to those of others, theoretically without limit.”\textsuperscript{31} Rawls states the priority of justice over both efficiency and welfare (the latter two a part of utilitarianism).

The second principle of justice is lexically prior to the principle of efficiency and to that of maximizing the sum of advantages; and fair opportunity is prior to the difference principle. There are two cases: (a) an inequality of opportunity must enhance the opportunities of those with the lesser opportunity; (b) an excessive rate of saving must on balance mitigate the burden of those bearing this hardship.\textsuperscript{32}

The first part again echoes the motives of those in the original position: the principles must, above all, protect the least advantaged amongst us. The second part merely states that those doing the preserving should obtain a greater benefit

\textsuperscript{30}At this point, I should introduce Axiomatic Choice Theory. According to Plott, there are two relevant consequences of the theory. The first is the existence of a ‘dictator.’ While this person may not know he or she is a dictator, the decisions of this person are always carried out. This makes perfect sense in relation to the veil of ignorance: all decisions are made from a conservative viewpoint. Thus, the members in the original position assume the worst about their place in society and make rules to benefit themselves. The second consequence is the fact that the person who is always the worst off will always be the person who is worst off. In other words, the poorest will always be the poorest, even if their lot in life improves they will remain the poorest of the group.


\textsuperscript{32}Rawls, \textit{A Theory of Justice}, 266-267.
than the cost they bear, that is, they should receive a higher utility. The *just savings principle* does not exclusively refer to cash. According to Rawles, it refers to preserving those things in a society necessary to maintain utility for future generations.\(^{33}\) As everything in the theory of justice has based itself on justice for those currently alive, the just savings principle refers to justice *between* generations, for those yet to be born. So, the second part of this priority statement declares those who do help preserve for future generations should fairly gain from their efforts.

### 3.3 A Theory of Justice

This is Rawles’s theory of justice as conceived and written in *A Theory of Justice*. It is interesting to note here that he did not call the book *The Theory of Justice*. The use of the indefinite article hints that Rawles did not think his theory the only acceptable one, that others could be found in the future. O happy dagger! This decision both encouraged critics, as Rawles did not think the theory finished, and saved his theory, as he never declared it final. Critics of Rawles pored over the book and his writings. Rawles spent the next score years both defending and revising his work. His *Political Liberalism* served as a further explication of his theory of justice.

### 4 Further Work on Justice

Besides providing minor clarification and a more concise treatment, Rawles introduced two further points in *Political Liberalism*: overlapping consensus and the law of peoples. Overlapping consensus was briefly mentioned in *A Theory of Justice*, but it was not fully explained. In *Political Liberalism*, Rawles fully explores it in Lecture IV. The *Law of Peoples* was not mentioned at all in *A Theory of Justice*, and it was briefly mentioned in *Political Liberalism*. Instead, it received full treatment in an article in *Critical Inquiry* and in a book of the same name.\(^{34}\) The next two sections are dedicated to these two important topics.

#### 4.1 Overlapping Consensus

Throughout much of his theory, Rawles seems to disregard the results when a society does not fully agree on the basics of justice. He even goes so far as to assume that in “a nearly just society there is a public acceptance of the same principles of justice.”\(^{35}\) The reality is much different. Rawles really only requires

---

\(^{33}\)Ibid., 252.


a “consensus of reasonable comprehensive doctrines.”\footnote{Rawles, \textit{Political Liberalism}, 144.} What this means is that total agreement is not necessary, only an overlap between them, some principles in common. Even if the concepts of justice are not common, all is not lost, for it suffices if the judgments of the people are common.

This is a political consensus, not a metaphysical one. Members of a just society should be allowed to believe what they wish; that is a facet of justice. However, upon entering the public realm, in other words enter the political sphere, those members must adhere to the two principles of justice. Those who do not follow the consensus are guilty of tearing down the consensus of that just system. As a result, Rawles would advocate their removal from the public realm.

\section*{4.2 The Law of the Peoples}

His \textit{A Theory of Justice} exclusively examines people within a closed society. The people enter the society through birth and leave it through death. There is no other way to enter or leave. As such, the only participants are those within the society itself, and the only interactions are between members of that society. In his article, \textit{The Law of the Peoples}, Rawles seeks to extend his theory of justice to open societies and to interactions between societies.

In this work, he returns to the original position and applies it to the next level in the world. No longer do the parties in the original position merely represent members of society, they represent societies “well ordered by some liberal view meeting certain conditions (justice as fairness is an example).”\footnote{Rawles, “The Law of the Peoples,” 41.} The result is a theory of justice, a theory of interactions that applies across nations, not all types of nations, for certain ones are not compatible with the requirements, but across many of them. Those that are not allowed to enter this society of nations are the theocratic and radical governments.

\section*{5 Conclusion}

What is the central result of Rawles’s theory? Perhaps the best place to find it is to look, not in either of his two major works, but in the conclusion to a smaller article. In the closing remarks to \textit{The Law of the Peoples}, Rawles states,

Political liberalism holds that comprehensive doctrines have but a restricted place in liberal democratic politics in this sense: fundamental constitutional questions and matters concerning basic rights and liberties are to be settled by a public political conception of
justice, exemplified by the liberal political conceptions, and not by these wider doctrines. For given the pluralism of democratic societies – a pluralism that is best seen as the outcome of the exercise of human reason under free institutions and that can only be undone by the oppressive use of state power – affirming such a public conception and the basic political institutions it supports is the most reasonable basis of social unity available to us. 38

Thus Rawles clarifies both his intentions and his bias. His intention was to create a theory of justice that was based on, and created by a “public political” activity as supported by “the pluralism of democratic societies.”

This also exposes his bias. In creating his theory of justice, Rawles relied on many things. The first was his experience with utilitarianism. While he denounced its extreme consequences, he did not fully renounce all of its benefits. As such, he enveloped it in his larger theory. Second, in reaching back to give support to his contractarianism, Rawles relied heavily on the writings of John Locke. As such, Locke’s inherent liberalism found its way into Rawles’s writings. Finally, living his entire life in the US encouraged Rawles to see political liberalism as the best option available. While he did seek to move beyond liberalism, especially through his use of the original position, he never achieved a position outside the system. In the words of Stephen Lukes, “in the end, the ‘Archimedean point for judging the basic structure of society’ that Rawles seeks eludes him.” 39 He does create a theory of justice, but it is a theory of liberal democratic justice.

38Ibid., 68.
6 Selected Bibliography


7 Questions

1. Compare Rawles’ philosophy with that of Locke and Hobbes.

2. Rawles has been called a socialist. Is he?

3. Did Rawles eliminate Utilitarianism or merely transform it?

4. Economist theorists have determined that lexicographical ordering does not happen in the real world, there is a calculation involved among the several options. How does this fact affect Rawles’ philosophy?

5. The author notes that Rawles subtly changed the lexical ordering between A Theory of Justice and Political Liberalism. Comment on how this affects the strength of his earlier arguments. Also comment on how it affects the strengths of his latter arguments.

6. How did Rawles’ upbringing affect his world view?

7. Does the good of the many outweigh the good of the few?

8. Would Rawles hold that Reflective Equilibrium could be used to solve all problems in the world?

9. Did we ever exist under the “veil of ignorance”? How does this affect his arguments?

10. Compare and contrast Rawles with the Marx of the Communist Manifesto.