

# Revisiting Rawls:

*A Theory of Justice* in the light of Levi's theory of decision

by

ERIK ANGNER

Dept. of History and Philosophy of Science and Dept. of Economics

University of Pittsburgh, Pittsburgh PA 15260, USA

Home Phone: 412-361-1128 Fax: 412-624-6825

E-mail: [ejast4@pitt.edu](mailto:ejast4@pitt.edu)

Word count (including footnotes): 7286

Date: 17 December 2002

---

**Acknowledgments:** I thank Ben Eggleston, Andreas Føllesdal, David Gauthier, Isaac Levi, Peter Vanderschraaf, and an anonymous referee for helpful comments on earlier drafts. As usual, all errors remain my own.

# Revisiting Rawls:

## *A Theory of Justice* in the light of Levi's theory of decision

---

**Abstract:** The present paper revisits the issue of rational decision making in John Rawls' original position. Drawing on Isaac Levi's theory of decision, I discuss how we can defend Rawls against John C. Harsanyi's charge that maximin reasoning in the original position is irrational. The discussion suggests that systematic application of Levi's theory is likely to have important consequences for ethics and political theory as well as for public policy.

### *1. Introduction*

Given the recent death of John Rawls, and the fact that three decades have passed since the publication of *A Theory of Justice*, it is perhaps time to reevaluate some of the major themes of the book.<sup>1</sup> In this paper, I propose to revisit the pivotal issue of rational decision making in the original position. As is well known, Rawls argues that agents in the original position should rely on maximin reasoning when choosing principles of justice. The proposition has been sharply criticized by among others John C. Harsanyi, who argues that it would be irrational of these agents to use anything other than subjective expected utility maximization.<sup>2</sup> I think it is

---

<sup>1</sup> John Rawls, *A Theory of Justice* (Cambridge: Harvard University Press, 1971), Revised Edition (Cambridge: Harvard University Press, 1999).

<sup>2</sup> John C. Harsanyi, "Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls' Theory," *American Political Science Review* 69 (1975): 594-606.

fair to say that, according to current consensus, the argument was settled in Harsanyi's favor. As Ken Binmore wrote in 1994: "All the big guns are on Harsanyi's side."<sup>3</sup>

Here, I examine the debate between Rawls and Harsanyi in the light of more recent developments in decision theory. Relying on Isaac Levi's theory of decision,<sup>4</sup> I discuss how one may defend Rawls as against the "big guns." I argue that if we accept Levi's theory of decision, we can reject Harsanyi's charge that maximin reasoning in the original position must be irrational. According to Levi's theory, there are conditions under which it is rationally permissible to have indeterminate probabilities and utilities, and to let choice be guided by some form of the maximin criterion rather than by expected utility maximization. I claim that those conditions (under which maximin reasoning on Levi's account is rational) in fact obtain in the original position as Rawls describes it. Moreover, I discuss how one might go about showing – as Rawls aspires to do – that maximin reasoning in the original position is not only rational, but uniquely rational. This can be done, I suggest, either by strengthening Levi's theory of decision or by changing the description of the original position, but I also discuss the possibility of moderating Rawls' aspiration.

The aim of my paper is not to argue that Rawls' conception of justice is correct, or that Levi does or should endorse it. However, by showing how one might go about defending Rawls against Harsanyi, I do hope to shed some new light on a fundamental question in Rawls' theory of justice, and to demonstrate that, on further reflection, some harshly criticized aspects of his thought may be eminently defensible. Furthermore, I want to provide evidence

---

<sup>3</sup> Ken Binmore, *Game Theory and The Social Contract Volume I: Playing Fair* (Cambridge: MIT Press, 1994), p. 291.

<sup>4</sup> Isaac Levi, *The Enterprise of Knowledge: An essay on knowledge, credal probability, and chance* (Cambridge: MIT Press, 1980), *Hard Choices: Decision making under unresolved conflict* (Cambridge: Cambridge University Press, 1986), and *The Covenant of Reason: Rationality and the commitments of thought* (Cambridge: Cambridge University Press, 1997).

to the effect that systematic application of Levi's theory of decision can have important consequences for both private and public decision making. Levi himself has argued that maximin reasoning is perfectly appropriate when determining whether to rely on nuclear power.<sup>5</sup> Substituting Levi's theory for expected utility theory can have a significant impact on how we think not only about public policy, but about ethics and political philosophy as well.

## ***2. The debate between Rawls and Harsanyi***

In this section I review the background to the debate about rational choice in the original position. I begin by quickly summarizing Rawls' own discussion of decision making in the original position. Then, I present Harsanyi's critique of Rawls' position, and point to some problems with it. The attention to old arguments is worthwhile, I believe, since it helps clarify the contrast with Levi's thinking.<sup>6</sup>

On Rawls' theory, the principles of justice are the terms of cooperation that rational people, under the right circumstances, would agree to follow in their interactions with each other.<sup>7</sup> The circumstances in question are those of the *original position*, in which individuals have been deprived of all morally irrelevant information about themselves, others, and the society in which they live. Rawls writes:

Among its essential features of this situation is that no one knows his place in society, his class position or social status, nor does any one know his fortune in the distribution of natural assets and abilities, his intelligence,

---

<sup>5</sup> Cf. "A Brief Sermon on Assessing Accident Risks in U.S. Commercial Nuclear Power Plants," in Levi, *The Enterprise of Knowledge*, pp. 431-444.

<sup>6</sup> Partially in response to criticism, the later Rawls substantially modified certain aspects of his system. Here, however, I restrict my attention to the 1971 position.

<sup>7</sup> Rawls, *A Theory of Justice*, Revised Edition, p. 10.

strength, and the like. I shall even assume that the parties do not know their conception of the good or their special psychological propensities.<sup>8</sup>

When Rawls says that individuals are behind a *veil of ignorance*, he means that they are in the situation just described.<sup>9</sup>

Having specified the conditions that constitute the original position, Rawls proceeds to discuss the principles of choice that a rational individual would use in the position. He writes:

Understood in this way the question of justification is settled by working out a problem of deliberation: we have to ascertain which principles it would be rational to adopt given the contractual situation. This connects the theory of justice with the theory of rational choice.<sup>10</sup>

That is, on Rawls' account, justifying principles of justice amounts to showing that they would be chosen by rational individuals in the original position. But what principles of rational choice should the agents use? In Rawls' view, a rational agent would use a *maximin* rule. "The maximin rule tells us to rank alternatives by their worst possible outcomes: we are to adopt the alternative the worst outcome of which is superior to the worst outcomes of the others," he tells us.<sup>11</sup> This rule is appropriate, among other things, because knowledge of actual probabilities of outcomes is impossible or insecure.<sup>12</sup>

---

<sup>8</sup> Ibid., p. 11.

<sup>9</sup> Ibid.

<sup>10</sup> Ibid., p. 16.

<sup>11</sup> Ibid., p. 133.

<sup>12</sup> Ibid., p. 134.

Harsanyi attacks Rawls by claiming that it would be irrational to use maximin reasoning in the original position.<sup>13</sup> His argument has two parts, nicely outlined in the following quote:

(a) The only alternative to using subjective probabilities, as required by Bayesian theory, would be to use a decision rule chosen from the maximin-principle family; and ... all these decision rules are known to lead to highly irrational decisions in important cases.

(b) Bayesian decision theory shows by rigorous mathematical argument that any decision maker whose behavior is consistent with a few—very compelling—rationality postulates simply *cannot help* acting as if he used subjective probabilities.<sup>14</sup>

Let me discuss these two lines of argument in order.

According to the first argument, decision making in the original position is a case of decision making under uncertainty, and the use of maximin reasoning in such cases “cannot fail to have highly paradoxical implications.”<sup>15</sup> The claim is illustrated by the following example:

Suppose you live in New York City and are offered two jobs at the same time. One is a tedious and badly paid job in New York City itself, while the other is a very interesting and well-paid job in Chicago. But the catch is that, if you wanted the Chicago job, you would have to take a plane from New York to Chicago (e.g. because this job would have to be taken up the very next day). Therefore there would be a very small but positive probability that you might be killed in a plane accident.<sup>16</sup>

---

<sup>13</sup> Harsanyi, “Can the Maximin Principle Serve as a Basis for Morality?” Harsanyi has other objections to Rawls’ use of maximin reasoning, as well as to Rawls’ own response, “Some Reasons for the Maximin Criterion,” *American Economic Review* 64 (1974): 141-146. I will not discuss these objections here.

<sup>14</sup> Harsanyi, “Can the Maximin Principle Serve as a Basis for Morality?,” p. 599; italics in original.

<sup>15</sup> *Ibid.*, p. 595.

<sup>16</sup> *Ibid.*

What should you do? If you use maximin reasoning, as Harsanyi points out, you should order alternatives according to their worst possible outcomes. Since Harsanyi posits that dying in a plane crash is worse than anything that could happen in the streets of New York, maximin reasoning favors staying there. Harsanyi continues:

Indeed, you must not choose the Chicago job *under any condition*—however unlikely you might think a plane accident would be, and however strong your preference might be for the excellent Chicago job.

Clearly, this is a highly irrational conclusion. Surely, if you assign a low enough probability to a plane accident, and if you have a strong enough preference for the Chicago job, then by all means you should take your chances and choose the Chicago job.<sup>17</sup>

Thus, in summary:

Conceptually, the basic trouble with the maximin principle is that it violates an important continuity requirement: It is extremely irrational to make your behavior wholly dependent on some highly unlikely unfavorable contingencies regardless of how little probability you are willing to assign to them.<sup>18</sup>

Assuming that it would indeed be irrational to choose the New York job under the circumstances described, Harsanyi concludes that the use of maximin reasoning can be highly irrational.

There is something awkward about this argument, however. Harsanyi begins the discussion by specifying that the choice of first principles in the original position is a matter of decision making under uncertainty.<sup>19</sup> According to a classic definition, decision making under uncertainty obtains when the probabilities of outcomes are “completely unknown or are not

---

<sup>17</sup> Ibid.; italics in original.

<sup>18</sup> Ibid.

<sup>19</sup> Ibid., p. 594.

even meaningful.”<sup>20</sup> This appears to describe the original position rather well, as Rawls makes it explicit that “the veil of ignorance excludes all knowledge of likelihoods.”<sup>21</sup> Moreover, maximin is usually presented as a rule for decision making under uncertainty alone.<sup>22</sup> To the best of my knowledge, at least, no one has seriously suggested substituting the maximin principle for the principle of maximizing expected utility under conditions of risk, i.e. when probabilities are known.<sup>23</sup> Yet, Harsanyi argues that maximin reasoning in his example is irrational because *no matter* how unlikely a plane crash (or how much we desire the Chicago job), maximin reasoning still obliges us to take the job in New York. This does not sound like decision making under uncertainty at all, since the relevant probabilities are assumed to be both meaningful and known. If the probability of a plane crash in Harsanyi’s example is known and negligibly low, it is uncontroversial enough that the Chicago job is the better choice.

Since the example appears to be an illustration of decision making under risk, it fails to show that maximin reasoning must have paradoxical consequences under conditions of uncertainty. Of course, Harsanyi could object that any case of decision making under uncertainty becomes a case of decision making under risk, when agents assign determinate probabilities to the possible states (as he would argue they are required to do).<sup>24</sup> This, of course, presupposes a different distinction between risk and uncertainty. Anyway, as we will see below, it is quite possible to deny that agents are rationally required to assign determinate probabilities to states. In the very least, Levi does not think they are. Incidentally, it is

---

<sup>20</sup> R. Duncan Luce and Howard Raiffa, *Games and Decisions* (New York: Wiley and Sons, 1957), p. 13.

<sup>21</sup> Rawls, *A Theory of Justice*, Revised Edition, p. 134.

<sup>22</sup> See e.g. Luce and Raiffa, *Games and Decisions*, p. 278.

<sup>23</sup> *Ibid.*, p. 13.

interesting to consider what would be the rational course of action in Harsanyi's example if we really did not have any idea about the probability of a plane crash. In this case it seems, at least to me, that staying in New York may not be such a bad idea.

Harsanyi's second charge is that we are logically compelled to maximize subjective expected utility since we want our choices to conform to certain "rationally compelling" principles of rationality.<sup>25</sup> Thus, "a rational decision maker simply *cannot help* using subjective probabilities, at least implicitly."<sup>26</sup> No matter what the rational person has in mind when making decisions, in this view, her choices will always reflect and reveal some consistent assignment of probabilities and utilities. Harsanyi is right, of course, to maintain that *if* our choices satisfy certain assumptions, *then* we can be ascribed a utility function and a probability function such that our choices can be described as maximizing subjective expected utility. This much is established by any one of a number of representation theorems.<sup>27</sup>

Yet, the issue is whether the axioms of rational preference are quite so compelling as Harsanyi suggests. By now, a rather extensive literature questions the normative status of these axioms. Suppose, for example, that I offer you the choice between a holiday in Florida and a holiday in California.<sup>28</sup> Suppose, moreover, that given all you know about hotel standards, weather conditions, and whatnot, you do not have a strict preference either way. Now, suppose I were to offer you the choice between a holiday in Florida and one apple, and a holiday in California. Provided that you like apples, and that the utility of apples does not

---

<sup>24</sup> See Levi, *Covenant of Reason*, p. 188-189.

<sup>25</sup> Harsanyi, "Can the Maximin Principle Serve as a Basis for Morality?," p. 599.

<sup>26</sup> *Ibid.*; italics in original.

<sup>27</sup> See e.g. Luce and Raiffa, *Games and Decisions*, section 2.5.

<sup>28</sup> The example is adapted from Sven Danielsson, "Numerical Representations of Value-Orderings: Some basic problems," in Christoph Fehige and Ulla Wessels (eds) *Preferences* (Berlin: de Gruyter, 1998), pp. 116-117.

decrease when bundled with holidays in Florida (or the other way around), the axioms of orthodox decision theory require you to accept the Florida holiday with an apple. For, if you do not have a strict preference for either one of the original alternatives, completeness implies that you are indifferent between the two. Since you strictly prefer {Florida, apple} to {Florida}, and you are indifferent between {Florida} and {California}, transitivity requires you to strictly prefer {Florida, apple} to {California}. Yet, it seems perfectly rational to not develop a strict preference even after you are offered the apple if you go to Florida. If so, at least one of the axioms must be rejected, and the choices made by a rational agent need not be compatible with any consistent assignment of probabilities and utilities.

### ***3. Isaac Levi on maximin reasoning***

Having pointed to various apparent deficiencies in Harsanyi's argument, I now wish to outline a more positive case for maximin reasoning in the original position. I rely on the theory of decision developed by Isaac Levi, according to which maximin reasoning is perfectly rational under certain conditions. I will claim that those conditions indeed hold in the original position, but I begin by providing a short outline of Levi's theory of decision.<sup>29</sup>

On the standard theory of rational choice under risk, each agent has a unique probability distribution over states of the world, and a unique utility function over outcomes.<sup>30</sup> The probability distribution reflects the subjective probabilities that the agent assigns to the occurrence of each state, and the utility function reflects the agent's valuations of the various

---

<sup>29</sup> For a somewhat more extensive discussion of Levi's theory, in the context of his claim that it can accommodate a number of paradoxical experimental results, see Erik Angner, "Levi's Account of Preference Reversals," *Economics and Philosophy* (in press).

<sup>30</sup> Cf. Levi, *The Covenant of Reason*, pp. 119-120.

outcomes. Given a probability distribution and a utility function, both with the appropriate properties, the agent can calculate the expected utility of each alternative course of action. She then chooses the optimal alternative with respect to expected utility, if there is a uniquely optimal one, or one of the optimal alternatives, if there is more than one.

Levi's theory of decision is a generalization of the standard one. According to this theory, each agent has a set of probability distributions, called *permissible* probability distributions, and a set of utility functions, called *permissible* utility functions.<sup>31</sup> When facing a number of alternatives, the agent first calculates which alternatives are *E-admissible*, where an alternative *A* is *E-admissible* if and only if there is some permissible probability distribution and some permissible utility function such that *A* is optimal relative to the probability distribution and utility function. If there is only one *E-admissible* alternative, the agent chooses that one. In case there is more than one *E-admissible* alternative, the agent relies on some secondary criterion. In particular, she may choose some *E-admissible* alternative that maximizes security, using a lexicographical maximin (*leximin*) criterion to assess security levels.<sup>32</sup> That is, if the worst possible outcome of the one is better than the worst possible outcome of the other, the former will be chosen. If the worst possible outcomes of the two alternatives are equivalent, the agent will consider the second worst outcome, etc. The agent is not, however, rationally *required* to rely on security considerations to break ties; other secondary criteria may be used as well. In Levi's account, the choice of a secondary criterion is best seen as a value commitment on the part of the agent.<sup>33</sup>

---

<sup>31</sup> Ibid., p. 126.

<sup>32</sup> Ibid., p. 194.

<sup>33</sup> Personal communication.

Since a rational agent may have several probability distributions and utility functions, in Levi's view, there may be a certain indeterminacy in utility and probability judgments. What is the rationale for making this assumption?

A feature of any acceptable view of inquiry ought to be that during an inquiry points under dispute ought to be kept in suspense pending resolution through inquiry ... This consideration ought to lead to a form of indeterminacy in probability judgment, utility judgment and assessments of expected utility.<sup>34</sup>

There are times, such as during inquiry, when a rational agent should suspend judgment and abstain from assigning a unique probability to the occurrence of an event or a unique utility to an outcome. When she lacks sufficient grounds to make such an assignment, she should assign a set of values to the probability of an occurrence or to the utility of an outcome. "I interpret indeterminacy in probability and value judgment as doubt or suspense with respect to probability or value," Levi says.<sup>35</sup> Levi's rejection of determinacy is supported also by examples such as the one about apples and holidays, for his theory allows us to say that your probabilities or utilities were, and remained, indeterminate.

Levi's theory permits us to draw a sharper distinction between decision making under uncertainty and decision making under risk. Indeed, he argues that this is one of the advantages of the theory.<sup>36</sup> As he writes:

I have tended to think of decision making under risk as arising in any context where an agent legitimately assigns numerically definite credal probabilities to hypotheses about the outcomes of feasible options. In other words,

---

<sup>34</sup> Levi, *The Covenant of Reason*, p. 185.

<sup>35</sup> *Ibid.*, p. xi.

<sup>36</sup> Levi, *Enterprise of Knowledge*, p. 145.

decision making under risk is any context in which maximizing expected utility is the appropriate criterion for identifying admissible options.<sup>37</sup>

Thus, uncertainty obtains just in case it is inappropriate to assign definite probabilities to possible states, so that some secondary criterion needs to be invoked in order to break ties. Below, I will argue that the circumstances of Rawls' original position make it precisely such a case.

On this view, an agent has a preference for some alternative  $A$  over some other alternative  $B$  if and only if for all permissible probability distributions and utility functions, the expected utility of  $A$  is higher than that of  $B$ .<sup>38</sup> This implies that the preference relation may be incomplete; if two alternatives are  $E$ -admissible then the agent simply lacks a preference over them.<sup>39</sup> It should be noted that a lack of preference does not inhibit the agent from making a rational choice between alternatives, since the secondary criterion will tell her what to do. It follows that on Levi's theory a choice of  $A$  rather than  $B$  need not reveal a preference for  $A$  over  $B$ , as Harsanyi assumes in his argument against Rawls.

As an example of how Levi's theory can be used in decision making, consider the decision problem of the job candidate considered above. Let us assume, to begin with, that her values over the various outcomes are clear, that is, that the utilities are determinate. Suppose the utility of taking the job in New York is one, that of taking the job in Chicago ten, and that of dying in a plane crash -890. Then the decision problem can be represented on matrix form as in Table 1.

---

<sup>37</sup> Levi, *The Covenant of Reason*, p. 190.

<sup>38</sup> Such preferences are referred to as *categorical preferences*. See Levi, *Hard Choices*, p. 91.

<sup>39</sup> Cf. *ibid.*, p. 196.

\*\*\* INSERT TABLE 1 ABOUT HERE \*\*\*

Now, the rational course of action depends on the probabilities assigned to the two relevant events. If the agent has determinate probabilities over the two events, it is easy to confirm that she should take the job in Chicago so long as the probability of a plane crash is less than one percent, and stay in New York if it greater than one percent; if the probability is exactly one percent, she is indifferent, and is rationally permitted to take either job. If the agent has indeterminate probabilities, we can also identify three different cases. First, the range of probabilities may be well below one percent. In this case taking the Chicago job is favored by all probability functions, so that Chicago is uniquely *E*-admissible. This corresponds to the case when the probability of a plane crash is low enough that it can be tolerated. Second, the range of probabilities may be well above 1 percent. In this case taking the New York job is favored by all probability functions, so that New York is uniquely *E*-admissible. This corresponds to the case when the probability of a plane crash is too high to be tolerated. Third, the range of probabilities may include 1 percent. If so, both alternatives come out maximal for some probability function, and both are *E*-admissible. This corresponds to the case when expected utility considerations fail to determine whether the probability of a plane crash is low enough to be tolerated or not. In this case, according to Levi's theory, the agent may very well use maximin reasoning and choose New York (something that does not imply that she *prefers* New York). The example illustrates that Levi can express the conditions under which maximin reasoning is rational in a quite precise manner.

It has been objected that maximin reasoning is an expression of paranoia or excessive pessimism.<sup>40</sup> This is not, however, an adequate criticism of Levi's account.<sup>41</sup> Note that the agent who uses security considerations to break ties between *E*-admissible alternatives does *not* assume that the worst outcome is most likely to happen. For in such cases, by assumption, considerations of probability and utility fail to reach a verdict. It may of course be true that in some situations, the maximin criterion favors that alternative which the paranoid pessimist would choose. Yet, behaving like a paranoid pessimist under some narrowly defined set of conditions does not make you one.

Now, how can Levi's theory illuminate rational decision making in the original position? Clearly, on Levi's account, maximin reasoning in the original position can be justified if the beliefs and values of agents are sufficiently indeterminate to make several alternatives *E*-admissible. Consider the following highly simplified case. Suppose that an agent in the original position is supposed to choose between the principles of justice favored by Rawls, and those favored by Harsanyi. The agent is concerned that she may be poor, and knows that Harsanyi's principle permit quite substantial welfare differences for the benefit of total utility. What is the probability that she will be poor? Knowledge of probabilities is hidden behind the veil of ignorance, but on the basis of some general considerations she may be happy to assign a probability somewhere between, say, 20 and 80 percent. Similarly, she has only a very rough idea concerning her welfare given the different outcomes, and assigns a range of utilities to each outcome. Thus, we can present the decision problem as in Table 2, where ranges indicate that the utilities are indeterminate.

---

<sup>40</sup> E.g. in Binmore, *Game Theory and the Social Contract*, p. 318.

\*\*\* INSERT TABLE 2 ABOUT HERE \*\*\*

Given this characterization of the decision problem, it is a simple matter to find probabilities and utilities such that adopting Rawls' principles is optimal; similarly, it is straightforward to find probabilities and utilities such that adopting Harsanyi's principles is optimal. Thus, both alternatives are *E*-admissible, and a choice between them may be made on the basis of maximin considerations. If so, Rawls' principles are uniquely admissible. In what follows I argue that the picture this example gives of the original position is adequate. Provided Rawls' characterization of the original position, agents' probabilities and utilities may be sufficiently indeterminate to justify the use of maximin reasoning.

First of all, let us consider what kind of consideration would be relevant to the decision that must be made in the original position. We can assume, as Rawls does, that the parties to the original agreement are free and rational, and concerned with furthering their own interests.<sup>42</sup> In order to make a fully informed decision about what principles of justice to adopt, agents would need to know what their position is in society, what their endowments are, how various privileges are distributed, and so on. Moreover, they would need to know just how much they appreciate or dislike different principles of association given their position, tastes, etc. However, this kind of information is precisely the kind that is hidden behind the veil of ignorance. Rawls writes:

It is assumed, then, that parties do not know certain kinds of particular facts. First of all, no one knows his place in society, his class position, or social status; nor does he know his fortune in the distribution of natural assets and abilities, his intelligence and strength, and the like. Nor, again, does anyone know his conception of the

---

<sup>41</sup> Levi, *Covenant of Reason*, p. 197.

good, the particulars of his rational plan of life, or even the special features of his psychology such as his aversion to risk or liability to optimism or pessimism. More than this, I assume that the parties do not know the particular circumstances of their own society. That is, they do not know its economic or political situation, or the level of civilization and culture it has been able to achieve. The persons in the original position have no information as to which generation they belong.<sup>43</sup>

Indeed, the only things agents know in the original position are some fairly general facts about humankind and the bounded amount of resources available at any point in time. “As far as possible ... the only particular facts which the parties know is that their society is subject to the circumstances of justice and whatever this implies,” Rawls writes.<sup>44</sup> He clarifies that “the circumstances of justice obtain whenever persons put forward conflicting claims to the division of social advantages under conditions of moderate scarcity.”<sup>45</sup>

In Harsanyi’s view, it is quite possible – indeed necessary – for the rational agent to adopt precise utilities and probabilities in the original position. According to his account, agents in the original position should assign determinate probabilities to the various positions in which they may find themselves, assign utilities to the specific outcomes, and choose principles of cooperation on the basis of expected utility considerations. Specifically, they should assign a probability of  $1/n$  to each of the cases “I am the best-off individual,” “I am the second-best-off individual,” and so on.<sup>46</sup> Then, assuming the right kind of utility function, the agent can choose between alternative social systems following the principle of expected utility.

---

<sup>42</sup> Rawls, *A Theory of Justice*, Revised Edition, p. 10.

<sup>43</sup> *Ibid.*, p. 118.

<sup>44</sup> *Ibid.*, p. 119.

<sup>45</sup> *Ibid.*, p. 110.

<sup>46</sup> Harsanyi, “Can the Maximin Principle Serve as a Basis for Morality?,” p. 598.

However, once we give up the assumption that probabilities and utilities need to be determinate, we can ask whether it is justified to adopt precise probability and utility functions under the conditions Rawls describes. It may seem that every such attempt would be frustrated by the crucial lack of information imposed by the veil of ignorance. Thus, it should be entirely appropriate to say that the agents in the original position are in doubt or suspense. Consider the assignment of probabilities and utilities, in turn.

First, it is unclear on what basis agents should assign probabilities to the various possibilities. An agent might want to assign a certain probability to the prospect of being poor, by some measure. Yet, lacking information about how many people in society are poor, this is a challenging task. The same thing should be true for the probability that she is religious, chronically ill, well educated, illiterate, or whatnot. Given the vast amount of uncertainty about the possibilities, it seems entirely appropriate to say that the agent indeed is in suspense concerning these probabilities, and let her leave them radically indeterminate.

Furthermore, even if agents do find some way to assign probabilities, à la Harsanyi, it is unclear on what basis they should assign utilities. They do not know just how badly off the worst-off individuals are, or how well off the best-off are. Neither do they know how wealth and other resources are distributed. Moreover, and more importantly, they do not know their conception of the good. That is, they do not know how much they value material goods, religious freedom, or freedom of choice more generally. Thus, they lack information that is crucial to assigning utilities to outcomes, and they can rightly be said to be in suspense with respect to their value judgments. While they can assume a few things about themselves, such as the fact that they have a conception of the good, this is not much. Pending more information about the society in which they live and their conception of the good, it seems perfectly appropriate for the agents to leave their utilities radically indeterminate.

Given the combination of radically indeterminate probabilities and radically indeterminate utilities, the set of *E*-admissible alternatives is likely to be large. Excluding alternatives that are inferior on all accounts, it seems likely that one competitor will be optimal given one pair of probability and utility function, and that another competitor will be optimal given another pair. Thus, using Levi's terminology, all alternatives will be *E*-admissible. Since expected utility considerations fail, agents may perfectly rationally resort to maximin reasoning when choosing between them. In short, if we accept Levi's theory of decision, an agent behind the veil of ignorance is rationally permitted to rely on maximin reasoning, and Harsanyi's critique fails. Binmore writes: "The circumstances of the original position look to me as though they might have been tailor-made for applying Bayesian decision theory."<sup>47</sup> In contrast, it seems to me as though the original position was tailor-made for applying Levi's theory of decision, and a paradigmatic case when indeterminate probabilities and utilities, and therefore maximin reasoning, are perfectly rational.

In an interesting though brief paper called "Four types of ignorance," Levi himself addresses the debate between Rawls and Harsanyi.<sup>48</sup> Levi notes that a main point of contention between Rawls and Harsanyi concerns the state of ignorance that affects agents in the original position. In Levi's terms, Harsanyi takes the state of ignorance in question to be one of *probabilistic ignorance*: the agent assigns the same probability (in this case  $1/n$ ) to each hypothesis about his position in society.<sup>49</sup> Meanwhile, Rawls takes the agent in the original position to be in a state of *extreme probabilistic ignorance*: he or she has not ruled out any

---

<sup>47</sup> Binmore, *Game Theory and the Social Contract*, p. 316.

<sup>48</sup> Isaac Levi, *Decisions and Revisions: Philosophical essays on knowledge and value* (Cambridge: Cambridge University Press, 1984), chapter 9. I am grateful to Levi for calling my attention to this essay.

<sup>49</sup> *Ibid.*, pp. 130-131.

probability distribution that satisfies the requirements of the probability calculus.<sup>50</sup> The example serves to show that notions of ignorance matter:

Thus our efforts to distinguish between different views of what constitutes a state of ignorance amounts to rather more than an elaboration of entries in some philosophically sophisticated lexicon. In Harsanyi's case, his conception of ignorance is critical to his assessment of the conditions under which moral judgments concerning social welfare may be made.<sup>51</sup>

Levi does not, however, discuss what features of Rawls' original position makes extreme probabilistic ignorance more appropriate, or detail the manner in which the rational agent's state of ignorance depends jointly on our description of the original position and on our conception of rationality.

The adoption of Levi's theory of decision does not, unfortunately, by itself go all the way toward defending Rawls' claims about rational choice in the original position. In his book, Rawls aspires to show that the choice of his principles of justice behind the veil of ignorance is not only rational, but uniquely rational. He writes: "Ideally, anyway, I should like to show that their acknowledgement is the only choice consistent with the full description of the original position."<sup>52</sup> My argument up to this point has only shown that, given Levi's theory of decision, maximin reasoning (and therefore the choice of Rawls' principles of justice) is consistent with rationality. As we have seen, Levi allows agents to rely on secondary criteria other than maximin reasoning if they so desire, and he sees the choice of a secondary criterion as a value commitment on the part of the agent. Incidentally, Rawls does not believe his own argument – at least as developed in *A Theory of Justice* – meets his expectations. By the end of

---

<sup>50</sup> Ibid., pp. 132-133.

<sup>51</sup> Ibid., p. 131.

<sup>52</sup> Rawls, *A Theory of Justice*, Revised Edition, p. 104.

the paragraph quoted above, he writes: “Unhappily, the reasoning I shall give will fall far short of this.... Yet it is essential to have in mind the ideal one would like to achieve.”<sup>53</sup> Thus, if the argument outlined so far could be extended to show maximin reasoning uniquely rational, that would constitute a significant improvement over Rawls’ original treatment.

How, then, might one go about modifying the account so as to make maximin reasoning uniquely admissible? I see two main ways to achieve this goal. The first strategy involves strengthening Levi’s theory of decision so as to make maximin reasoning the only secondary criterion consistent with rationality. There are, of course, reasons to favor the maximin criterion as a principle of choice under uncertainty.<sup>54</sup> The issue would be solved if we could develop an argument to the effect that a maximin criterion is the only secondary criterion compatible with rationality. The problem, of course, is that knock-out arguments in favor of criteria of decision making under uncertainty are hard to come by. The second strategy involves changing the description of the original position, so as to include the assumption that agents in fact have committed to using maximin reasoning as a secondary criterion. On Levi’s view, this amounts to ascribing to the agents a certain value judgment. As Rawls notes, we have already made a number of assumptions about their “beliefs and interests.”<sup>55</sup> All we need to do is to add one more assumption of this kind. The difficulty here is that it is unclear on what grounds one may justify a particular characterization of the original position.<sup>56</sup>

---

<sup>53</sup> Rawls, *A Theory of Justice*, Revised Edition, p. 105.

<sup>54</sup> See Luce and Raiffa, *Games and Decisions*, chapter 13, for a classic discussion of decision making under uncertainty.

<sup>55</sup> Rawls, *A Theory of Justice*, Revised Edition, p. 104.

<sup>56</sup> Cf. Levi, “Four Types of Ignorance,” pp. 133-134.

Perhaps a more promising approach involves rejecting the notion that the principles of justice have to be the unique rational choice by agents in the original position.<sup>57</sup> Instead, we could adopt e.g. some version of T. M. Scanlon's idea that the principles of justice only need be such that they cannot reasonably be rejected by the agent, no matter what his type turns out to be.<sup>58</sup> Scanlon writes:

Whatever rules of rational choice this single individual, concerned to advance his own interests as best as he can, is said to employ, this reduction of the problem to the case of a single person's self-interested choice should arouse our suspicion.... [It] is important to ask whether this single individual is held to accept a principle because he judges that it is one he could not reasonably reject whatever position he turns out to occupy, or whether, on the contrary, it is supposed to be acceptable to a person in any social position because it would be the rational choice for a single self-interested person behind the veil of ignorance.<sup>59</sup>

Of course, at least on the face of it, Rawls' argument is of the latter kind. Scanlon continues: "I would claim, however, that the plausibility of Rawls' arguments favouring his two principles over the principle of average utility is preserved, and in some cases enhanced, when they are interpreted as instances of the first form of contractualist argument."<sup>60</sup> Presumably, it is easier to argue that a given set of principles cannot be reasonably rejected, than to argue that they are rationally required behind the veil of ignorance. Scanlon's approach too may have its drawbacks. For one thing, since the test he proposes is somewhat less stringent than Rawls' own, it is less likely to identify a unique set of principles of justice. In Scanlon's words: "It seems likely that many non-equivalent sets of principles will pass the test of non-

---

<sup>57</sup> I thank Andreas Føllesdal for offering this suggestion.

<sup>58</sup> T. M. Scanlon, "Contractualism and Utilitarianism," in Amartya Sen and Bernard Williams (eds) *Utilitarianism and Beyond* (Cambridge: Cambridge University Press, 1982), pp. 103-128.

<sup>59</sup> *Ibid.*, pp. 124-125.

<sup>60</sup> *Ibid.*, pp. 125; cf. pp. 125-128.

rejectability.”<sup>61</sup> To what extent this is a problem for the theory of justice is an issue that falls outside the scope of the present paper.

Before closing, I should note that I have not offered a complete argument in defense of Rawls’ claims about rational decision making in the original position. Most importantly, I have not developed a systematic defense of Levi’s theory of decision, even though the defense of Rawls I suggest relies heavily on its correctness. Nevertheless, I believe that I have shown that Rawls’ account cannot be dismissed quite so easily as Harsanyi and Binmore think. Moreover, I hope my discussion can serve as a road map of sorts, to be used by somebody who aspires to develop a more complete defense of Rawls’ theory. In the course of the discussion, I have tried to identify and discuss a number of routes that can be taken in order to pursue that goal.

#### ***4. Conclusion***

In this paper, I have critically reviewed John C. Harsanyi’s critique of John Rawls’ theory of justice. In particular, I have focused on Harsanyi’s claim that it would be irrational of agents in the original position to rely on maximin reasoning when choosing principles of justice. Though I believe that Rawls’ own response to the critique is unconvincing, I have pointed out a number of apparent deficiencies in Harsanyi’s arguments. First, Harsanyi fails to show that the principle of maximin commits agents to irrational decisions under conditions of uncertainty, since his examples appear to concern cases of decision making under risk. Second, the axioms of orthodox decision theory are not as rationally compelling as Harsanyi

---

<sup>61</sup> Ibid., p. 112.

believes, and that representation theorems therefore fail to prove that every rational individual acts as if she maximized subjective expected utility.

Furthermore, I have outlined a more positive argument to the effect that maximin reasoning is rational in the original position. Relying on Isaac Levi's theory of rational decision, I have argued that the conditions under which it is rational to use maximin reasoning indeed hold in the original position. If we accept Levi's theory, then, it follows that maximin reasoning is entirely appropriate in the original position. Also, I have discussed how one may go about arguing that maximin reasoning in the original position is uniquely rational. That goal can be achieved, I have suggested, either by strengthening the theory, or by modifying the characterization of the original position. Even so, the most promising approach may be to reject Rawls' notion that the principles of justice need to be the unique choice by agents behind the veil of ignorance. Though I have not aspired to defend Rawls' theory of justice, the argument suggests strategies that may be taken by somebody who does.

Whether or not Levi's theory ultimately helps Rawls rescue his argument from *A Theory of Justice*, the preceding discussion strongly suggests that substituting Levi's theory of decision for expected utility theory can have important consequences not only for public policy, as suggested by Levi's discussion of nuclear power, but for ethics and political philosophy as well. Whenever there is doubt in probability or value, something that is common in many contexts of decision making, Levi's theory may give substantially different – and arguably superior – advice about the correct course of action.

### ***References***

Angner, Erik. "Levi's Account of Preference Reversals," *Economics and Philosophy* (in press).

Binmore, Ken. *Game Theory and The Social Contract Volume I: Playing Fair*. Cambridge: MIT Press, 1994.

- Danielsson, Sven. "Numerical Representations of Value-Orderings: Some basic problems." In Christoph Fehige and Ulla Wessels (eds) *Preferences*. Berlin: de Gruyter, 1998, pp. 114-122.
- Harsanyi, John C. "Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls' Theory," *American Political Science Review* 69 (1975): 594-606.
- Levi, Isaac. *The Enterprise of Knowledge: An essay on knowledge, credal probability, and chance*. Cambridge: MIT Press, 1980.
- Levi, Isaac. *Decisions and Revisions: Philosophical essays on knowledge and value*. Cambridge: Cambridge University Press, 1984.
- Levi, Isaac. *Hard Choices: Decision making under unresolved conflict*. Cambridge: Cambridge University Press, 1986.
- Levi, Isaac. *The Covenant of Reason: Rationality and the commitments of thought*. Cambridge: Cambridge University Press, 1997.
- Luce, R. Duncan and Howard Raiffa. *Games and Decisions*. New York: Wiley and Sons, 1957.
- Rawls, John. *A Theory of Justice*. Cambridge: Harvard University Press, 1971.
- Rawls, John. "Some Reasons for the Maximin Criterion," *American Economic Review* 64 (1974): 141-146.
- Rawls, John. *A Theory of Justice*. Revised Edition. Cambridge: Harvard University Press, 1999.
- Scanlon, T. M. "Contractualism and Utilitarianism." In Amartya Sen and Bernard Williams (eds) *Utilitarianism and Beyond*. Cambridge: Cambridge University Press, 1982, pp. 103-128.

\*\*\* TABLES \*\*\*

	<b>Plane crashes</b>	<b>Plane does not crash</b>
<b>Take job in NYC</b>	1	1
<b>Take job in Chicago</b>	-890	10

*Table 1.*

	<b>Decision maker is poor</b>	<b>Decision maker is rich</b>
<b>Adopt Rawls' principles</b>	1-10	2-20
<b>Adopt Harsanyi's principles</b>	0-2	10-100

*Table 2.*