
Von Neumann/Morgenstern Decision Making and Harsanyi's Theory of Justice: After Rawls and Nozick, If It Risks Personal Freedom and Individual Liberty, Can It Be Really Just?

Author(s): Will Carrington Heath

Source: *The American Journal of Economics and Sociology*, Jul., 1988, Vol. 47, No. 3 (Jul., 1988), pp. 355-362

Published by: American Journal of Economics and Sociology, Inc.

Stable URL: <https://www.jstor.org/stable/3486484>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



is collaborating with JSTOR to digitize, preserve and extend access to *The American Journal of Economics and Sociology*

JSTOR

von Neumann/Morgenstern Decision Making and Harsanyi's Theory of Justice:

After Rawls and Nozick, If It Risks Personal Freedom and Individual Liberty, Can It Be Really Just?

By WILL CARRINGTON HEATH*

ABSTRACT. John C. *Harsanyi* has developed a theory of *justice* based on the *von Neumann-Morgenstern* (vN-M) theory of *decision making* under *uncertainty*. Instead of applying the vN-M theory strictly, however, Harsanyi proposes to modify his approach so as to exclude decisions which seem to be the result of rashness, strong excitement and other circumstances which might hinder *rational choice*. The result is a theory of justice which is internally inconsistent and potentially hostile to *liberty* and *individual freedom*.

I

THEORIES OF JUSTICE that have some basis in economic theory have attracted much attention in recent years. John Rawls' book, *A Theory of Justice*, which appeared in 1975, has been extremely influential among economists. Five years later economist Joseph W. Spengler added to the revival of interest in this area with the publication of his book, *Origins of Economic Thought and Justice*. The libertarian philosopher Robert Nozick also created quite a stir among economists with his controversial book, *Anarchy, State and Utopia*.

The list of philosophers who have written on economics and justice would not be complete without the name of John C. Harsanyi. Harsanyi had been studying the welfare economics of A. C. Pigou since the 1950s and in 1953 and 1955 he published articles on cardinal utility, welfare and ethics in the University of Chicago journal, *The Journal of Political Economy*.¹ On the appearance of Rawls' book Harsanyi published a critique, and a year later came out with his own book on the subject, *Essays on Ethics, Social Behavior and Scientific Explanation*.

While Harsanyi has not received the notoriety of Rawls or Nozick, neither has he been overlooked by economists. His work has been favorably reviewed by such eminent economists as Kenneth Arrow, who wrote the foreword to Harsanyi's book,² and Paul Samuelson, who went so far as to say that Harsanyi had achieved "one of the few quantum leaps" in welfare economics.³ Harsanyi has

* [Will Carrington Heath, Ph.D., is assistant professor of economics, Birmingham Southern College, 800 Eighth Avenue, West, Birmingham, AL 35254.]

American Journal of Economics and Sociology, Vol. 47, No. 3 (July, 1988).

© 1988 American Journal of Economics and Sociology, Inc.

his critics as well, including John Rawls, Amartya Sen and Leland Yeager, among others. Generally speaking, the critics have tended to focus on three aspects of Harsanyi's approach: its allegedly anti-egalitarian bias, the von Neumann-Morgenstern analysis that underlies so much of Harsanyi's theorizing, and Harsanyi's willingness to make interpersonal comparisons of utility.

We shall look at their criticisms only briefly, for there is a more serious flaw in Harsanyi's theory yet to be considered. It will be argued here that any attempt to put his theory into practice would likely mean violating one of its own most fundamental ethical principles. Beyond this problem of internal inconsistency, there are also disturbing implications for personal freedom and individual liberty.

II

HARSANYI BEGINS with the principle that correct value judgments for a theory of social welfare must be made on the basis of moral preferences, as distinguished from personal preferences. According to Harsanyi, a person has "two different sets of preferences: he will have a set of personal preferences, which may give particularly high weight to his personal interests (and to those of his close associates); and he will have a set of moral preferences, based on a serious attempt to give the same weight to the interests of every member of society. . . ."⁴

Judgments based entirely upon moral preferences are highly improbable in real life, of course. So Harsanyi constructs a hypothetical situation in which these kinds of judgments are supposed to be guaranteed. He imagines a rational individual (or "sympathetic observer") making choices among different social arrangements, constrained by two important stipulations. First, the individual faces an equal chance of taking the place of any member of society.⁵ Second, all choices must be made in accordance with the feelings and attitudes (*i.e.*, the utility functions) of those whose situations are being considered.⁶

The first stipulation makes it impossible to know in advance what the individual's station in society will turn out to be. When one's own position cannot be predicted, rationality demands that all possibilities be considered. This stipulation is therefore meant to insure that all judgments will be impersonal in the sense that they will not be tailored to the contingencies of one particular set of circumstances.

The second stipulation is of crucial substantive importance. It serves, in effect, to invoke the principle commonly known as the Golden Rule:

The hypothetical sympathetic observer must judge the consequences that a given action has "sympathetically" in the literal sense of the term, *i.e.*, in terms of the attitudes, wants, desires, preferences of these people themselves. . . . Only this theory can do full justice to the

traditional principle that morality essentially consists in following the rule: "Treat people the same way as you want to be treated yourself."⁷

Harsanyi argues that placing an individual in this sort of hypothetical situation (his version of the Rawlsian "original position") amounts to presenting one with a choice among lottery tickets.⁸ Each ticket represents a different society, with different opportunities for different people. For instance, one society might offer a relatively high per capita income, but tolerate the practice of slavery. Another society might offer a lower level of income, but guarantee equality before the law for everyone.

Which society would be chosen? The rationale for decision making in precisely this kind of situation has been developed in economic theory by Messrs. John von Neumann and Oskar Morgenstern.⁹ If the von Neumann-Morgenstern (vN-M) analysis of choice under uncertainty is correct, then the rational person would be expected to choose the society that offers the most promising prospects overall—technically, the greatest average expected utility.

A social welfare function defined to include all individuals in society would thus amount to an aggregate vN-M utility function to be maximized in terms of average utility. Harsanyi takes such a utility function as the starting point for building the Good Society. His rule for society's architects is at once simple and complex: Give us the kind of society that will maximize average utility:

In my model, every person making a moral value judgment will evaluate any institutional arrangement in terms of the average utility level it yields for the individual members of the society, *i.e.*, in terms of the arithmetic mean of these individuals' vN-M utility functions. This means that, under my theory, people's vN-M utility functions enter into the very definition of justice. . . .¹⁰

A brief digression on the meaning of utility is in order, even at the risk of raising more questions than we answer. Von Neumann and Morgenstern use the term "utility" in the context of modern microeconomics, not 18th century psychology or Benthamite utilitarianism. Modern economic theorists regard "utility" as a convenient term for the explicit value of a function which describes human behavior in a choice-theoretic framework. As Milton Friedman puts it:

We observe that people choose; if this is to be regarded as a deliberative act, it must be supposed that the various things among which choice is made can be compared; to be compared, they must have something in common . . . [W]e call this common characteristic utility. . . .¹¹

Friedman's colleague at the University of Chicago, George Stigler, contends that "It does not affect the formal theory of demand in the least whether the individual maximizes wealth, religious piety, the annihilation of crooners or one's waistline."¹² Clearly there is no presumption that utility is to be measured in some psychological unit. Choice is the essence of "utility" in microeconomic

theory, not the “pleasure” and “pain” of utilitarianism, and the vN-M utility function is not meant to say anything about utility in the psychological sense. Rather, it is intended to describe human choice in circumstances of risk and uncertainty.

John Rawls questions whether such a utility function has any place in a theory of justice.¹³ Harsanyi argues that it does.

To be sure, the vN-M utility function of any given individual is estimated from his choice behavior under risk and uncertainty. But this does not mean that his vN-M function is merely an indication of his attitudes toward risk taking. Rather, as its name shows, it is a utility function. This means that the primary task of a vN-M utility function is not to express a given individual's attitude toward risk taking; rather, it is to indicate how much utility, *i.e.*, how much subjective importance, he assigns to various goals. . . . Consequently, vN-M utility functions have a completely legitimate place in ethics because they express the subjective importance people attach to their various needs and interests. For example, I cannot see anything wrong with a concept of justice which assigns high priority to providing university education for a given individual partly on the ground that he attaches very high utility to receiving such an education . . . as shown by the fact that he would be prepared to face very considerable personal and financial risks, if he had to, in order to obtain a university education.¹⁴

III

AMARTYA SEN HAS ARGUED that Harsanyi's welfare function is not sufficiently egalitarian to be a morally acceptable guide for social organization.¹⁵ Harsanyi replies that his welfare function simply reflects the egalitarianism of society as a whole.¹⁶ Remember, the sympathetic observer takes as “given” each person's attitudes about everything, including the distribution of wealth and income. Those who believe Harsanyi's theory is not sufficiently egalitarian either do not understand this point, or they just reject the views of non-egalitarians in society. If the latter is the case, then they would impose their own (egalitarian) values in a way that is antithetical to Harsanyi's whole approach.

Leland Yeager focuses on a different aspect of Harsanyi's theory. Yeager has doubts about the rationality of the vN-M analysis itself, and questions the conclusion that rational persons will always opt for the greatest expected utility.¹⁷ Indeed, the conclusions of the vN-M analysis are not immediately obvious. Fortunately, the principal insights of the vN-M analysis can be made more accessible by replacing the original analysis with an equivalent set of postulates that are a bit more straightforward. Several writers have offered such postulates, perhaps the best known being those developed by Jacob Marshack.¹⁸ They can be summarized as follows:

- (1) An individual's preferences are completely ordered. Any two “prospects can be ordered in the sense that prospect A is preferred to B, B is preferred to A, or the individual is indifferent between them.

- (2) If A is preferred to B and B to C, then there exists a probability p between 0 and 1 such that the gamble A with probability p and C with probability $1 - p$, which can be stated as $[pA + (1 - p)C]$, will be indifferent (*i.e.*, equally preferred) to B.
- (3) For any object of choice or "prospect" A and for any probability p one can specify another prospect B such that A will not be indifferent to the probability combination $[pA + (1 - p)B]$. The significance of this is that a gamble in which the probability of a given prospect of A is not 1, however close to 1 it may be, cannot be regarded as the equivalent of A with certainty.
- (4) If A and B are indifferent and p is between 0 and 1, then $[pA + (1 - p)C]$ is indifferent to $[pB + (1 - p)C]$.

These postulates are certainly straightforward, and they seem to be consistent with rational behavior. If they are, then the vN-M analysis correctly describes rational behavior under uncertainty.

A different, and more controversial, presupposition of Harsanyi's is that we can meaningfully add up different persons' utility functions, *i.e.*, make inter-personal comparisons in terms of a common utility metric.¹⁹ The widely-held view among economists is that adding up different persons' utility functions is theoretically meaningless. We might know that two individuals both prefer A to B, B to C, and C to D. And we might even know (perhaps through extensive observation) that they assign the same relative importance to each of the different possibilities, *e.g.*, that both like A twice as much as B, B three times as much as C, and C five times as much as D. What we can never know is the absolute amount of utility involved, *i.e.*, the scale in which either one's utility ought to be measured. Therefore any attempt to sum the two utility functions is purely a matter of guesswork.

IV

AS INDICATED in the introduction, a very serious difficulty with Harsanyi's theory of justice has been ignored by his critics. It is suggested by Harsanyi himself:

[The vN-M utility criterion] is, however, subject to an important qualification or rather clarification. Common sense does distinguish between sensible and foolish wants or preferences, and it would be absurd to suggest that our moral code should not take account of this distinction in one way or another.²⁰

Harsanyi is not clear as to what he means by "common sense," or precisely how it operates in his theory of justice. Presumably he has in mind something other than the vN-M rationality itself. But why bring "common sense" standards into a theory that is supposed to be based upon vN-M utility? Judging one's

actions on the basis of “common sense” introduces the extraneous criteria that Harsanyi’s approach is meant to avoid.

It is necessarily true (because tautologically true) that one’s actions conform with one’s vN-M utility function, regardless of whether or not one’s preferences seem to violate the standards of common sense. Harsanyi is not, however, questioning this fact. He is raising the more interesting possibility that one’s vN-M utility function could become temporarily distorted, due to “carelessness, rashness, strong excitement, or other conditions which hinder rational choice.”²¹ He insists that such false utility functions be ignored for moral considerations, and that ethical choices be based only on the individual’s “normal” or “true” utility function.²² In this way Harsanyi attempts to preserve the internal consistency of his theory by arguing that we can still define a person’s utility function without reference to any standard outside the person’s own true preferences.

It is not clear how the sympathetic observer is to distinguish “normal” and “true” utility functions from those that are abnormal and false. The only guidance Harsanyi offers is less than adequate:

We may define a person’s “true” preferences as those preferences that this person would manifest on due reflection and in possession of all relevant information, (including information on the pleasures and pains resulting from alternative courses of action), making the moral rules enjoin acts that tend to produce consequences conformable to people’s ‘true’ preferences rather than to their actual explicit preferences (which may not always represent their ‘true’ preferences).²³

Harsanyi is merely begging questions. How much reflection constitutes “due reflection”? How is the sympathetic observer to know what information is “relevant”? Are we to assume the observer has lower information costs than anyone else, and for this reason is likely to make better-informed choices? If “actual explicit” preferences are judged to be false, then is the observer meant to consider “potential implicit” preferences? How? Defining “true” preferences as those which enjoin acts that are conformable to true preferences involves a serious problem of circularity.

It seems impossible, in fact, for the sympathetic observer to avoid imposing independent standards implicitly, perhaps even subconsciously, in every decision made. Leland Yeager addresses this issue in his critique of Harsanyi, but comes to a different conclusion:

It is not nonsensical to imagine myself in the position of another person, with his values and tastes, and consider how my utility would then be affected by some event or set of circumstances (perhaps a distribution of wealth or income). We make such judgments all the time.²⁴

Yeager is missing the point. He suggests that Peter would have no real trouble pretending to be Paul. But Harsanyi suggests there may be many different Pauls—

some who are reasonable and others who are not. So Peter cannot simply pretend to be Paul, he must decide which Paul to pretend to be. It may not be nonsensical to imagine having someone else's values and tastes, however bizarre these may be, but it is nonsensical to imagine one's judging the "correctness" of those values and tastes without appealing to some sort of independent standard.

Moreover, if one is to make an honest personal judgment, those standards have to be one's own. So much for the idea that the hypothetical observer must judge consequences "sympathetically" in the literal sense of the term, *i.e.*, in terms of the attitudes, wants, desires, preferences of these people themselves. . . .²⁵ How is it possible for one honestly to accept as true or rational those values and tastes which violate one's own standards of truth and rationality?

V

ANY PRACTICAL APPLICATION of Harsanyi's theory of justice would inevitably have some people deciding what is good for other people. This would seem to violate Harsanyi's own principle of "sympathetic" value judgments. But the implications go beyond problems of internal inconsistency. Disturbing questions arise as to what Harsanyi's approach would mean for personal liberty and freedom. The fundamental question from the libertarian perspective is this: Does justice demand that one always accept the most rational course of action? Robert Andelson has articulated the libertarian position well in his critique of Rawls.

[A] person is under no moral liability to accede to even the fairest and most reasonable bargain, provided that his refusal to do so does not positively interfere with the freedom of anybody else. King Ahab offered Naboth a good deal for his vineyard—cash, or, if he preferred, a better piece of land in trade. To have accepted this fair proposal would have been the rational thing to do, but Naboth had a right to be irrational.²⁶

In Harsanyi's scheme, when individuals decline "reasonable" prices for their property, would they be forced to sell, on the grounds that their utility functions had obviously become temporarily distorted? Would those who decide not to participate in "obviously advantageous" insurance programs be forced to sign up against their own explicit (but "false") wishes? Would twice-yearly visits to the dentist be mandatory?

In an attempt to hedge against consequences that might arise due to "conditions which hinder rational choice," Harsanyi has compromised the principle of sympathetic value judgments. Unfortunately, such hedging presents its own risks, especially for personal liberty and freedom. Either Harsanyi should accept the consequences of his principles or he should give up his principles. As John Stuart Mill reminds us in a slightly different context:

Strange it is that men should admit the validity of the arguments for free discussion, but object to their being 'pushed to an extreme,' not seeing that unless the reasons are good for an extreme case, they are not good for any case.²⁷

Mill wrote this in an essay aptly entitled "On Liberty." For justice and liberty are inseparable.

Notes

1. J. Harsanyi, "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking," *Journal of Political Economy*, Vol. 61, No. 5 (1953), pp. 434–38; J. Harsanyi, "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy*, Vol. 63, No. 4 (1955), pp. 309–21.
2. J. Harsanyi, *Essays on Ethics, Social Behavior, and Scientific Explanation* (Boston: Reidel, 1976), pp. vii–viii.
3. Paul Samuelson, "Complementarity—An Essay on the 40th Anniversary of the Hicks-Allen Revolution on Demand Theory," *Journal of Economic Literature*, Vol. 12 (1974), pp. 1255–89.
4. J. Harsanyi, *Essays &c. op. cit.* note 2, p. 46.
5. *Ibid.*, p. 45.
6. *Ibid.*, p. 31.
7. *Ibid.*
8. *Ibid.*, p. ix.
9. John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior* (Princeton: Princeton Univ. Press, 1944), especially pp. 15–218.
10. Harsanyi, *Essays &c. op. cit.* note 2, pp. 48–49.
11. M. Friedman, *Price Theory* (New York: Aldine, 1976), p. 35.
12. G. Stigler, *The Theory of Price* (New York: MacMillan, 1946), p. 64.
13. John Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard Univ. Press, 1971), pp. 172 and 323.
14. Harsanyi, *op. cit.* note 2, p. 49.
15. A. Sen, *On Economic Inequality* (Oxford: Clarendon Press, 1973), pp. 18 and 55–65.
16. Harsanyi, *Essays &c. op. cit.* note 2, pp. 64–88.
17. L. Yeager, "Utility and the Social Welfare Function" (unpublished manuscript, Southern Economics Association Meetings, November, 1985), pp. 14–17.
18. J. Marshack, "Rational Behavior, Uncertain Prospects, and Measurable Utility," *Econometrica*, Vol. XVIII (1950), pp. 111–41.
19. Harsanyi, *Essays &c. op. cit.* note 2, pp. 77–80.
20. *Ibid.*, p. 31.
21. *Ibid.*, p. 32.
22. *Ibid.*
23. *Ibid.*
24. Yeager, *op. cit.*, p. 5.
25. Harsanyi, *op. cit.* note 2, p. 32.
26. Robert Andelson, "Vive La Difference? Rawls' 'Difference Principle' and the Fatal Premise upon Which It Rests," *The Personalist*, Vol. 56, No. 2 (1975), p. 210.