Two Problems in the Theory of Fairness

by

Hal R. Varian

Number 163  July, 1975

massachusetts institute of technology

50 memorial drive

cambridge, mass. 02139
Two Problems in the Theory of Fairness

by

Hal R. Varian

Number 163 July, 1975

This is a revised version of an earlier paper of the same title which appeared in July 1974. The current version will be published in the Journal of Public Economics.
Title: Two Problems in the Theory of Fairness

Author: Hal R. Varian

Affiliation: Department of Economics, M.I.T.

Cambridge, Massachusetts 02139

Abstract:

An allocation is called fair iff it is pareto efficient and no agent prefers the bundle of any other agent to his own. This paper briefly discusses the conceptual foundations of this definition and then considers two problems in the theory of fairness. The first involves clarifying the sense in which equal income competitive equilibria are "especially fair", and the second involves a suggestion for using the fairness idea for "second best" comparisons of arbitrary allocations.
Two Problems in the Theory of Fairness*
by Hal R. Varian
July 1974
(Revised July 1975)

An allocation \( x \) is called **equitable** iff no agent prefers another agent's bundle to his own. If at a given allocation some agent does prefer another agent's bundle to his own, we will say that the first agent envies the second. An allocation \( x \) is called **fair** iff it is both equitable and pareto efficient. These and related definitions have been analyzed by several authors, among them Feldman and Kirman (1974), Foley (1967), Kolm (1972), Schmeidler and Vind (1972), Pazner and Schmeidler (1974), and Varian (1974), (1975). There are many interesting relationships between these ideas and the classical concepts of general equilibrium analysis. In this paper, we will consider briefly the conceptual foundation of the notion of equity and then discuss two problems in the theory of fairness. The entire discussion will be limited to the case of a pure exchange economy.

The Concept of Equity

How does the formal definition of equity relate to our intuitive notions of distributive justice? I believe that the fundamental feature of the definition of equity is its emphasis on symmetry: we require each agent to find his own position at least as good as that of any other agent. The nice thing about this particular kind of symmetry is that it is an internal measure, depending specifically on the individual tastes of the agents involved. Thus the theory of fairness, based on symmetry, emphasizes yet another aspect of the concept of justice and can fruitfully
be compared with, and contrasted to, the basically utilitarian theory of welfare economics, the contractual theory of Rawls (1971), and the procedural theory of Nozick (1973).

I have discussed the relationship of fairness to Nozick's theories in Varian (1975), and will consider some relationships to welfare economics in the last section of this paper. Here I want to describe an interesting viewpoint on Rawls's theory that stems from the idea of equity.

Rawls has provided us with a rich theory of justice which involves many concepts. A particular concept that has achieved much attention among economists is the maximin criterion (Rawls (1971), (1974), Arrow (1973), and Alexander (1974)). Rawls suggests that a just allocation is one that maximizes the utility of the worst off individual. This rule has been criticized by several authors, but this is not our purpose here. I am not so concerned with Rawls's justification for the maximin rule, but rather with its operational meaning. How are we to determine who is the worst off individual in some given circumstances? Any such determination would, it seems, require some "external" comparison of utilities, a task which is often quite difficult. It would be nice if we had a version of the maximin rule which involved only an "internal" interpersonal comparison in the way that the equity comparison is internal.

I suggest the following definitions: at any allocation we will define the worst off person as being one whom no one envies; similarly the best off person will be one that envies no one. It can be shown that at any pareto efficient allocation there is someone that no one envies, and someone that envies no one; thus, the best off and the worst
off person are well defined, at least if we limit ourselves to pareto efficient allocations. (See Varian (1974), Theorem 1.1.)

The nice property of these definitions is that they are internal: interpersonal comparisons are being made, but they are being made by the agents involved, and are not being imposed by any external observer. One can now interpret the maximin rule as suggesting that a just allocation is a pareto efficient allocation where the worst off agent — the one that no one envies — is made best off — that is, envies no one. In other words, a just allocation would be one where all agents were unenvious, that is, a fair allocation. In this sense, fair allocations and "maximin" allocations are compatible notions of distributive justice.¹

Fairness and Equal Incomes

It is clear that an equal income competitive equilibrium is a fair allocation. However, the converse is not true; in general there will be many other fair allocations that do not have equal incomes. Nevertheless, it seems somehow that the equal income fair allocation is, in several senses, "especially fair". One sense of this has already been examined in a previous paper: if we require not only that no individual envies another, but also that no group of agents envies the average bundle of any other group, the only fair allocations in large economies are the equal income competitive equilibria. (Varian (1974).)

However, the "coalition-envy" concept is by no means as compelling as the original envy concept. Are there other senses in which equal income allocations are privileged?
Opportunity Fairness

Let us refer to Figure 1 where we have an example of a fair allocation with unequal incomes. Since the allocation is efficient we know that it can be supported by prices \( p \); since neither bundle lies in the preferred set of the other agent the allocation is fair.

If we imagine now that there is a market so that agents can trade their bundles at competitive prices, this allocation loses some of its normative significance. For, even though agent 2 doesn't envy agent 1 directly, he does envy agent 1's possibilities: there are points in agent 1's budget set that agent 2 prefers to his own bundle \( x_2 \).

It is clear that we can define a concept of "opportunity-fairness" which requires that each agent prefers his bundle to any bundle in any other agent's budget set. A moment's reflection will convince us that under an assumption of local nonsatiation of preferences, an allocation is opportunity-fair if and only if it is a competitive equilibrium with equal incomes.

The original notion of fairness still makes sense here: if all we know about the problem are the agents preferences and what are the feasible bundles, all fair allocations seem equally reasonable. If we know more about the problem — for example, what agent's possibilities are under some particular institutional setup — we may well want to restrict the set of fair allocations further.

Robustness of Fair Allocations

Let us refer again to the example in Figure 1. It is well known that if all agents have identical, strictly convex preferences, the unique fair allocation is equal division. Thus, in a sense, it is the differences
in tastes that allow for the existence of these nonequal income fair allocations.

But are these nonequal income allocations really robust against a variety of tastes? Imagine agents being continually born and being assigned randomly to one of the two income classes. Imagine further that the agents have considerable variation in preferences in the following sense: at given prices and income, there is some probability that any point on the budget hyperplane will be demanded by some agent.

Then eventually some agent will be assigned to the higher income level and demand a point lying in the preferred set of one of the lower income agents. If on the other hand an equal income fair allocation were specified, the fairness of the allocation would be robust against adding new agents.

In this sense the equal income allocation is especially fair in that it works no matter what the agents distribution of tastes is, or how that distribution changes. It therefore places many fewer informational demands on the allocative system.

Equal Incomes and a Continuum of Tastes

Some further consideration of Figure 1 offers this analysis: the problem is that there are gaps in the taste patterns of the agents. For it is clear that if an agent has tastes very similar to mine and a higher income, then I will envy his chosen consumption bundle. Now if there were a continuum of agents each similar to his neighbors, then certainly each agent could have no larger income than his neighbors - or else envy would result.
The natural model for this situation is that of a continuum of economic agents. There is an extensive literature on such economies which involves considerable use of measure theory (see for example Aumann (1966) and Hildenbrand (1969)); here I will concentrate on a simpler model emphasizing the topological properties of the continuum. The set of agents is regarded as the open unit interval I. Agent t in I will be endowed with a continuously differentiable strictly concave utility function \( u_t \). We want nearby agents to have similar tastes; I therefore assume that:

1. \( \mathbf{u: \mathbb{R}^k_+ \to \mathbb{R}} \) defined by \( u(t,x) = u_t(x) \) is a continuous function.

Let \( P = \{x \in \mathbb{R}^k_+: w^i - x^i > 0 \text{ for } i = 1, \ldots, k \} \) be the possible consumption set of the agents; here \( w \) is the original bundle to be divided.

**DEFINITION.** A fair allocation is a function \( x: I \to P \) such that

1. \( x \) is pareto efficient
2. \( u_t(x(t)) \geq u_t(x(s)) \) for all \( s, t \) in I

For the purposes of this paper, I will define pareto efficiency in the following way:

**DEFINITION.** An allocation \( x \) is pareto efficient iff it is feasible and there is no set of agents \( A \) and no feasible allocation \( y \) such that

\[ u_t(y(t)) > u_t(x(t)) \text{ for all } t \text{ in } A, \text{ and } \sum_{t \in A} y(t) = \sum_{t \in A} x(t) \text{ (or } \int y(t) = \int x(t), \text{ whichever is appropriate to the cardinality of } A). \]

That is, an allocation is pareto efficient iff there is no set of agents - of any size - that can benefit from trade.
It is intuitively clear that under appropriate assumptions this will require that all marginal rates of substitution between each pair of goods be equal to a common value which we can take to be the competitive price ratio between the two goods. Rather than spell out these conditions precisely, I will simply make the following assumption:

A2. *x* is a competitive equilibrium for some price vector *p* and some income distribution *y*: \( I \rightarrow \mathbb{R} \). That is, for each *t*, \( x(t) \) maximizes \( u_t \) on agent *t*’s budget set, \( B_t = \{z \in P: p'z = y(t) = p'x(t)\} \).

A natural requirement of an equity concept is that equals be treated equally, or, even more, that nearly equals be treated nearly equally. That is, if *x* is a fair allocation and agent *s* is similar to agent *t*, then \( x(s) \) should be close to \( x(t) \). It is reassuring to discover that fair allocations have this property:

**Lemma.** Let \( x: I \rightarrow P \) be a fair allocation; then \( x \) is a continuous function.

Unfortunately, continuous functions can still exhibit quite perverse behavior. To get the next result, we need to specialize a bit more.

**Theorem.** Let \( x: I \rightarrow P \) be a fair allocation which is differentiable; then \( y(t) = p'x(t) \) is a constant function.

In other words, a differentiable fair allocation must award equal incomes to all agents. The differentiability of preferences is an important hypothesis of the above proposition, as it is easy to construct counterexamples if indifference curves have kinks. I do not know if the differentiability of the allocation is equally important.
There is an easy geometrical argument for this proposition. Under the assumptions of the theorem, the fair allocation will be some curve in the commodity space $P$. Consider some point on this curve, say $x(t)$, and consider agent $t$'s indifference curve through this point. This indifference curve cannot cross the allocation curve or else envy would result; hence it is tangent to the allocation curve, and the allocation curve is the envelope of all of the agents' indifference curves.

But remember that the fair allocation is pareto efficient, and therefore each agent's indifference curve is tangent to his budget line. This implies that the allocation line itself is tangent to each agent's budget line; which in turn implies that all agents have equal incomes.

In summary, we have found three reasons why equal income allocations are especially fair: they provide equal opportunities for all agents, they are robust against changes in tastes or adding agents with new tastes, and they are the only kind of fair allocation when tastes vary continuously across many agents.

Second best fairness

The fairness approach to normative economics specifies a globally "best" allocation by appealing to the agents' preferences and to considerations of symmetry. The more classical approach of welfare economics specifies a "best" allocation by maximizing an a priori given welfare function over the set of feasible allocations.

Both of these approaches have their advantages and disadvantages: fairness gives an explicit solution to the problem while the welfare function approach simply pushes the problem back one stage to the specification of
the welfare function; on the other hand the fairness criterion is of no use in comparing the social value of two arbitrary allocations, while the welfare function handles this problem easily. When we use the welfare function approach we have the problem of specifying the form of the function; while in the fairness approach we have the problem extending the fairness criterion to order all alternatives.

Perhaps by combining these two approaches we can achieve a reasonable solution to both problems. Suppose that we limit our choice of welfare functions to those of the linear-in-utility form: \( W(x) = \sum_{i=1}^{n} a_1 u_1(x_1) \). This form is especially simple, and can in fact be axiomatically characterized; see Fishburn (1973), Fleming (1952), Harsanyi (1969) and Vickrey (1969).

Vickrey's characterization is especially interesting for our purposes. Imagine an economic agent in a Rawlsian "original position" of ignorance as to his role in society and even as to his own particular utility function. When faced with this situation an agent might well attempt to choose that distribution of goods that maximizes his "expected utility" - the expectation being taken over the possible cardinal utilities involved in the various outcomes. This formulation of the problem has the advantage of introducing the particular von-Neuman-Morgenstern cardinalization of the utility functions at the outset, as well as leading to a welfare function that is a weighted sum of utilities.

This form does have certain other advantages which make it quite useful; namely, that under certain assumptions:

(1) every maximum of such a function is pareto efficient, and

(2) every pareto efficient allocation is a maximum of such a function for some choice of the weights, \( a_1, \ldots, a_n \).
The first proposition is trivial; the second rests on the fact that under convexity assumptions a necessary and sufficient condition for \( \hat{x} \) to maximize \( W(x) \) on the set of feasible allocations is that there exist a price vector \( \hat{p} \) such that \( a_i Du_i(\hat{x}) - \hat{p} = 0 \). But since \( \hat{x} \) is pareto efficient, each agent is maximized on his budget set so \( Du_i(\hat{x}_i) = \lambda_i \hat{p} \). Choosing \( a_i = \frac{1}{\lambda_i} \) then shows \( \hat{x} \) maximizes \( W(x) \). (For a more detailed proof, the reader should consult Negishi (1960).)

We still have the problem of how to choose the weights. If we adopt the view of fairness, a fair allocation is the "best of all possible allocations". Hence we would want to require that \( W(x) \) reach a global maximum at a fair allocation. We have already seen that an equal income competitive allocation is necessarily fair, and, furthermore, this type of allocation is especially fair in several important senses.

Since an equal income competitive allocation is efficient, it maximizes \( W(x) = \sum_{i=1}^{n} a_i u_i(x_i) \) for some set of \( (a_i) \). We therefore normalize our welfare function by choosing that set of \( (a_i) \) as weights. The resulting welfare function can be used to make comparisons among arbitrary allocations.

What is the economic interpretation of these weights? The above argument tells us that these weights are the inverses of the marginal utilities of income evaluated at the equilibrium values. Thus:

\[
a_i = \frac{1}{\text{D}_y u_i(x_i(p^*, y^*))} = \frac{1}{\lambda_i^*}
\]

where

- \( x_i = \text{ith agents demand function} \)
- \( p^* = \text{equilibrium vector of prices} \)
- \( y^* = \text{equilibrium income; equal for all agents} \)
Notice that the natural units of $a_i$ are in dollars/util. The welfare evaluation of an arbitrary allocation is then measured in dollars: our criterion leads us to evaluate utilities at any allocation in terms of dollar evaluations that are appropriate at a "utopian" position; in this sense, it is a kind of "consumer surplus" measure.

This welfare function has a property which is very convenient for evaluating small changes around the fair allocation: suppose we are currently at the fair allocation $(x,p)$ and we are considering undertaking a small new project which will result in a new allocation, $z$. Then this project increases social welfare if and only if $\sum_{i=1}^{n} p_i x_i < \sum_{i=1}^{n} p_i z_i$; i.e., if and only if national income is greater when evaluated at the current prices.

We can verify this proposition by expanding the welfare function in a Taylor series around the original allocation $x$:

$$W(z) - W(x) = \sum_{i=1}^{n} \frac{u_i(z_i)}{\lambda_i} - \sum_{i=1}^{n} \frac{u_i(x_i)}{\lambda_i} = \sum_{i=1}^{n} \frac{Du_i(x_i)}{\lambda_i} \cdot (z_i - x_i) + o(z - x)$$

But since the original allocation is an equal income competitive equilibrium, each agent is maximized on his budget set and therefore $Du_i(x_i) = \lambda_i p_i$. Inserting this into the above equation we get

$$W(z) - W(x) = \sum_{i=1}^{n} \frac{u_i(z_i)}{\lambda_i} - \sum_{i=1}^{n} \frac{u_i(x_i)}{\lambda_i} = \sum_{i=1}^{n} p_i (z_i - x_i) + o(z - x)$$

which says that for small changes aggregate welfare is greater if and only if national income is greater.

Of course this property follows from the assumption of efficiency, not the assumption of equity. At any efficient allocation, the economy can be viewed as implicitly maximizing a weighted sum of utilities,
with poorer agents - those with higher marginal utilities of income - receiving less weight. Any small project that increases national income increases this implicit welfare function.

But at an arbitrary nonfair allocation the implicit welfare function may not be the right welfare function. An allocation that increases national income may well decrease a welfare function normalized on a fair allocation. If the utopian goal of society is equal incomes then cost-benefit studies must be done with care - the appropriate criterion is not whether benefits exceed costs but whether weighted benefits exceed weighted costs.

What are the defects of this approach? It certainly has an appeal when \( x \) is close to \( x^* \); however, when \( x \) is far from \( x^* \), the weights chosen may appear to be somewhat arbitrary. Secondly, if the endowment of the economy changes, the welfare function may change. Of course if preferences are homothetic and the endowment changes in a balanced way, the welfare function will remain constant. Thirdly, although the welfare function reaches a maximum at fair allocations, it may also reach a maximum at very unfair allocations.

For example, suppose that utility functions are concave, homogenous of degree one, and identical. In a pure exchange economy, the unique fair allocation will then be equal division; that is \( x^*_i = \frac{w_i}{n} \) for \( i = 1, \ldots, n \). In this case we can normalize all the \( \lambda_i \)'s to be equal to 1. For \( n = 2 \), and for arbitrary \( 1 > t > 0 \), we have:

\[
\begin{align*}
&u_1(x^*_1) + u_2(x^*_2) = tu_1(x^*_1) + (1 - t)u_2(x^*_2) = u_1(tx^*_1) + u_2((1 - t)x^*_2) \\
&= u_1(x^*_1) + u_2(x^*_2)
\end{align*}
\]
Thus there are very many nonegalitarian distributions which also achieve maximum welfare; the linear in utility welfare function ignores the distribution of utility. One way to remedy this situation is to consider a social welfare function of the following form:

$$V(x) = \sum_{i=1}^{n} a_i u_i(x_i) - b \sum_{i=1}^{n} \sum_{j=1}^{n} a_i (u_i(x_j) - u_i(x_i)) \delta_{ij}$$

where \((a_i)\) and \(b\) are parameters and \(\delta_{ij}\) is one when the preceding term is positive and zero otherwise.

This welfare function is actually much more in the spirit of fairness than the preceding one. The first term can be seen as expressing a measure of efficiency, while the second expression measures the total "envy" in the distribution. This method of introducing distributional characteristics into the welfare function is more appealing than measures of variance of income, etc., since it measures the distribution in utility terms.

Let \(x^*\) be the equal income fair allocation. Then a natural choice for \(a_i\) is again, \(\frac{1}{\lambda_i^*}\). This also ensures that \(V\) reaches its global maximum at such a point. For let \(y\) be any other feasible distribution; then

$$\sum_{i=1}^{n} \frac{1}{\lambda_i^*} u_i(x_i^*) \geq \sum_{i=1}^{n} \frac{1}{\lambda_i^*} u_i(y_i)$$

since \(W(x)\) is maximized at \(x^*\),

and

$$b \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{\lambda_i^*} (u_i(x_j^*) - u_i(x_i^*)) \delta_{ij} \leq b \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{\lambda_i^*} (u_i(y_j) - u_i(y_i)) \delta_{ij}$$

since the envy term is 0 at the fair allocation and non-negative elsewhere.

In particular, welfare will be strictly lower than \(V(x^*)\) at any nonequitable allocation.
The welfare function $V(x)$ is a function of a non-Bergsonian variety; that is, $V(x)$ is of the form $F(u_i(x_j))$ rather than $F(u_i(x_j))$. The welfare of a nonefficient allocation may well be higher than that of pareto dominating allocations; this type of function thus reveals the often mentioned but seldom analysed "equity-efficiency tradeoff".

Figure 1
Appendix

Equal Incomes and a Continuum of Tastes

**Lemma.** Let \( f: X \to Y \) be a function from a metric space to a compact metric space. Let \((x_i)\) be a sequence in \(X\) converging to \(x_0\), and let \((f(x_i))\) be the corresponding sequence in \(Y\). If every convergent subsequence of \((f(x_i))\) converges to \(f(x_0)\), then the function \(f\) is continuous at \(x_0\).

**Proof.** This proposition is well known, but it is easier to prove it than to find a suitable reference.

Suppose that \(f\) is not continuous at \(x_0\). Then there is some neighborhood of \(f(x_0), N\), such that the sequence \((f(x_i))\) does not remain in \(N\) as \(i\) goes to infinity. It is therefore possible to choose a subsequence of \((f(x_i))\) that converges to some point \(y\) which is not equal to \(f(x_0)\). \(\square\)

**Lemma.** Let \(x: I \to P\) be a fair allocation; then \(x\) is a continuous function.

**Proof.** Let \((t_i)\) be a sequence in \(I\) converging to \(t_0\), and let \((x(t_i))\) be the corresponding sequence in \(P\). We will show that every convergent subsequence of \((x(t_i))\) converges to \(x(t_0)\).

We first note that by compactness \((x(t_i))\) has a convergent subsequence that converges to some point \(x^*\) in the closure of \(P\).

Suppose that \(u_{t_0}^{x^*}(x^*) > u_{t_0}^{x(t_0)}(x(t_0))\). Then for \(x(t_i)\) close to \(x^*\), \(u_{t_0}^{x(t_i)}(x(t_i)) > u_{t_0}^{x(t_0)}(x(t_0))\), contradicting equity. Similarly, suppose
that \( u_{t_0}^*(x^*) < u_{t_0}^*(x(t_0)) \). Then for \( t_1 \) close to \( t_0 \) and \( x(t_1) \) close to \( x^* \), we have \( u_{t_1}^*(x(t_1)) < u_{t_1}^*(x(t_0)) \) which again contradicts equity. Therefore \( u_{t_0}^*(x^*) = u_{t_0}^*(x(t_0)) \).

Now if \( x^* \) is not equal to \( x(t_0) \), strict concavity of utility implies that \( u_{t_0}^*(\frac{1}{2} x^* + \frac{1}{2} x(t_0)) > u_{t_0}^*(x^*) = u_{t_0}^*(x(t_0)) \). Therefore for \( t_1 \) close to \( t_0 \), and for \( x(t_1) \) close to \( x^* \)

\[
u_{t_0}^*\left(\frac{1}{2} x(t_1) + \frac{1}{2} x(t_0)\right) > u_{t_0}^*(x(t_0))
\]

Since \( u \) is continuous in \( t \) we also have:

\[
u_{t_1}^*\left(\frac{1}{2} x(t_1) + \frac{1}{2} x(t_0)\right) > u_{t_1}^*(x(t_1))
\]

But these two inequalities contradict the hypothesis of pareto efficiency. Therefore \( x^* = x(t_0) \) and \( x \) is a continuous function. \( \square \)

**THEOREM.** Let \( x: I \rightarrow P \) be a fair allocation which is differentiable; then \( y(t) = p \cdot x(t) \) is a constant function.

**Proof.** Choose an agent \( t_0 \) and define a differentiable function \( v: I \rightarrow R \) by \( v(t) = u_{t_0}^*(x(t)) \). Since \( x \) is a fair allocation, \( v(t) \) reaches a maximum at \( t_0 \) and therefore its derivative must vanish. This implies

\[
Dv(t_0) = Du_{t_0}^*(x(t_0)) = Du_{t_0}^*(x(t_0)) \cdot Dx(t_0) = 0.
\] (1)

But \( x \) is a competitive equilibrium, so agent \( t_0 \) is maximized on his budget set, which means

\[
D_x u_{t_0}^*(x(t_0)) = \lambda_{t_0} p
\] (2)
Putting expressions (1) and (2) together we get
\[ p \cdot D_t x(t_0) = 0. \]  

(3)

Consider the income function \( y \) defined above, and use (3) to calculate its derivative:
\[ D_t y(t_0) = D_t [p \cdot x(t_0)] = p \cdot D_t x(t_0) = 0 \]

The choice of \( t_0 \) was arbitrary, which implies \( y \) is a differentiable function with everywhere zero derivative. It must therefore be a constant function. □
Footnotes

* I wish to thank Marty Weitzman and Kenneth Arrow for several helpful suggestions on these issues. Work supported in part by the Urban Institute grant, "Efficiency in Decision Making".

1. Pazner and Schmeidler (1973) suggest another way in which fairness and the maximin rule are compatible.

2. Feldman and Kirman (1974) have used the fairness idea to make second best comparisons in a quite different way.
References


