

The Methodology of Normative Economics

by

Steven E. Landsburg
University of Rochester

When economists distinguish between positive and normative issues, they traditionally add the caveat that economics has more to say about the former than the latter. While positive questions can be settled in principle by pure science, normative questions can be settled only by the introduction of values that come from outside any economic model. We can predict the consequences of a social planner's behavior, but we are silent about what he ought to maximize.

This paper argues otherwise. In the context of a very simple model, I will show that the social planner's objective function must arise endogenously, and that there is a strong argument that the endogenously determined objective function is the only "right" one. The reason for this is that is literally impossible for the planner to maximize anything else.

An extreme example will illustrate the flavor of the argument. Suppose that the planner's goal is "equity at any cost", in consequence of which he sets out to increase the welfare of the least well-off member of society. Having located that unfortunate soul, the planner asks what he can do to make him happier. The unfortunate soul replies that his greatest desire is to live in a world where social planners do not seek equity.

In this example, the only way for the planner to achieve his goal is: To abandon his goal. Despite the example's flamboyance, it illustrates a quite general phenomenon. If people care what the planner is up to, the rule for maximizing almost any social welfare function will entail switching to a different social welfare function. Typically, only a finite number of social welfare functions escape this paradox. In fortuitous circumstances, that finite number is one.

The key new assumption in all of this is that people care about what kind of society they live in, independent of its effect on their material well-being. Of course, they care

also about their consumption, which is affected by the planner's behavior. Therefore his behavior should enter the utility function in two ways: Directly, as an argument in its own right, and indirectly, through its effect on consumption.

1. Normative Economics

For concreteness, I will work within the welfarist tradition associated with Bergson (1938) and Samuelson (1947); that is, I will imagine a planner who chooses a social welfare function and I will model normative disputes as disagreements about the relative merits of various candidates for that welfare function. Of course, the social choice literature proceeding from Arrow (1950) and Vickrey (1960) allows us to imagine a more general setup where the planner is permitted to use mechanisms other than the maximization of a Bergson-Samuelson welfare function; presumably the ideas of the present paper can be generalized to such a context, but I have not carried out that exercise.

As eloquently discussed by Sen (1977) and others, the welfarist framework is not perfectly general; for example it appears not to allow for a doctrine of natural rights. Other limitations are associated with the interpersonal comparison of utilities; once again see Sen. But the welfarist setup is still adequate for a first pass at many of the issues (such as equity/efficiency trade-offs) that economists typically think of when they hear the word "normative".

That said, I will proceed to the model.

There are n individuals, with utility functions u^1, u^2, \dots, u^n . These utility functions might or might not have cardinal significance. They take as argument a vector $x = (x^1, \dots, x^p)$ which is permitted to range over some compact convex subset of some \mathbf{R}^p . The vector x typically has some *physical* interpretation, most commonly as an allocation of goods across individuals. In the simplest models, individual i cares only about those components of the vector x that affect his own consumption; in models allowing for altruism or envy, every individual potentially cares about all of the components.

Now introduce a social planner who must choose a vector x to maximize a social

welfare function $F(u^1, \dots, u^n)$. F can take any of a variety of forms, and I take it that much disagreement about the nature of the “just society” can be reduced to disagreement about the appropriate choice of the function F . A classical utilitarian (with faith in the meaningfulness of cardinal utility) might call for F to take the form

$$F(u^1, \dots, u^n) = \sum_{i=1}^n u^i$$

or more generally

$$F(u^1, \dots, u^n) = \sum_{i=1}^n \alpha_i u^i$$

where the α_i are weights chosen according to some (here unspecified) criterion. The strong advocate for equity urges us to take

$$F(u^1, \dots, u^n) = \min_i \{u^i\}.$$

(See Sen (1977) for extensive discussion of some of the philosophical baggage associated with such examples.)

At this point, the game is to marshal arguments for one F over another, and economists traditionally abandon the playing field to philosophers and other pundits (though not without reserving their right to coach from the sidelines). Substantial resources are devoted to these arguments in academic seminars, on newspaper editorial pages, and around the lunch table. From this we may infer that people care about the outcome; that is, people have preferences regarding the choice of F .¹ But if this is so, then the original problem must be misspecified, because F does not appear as an argument in any of the utility functions. To put this another way, the basic model, which asserts that only material outcomes (that is, x 's) matter, predicts that nobody will ever want to study the questions that the model is designed to address. The model predicts its own non-existence.²

¹ A possible objection is that preferences regarding F are not primary but derived from preferences regarding outcomes; I will address this objection in the concluding section.

² Exactly the same point was raised by Hahn (1982), in a paper that I discovered after the present paper was written. As I shall do in the remainder of this paper, Hahn proposed an individual utility function of the form $u^i = u^i(x, F)$. He described the dependence of u^i on F as a “difficulty” for traditional welfare economics. One contribution of the present paper is the resolution of that difficulty.

I propose to address the paradox by modifying the model. There remain n individuals with utility functions u^1, \dots, u^n . The allocation of resources is represented by a variable x which is typically allowed to range over some subset $\mathcal{X} \subset \mathbf{R}^p$. The social welfare function is denoted $F(u^1, \dots, u^n)$, and is chosen from some space \mathcal{S} of allowable functions. Each u^i can be written as a function of two variables

$$u^i = u^i(x, F).$$

(Warning: It is not the *value* of F but the entire *function* F that enters as an argument to u^i .) Given F , the planner chooses $(x, G) \in X \times \mathcal{S}$ to maximize

$$F(u^1(x, G), u^2(x, G), \dots, u^n(x, G)).$$

We say that F is *self-justifying* if there exists an $x \in X$ such that (x, F) is a solution to the maximization problem just described. Unless F is self-justifying, it is impossible for the planner to behave in a way that maximizes F .

2. Existence.

In this section, I will show that with some additional restrictions, there always exists a self-justifying welfare function.

There remain n agents indexed by $\{1, \dots, n\}$. Social welfare functions are parameterized by points in a given compact convex subset $\mathcal{S} \subset \mathbf{R}^m$, which is assumed to be the closure of its interior. A typical point is denoted $(\alpha^1, \dots, \alpha^m)$. Material outcomes (e.g. the allocation of some collection of goods) are parameterized by points in a given compact convex subset $\mathcal{X} \subset \mathbf{R}^p$, assumed to be the closure of its interior, with a typical point being denoted (x^1, \dots, x^p) . Agent i has the differentiable concave utility function

$$u^i(x^1, \dots, x^p; \alpha^1, \dots, \alpha^m).$$

The family of welfare functions is given by a differentiable function

$$\mathcal{F}(u^1, \dots, u^n; \alpha^1, \dots, \alpha^m).$$

For any fixed $\alpha = (\alpha^1, \dots, \alpha^m)$, the function

$$F_\alpha(u^1, \dots, u^n) = \mathcal{F}(u^1, \dots, u^n; \alpha^1, \dots, \alpha^m)$$

is concave and increasing in each variable.

Given α , the planner chooses $x(\alpha) = (x^1(\alpha), \dots, x^r(\alpha))$ and $\beta(\alpha) = (\beta^1(\alpha), \dots, \beta^m(\alpha))$ to maximize

$$F_\alpha(u^1(x; \beta), \dots, u^n(x, \beta))$$

and we say that F_α is *self-justifying* if for some value of the (possibly multiple-valued) function β , we have $\beta(\alpha) = \alpha$.

Theorem 1. *With the above hypotheses, a self-justifying welfare function exists.*

Proof. By Klein and Thompson (1984), 9.2.6, the correspondence $\alpha \mapsto \beta(\alpha)$ is both compact-valued and upper semi-continuous. It follows by standard arguments (e.g. as in Browder (1968)) that the correspondence has a closed graph. The concavity of each F_α easily implies that the correspondence is convex-valued also. This enables us to invoke Kakutani's fixed-point theorem in the form given by Klein and Thompson (1984), 8.2.5, and conclude that $\alpha \mapsto \beta(\alpha)$ has a fixed point.

q.e.d.

3. Uniqueness: Preliminaries

Uniqueness is a more delicate question than existence and requires additional hypotheses. In this section I will state those hypotheses and explain their significance.

I continue to use the notation and hypotheses of the preceding section. For convenience in stating the argument, I will take each F_α to be not just concave but *strictly* concave, so that the correspondence $\alpha \mapsto \beta(\alpha)$ is single-valued. I make the following additional hypotheses:

Hypothesis 0. *The matrix of partials*

$$\begin{pmatrix} \frac{\partial u^i}{\partial x^j} \end{pmatrix}$$

is everywhere of rank at least $n - 1$.

This requires in particular that the space from which the vector x is drawn have dimension at least $n - 1$. This is not hard to satisfy. If x represents the allocation of k goods among n agents, then the space has dimension exactly $k \cdot (n - 1)$.

Hypothesis 1. *For each i, j , and k , we have*

$$\frac{\partial u^i}{\partial x^j \partial \alpha^k} = 0.$$

More generally, it is enough to assume that u and v can be brought to additively separable form by monotonic transformations.

To interpret Hypothesis 1, note that agents care about the choice of social welfare function for two reasons: First because it affects the planner's choice of x (representing a material outcome) and second, for purely philosophical reasons. Hypothesis 1 states that the purely philosophical components of agents' preferences are independent of the current allocation of goods. The hypothesis can be viewed as formalizing the requirement that those preferences have genuine ethical content.

Hypothesis 2. *Let Q be the matrix given by*

$$Q_{ij} = \frac{\partial^2 \mathcal{F}}{\partial u^j \partial \alpha^i}.$$

Let \hat{U} be the matrix given by

$$\hat{U}_{ij} = \frac{\partial u^i}{\partial \alpha^j}.$$

Then the matrix $Q \cdot \hat{U}$ is negative semi-definite at every self-justifying optimum.

(I say that an $m \times m$ matrix A is negative definite if the quadratic form on \mathbf{R}^m given by $x \mapsto x^T A x$ is negative definite. In particular, there is no implication that A is symmetric.)

Hypothesis 2 is a formal way to require that there be substantial tension between agents' philosophical predilections (as mirrored by the partials of their utility functions

with respect to the α 's) and their selfish interests (as mirrored by the partials of their utility functions with respect to the x 's).

To make this clearer, consider an example where two agents with utility functions u and v must divide one unit of a single consumption good. The outcome is summarized by a single number $x \in [0, 1]$ representing the fraction of the good that is allocated to the first individual. The planner seeks to maximize a concave differentiable welfare function F_α where α is a parameter in $[0, 1]$. We write $u = u(x, \alpha)$, $v = v(x, \alpha)$, and $\mathcal{F}(u, v, \alpha) = F_\alpha(u, v)$. Then Hypothesis 2 reduces to the requirement that

$$\frac{\partial \mathcal{F}_1}{\partial \alpha} \frac{\partial u}{\partial \alpha} + \frac{\partial \mathcal{F}_2}{\partial \alpha} \frac{\partial v}{\partial \alpha} \leq 0 \quad (1)$$

at self-justifying optima, where \mathcal{F}_1 and \mathcal{F}_2 are partial derivatives with respect to u and v .

To help interpret equation (1), note that at a self-justifying optimum, the planner must be solving the equation

$$\mathcal{F}_1 \frac{\partial u}{\partial \alpha} + \mathcal{F}_2 \frac{\partial v}{\partial \alpha} = 0. \quad (2)$$

Using (2), the quotient rule for differentiation, and the observation that $\mathcal{F}_1 > 0$, we can rewrite (1) as

$$\frac{\partial v}{\partial \alpha} \cdot \frac{\partial(\mathcal{F}_2/\mathcal{F}_1)}{\partial \alpha} \leq 0 \quad (1')$$

which is therefore equivalent to Hypothesis 2 in this example.

To interpret (1'), note that the term $\frac{\partial v}{\partial \alpha}$ is positive if and only if the second agent's philosophy argues for an increase in α , whereas (speaking a bit loosely) the term $\frac{\partial(\mathcal{F}_2/\mathcal{F}_1)}{\partial \alpha}$ is positive if and only if an increase in α would lead the planner to give the second agent's utility more relative weight in the social welfare function. Thus Hypothesis 2 says that at a self-justifying optimum, the second agent's philosophy argues for a change in social policy that would cause his own utility to receive less weight.

Example 1. Suppose that

$$F_\alpha(u, v) = (1 - \alpha) \cdot u + \alpha \cdot v.$$

Then

$$\frac{\partial(\mathcal{F}_2/\mathcal{F}_1)}{\partial \alpha} = \frac{\partial}{\partial \alpha} \left(\frac{\alpha}{1 - \alpha} \right) = \frac{1}{(1 - \alpha)^2}$$

is unambiguously positive, so Hypothesis 2 requires $\frac{\partial v}{\partial \alpha} < 0$, i.e. the purely philosophical component of the second agent's preferences argues for a decrease in α —which is to say that the second agent believes that he deserves less weight than he is getting in the social welfare function. It is easy to show (using (2)) that this in turn is equivalent to first agent believing that *he* deserves less weight than he is getting.

Example 2. Suppose that

$$F_\alpha(u, v) = \begin{cases} (u^\alpha + v^\alpha) & \text{if } \alpha \in (0, 1]; \\ \log(u) + \log(v) & \text{if } \alpha = 0. \end{cases}$$

(Thus the classical utilitarian prefers high values of α , and the egalitarian prefers low values.) Then

$$\frac{\partial(\mathcal{F}_2/\mathcal{F}_1)}{\partial \alpha} = \log\left(\frac{v}{u}\right) \cdot \left(\frac{u}{v}\right)^{1-\alpha}$$

which is positive when and only when $v > u$. Now equation (1'), and hence also Hypothesis 2, is equivalent to the following statement: At any self-justifying optimum, $\frac{\partial v}{\partial \alpha} < 0$ if and only if $v > u$. (It is also equivalent to the symmetric statement that results when u and v are reversed.) Thus in this example Hypothesis 2 can be interpreted as follows: At any self-justifying optimum, whichever agent has the higher utility also has a philosophical preference for more equity—or, more imprecisely, that each agent believes he is better off than he deserves to be.

Remark 1. It is clear that something like Hypothesis 2 is necessary to generate a uniqueness theorem. If everyone's philosophical preferences tended to reinforce his selfish interests, then many welfare functions would be self-justifying. In the most extreme case, suppose that each individual believes on purely philosophical grounds that only his own utility should count in the social welfare function. (Call this the “solipsistic case”.) Then it is self-justifying to choose an arbitrary agent and make him dictator. There has to be tension between agents' philosophical beliefs and their material self-interest if there is to be any hope of uniqueness.

Remark 2. The use of Hypothesis 2 to generate a uniqueness theorem is a precise echo of a vague theme from the philosophical literature. Beginning with the assumption of a

conflict between moral judgment and individual self-interest, Kant (1785, 1788) concluded the existence of one universal moral law.

Remark 3. Suppose that $H(Z, \alpha)$ is a differentiable function that is monotonically increasing in Z for each α . Then the planner's problem remains unchanged if the function $\mathcal{F}(u, \alpha)$ is replaced by

$$\mathcal{G}(u, \alpha) = H(\mathcal{F}(u, \alpha), \alpha).$$

As one would hope, Hypothesis 2 is invariant under such transformations.

4. Uniqueness: Results

In this section I will prove two uniqueness theorems. Recall that n is the number of agents.

Theorem 2. *Suppose Hypotheses 1 and 2 hold and that $n = 2$. (In this case, Hypothesis 0 holds automatically.) Then the self-justifying social welfare function is unique.*

Theorem 3. *Suppose that Hypotheses 0, 1 and 2 hold and that $\partial^2 \mathcal{F} / \partial u^i \partial u^j = 0$ for all $i \neq j$. (In other words, the planner maximizes a weighted sum of utilities and the problem is to pick the weights.) Then the self-justifying social welfare function (that is, the self-justifying choice of weights) is unique.*

I will prove both theorems simultaneously, via a method which can be employed to generate additional uniqueness theorems of a similar flavor. First we need a considerable amount of notation.

Let

$$\mathcal{F}_i = \frac{\partial \mathcal{F}}{\partial u^i} \quad \hat{\mathcal{F}}_i = \frac{\partial \mathcal{F}}{\partial \alpha^i} \quad u_j^i = \frac{\partial u^i}{\partial x^j} \quad \hat{u}_j^i = \frac{\partial u^i}{\partial \alpha^j}$$

$$u_{jk}^i = \frac{\partial u^i}{\partial x^j \partial x^k} \quad \hat{u}_{jk}^i = \frac{\partial u^i}{\partial \alpha^j \partial \alpha^k}$$

In terms of this notation, the first-order conditions for the planner's problem are

$$\sum_{i=1}^n \mathcal{F}_i u_s^i = 0 \quad \text{for all } s \in \{1, \dots, r\} \quad (3)$$

$$\sum_{i=1}^n \mathcal{F}_i \hat{u}_t^i = 0 \quad \text{for all } t \in \{1, \dots, m\} \quad (4)$$

Now let capital letters stand for matrices, and define a set of matrices by the following equations:

$$H_{ij} = \mathcal{F}_{ij}.$$

$$U_{ij} = u_j^i.$$

$$\hat{U}_{ij} = \hat{u}_j^i.$$

$$Q_{ij} = \frac{\partial \mathcal{F}_i}{\partial \alpha_j}.$$

$$D_{ij} = \sum_{k=1}^n \mathcal{F}_k u_{ij}^k.$$

$$E_{ij} = \sum_{k=1}^n \mathcal{F}_k \hat{u}_{ij}^k.$$

Finally, set

$$A = U^T H U.$$

Now view x^i and β^j as functions of the α 's, chosen to maximize $\mathcal{F}(u^1(x; \beta), \dots, u^n(x; \beta); \alpha)$

Lemma. *If Hypotheses 0 and 1 hold, then*

$$\left(\frac{\partial x^i}{\partial \alpha^j}\right) = -Z^{-1}U^T Q^T \quad \left(\frac{\partial \beta^i}{\partial \alpha^j}\right) = -E^{-1}\Theta^T D Z^{-1}U^T Q^T \quad (5)$$

where Θ is a matrix such that

$$U \cdot \Theta = \hat{U} \quad (6)$$

and where

$$Z = A + D + A\Theta E^{-1}\Theta^T D. \quad (7)$$

Proof. Hypothesis (0), together with equations (3) and (4), guarantees the existence of Θ as in (6).

Differentiating (3) and (4) with respect to each α^j and taking account of Hypothesis 1 gives a matrix equation

$$\begin{pmatrix} A + D & A\Theta \\ \Theta^T A & \Theta^T A\Theta + E \end{pmatrix} \cdot \begin{pmatrix} (\partial x^i / \partial \alpha^j) \\ (\partial \beta^i / \partial \alpha^j) \end{pmatrix} = \begin{pmatrix} -U^T Q^T \\ -\Theta^T U^T Q^T \end{pmatrix}.$$

By inspection, one verifies that this equation is solved by (5).

q.e.d.

The self-justifying optimum will be unique under any hypotheses that suffice to guarantee that the matrix $(\partial \beta^i / \partial \alpha^j)$ is negative semi-definite. We will apply this observation to prove Theorems 2 and 3. The ambitious reader can apply it to prove additional theorems of a similar flavor.

First let $\mathcal{S} \subset \mathbf{R}^m$ be the space from which the vector $\alpha = (\alpha^1, \dots, \alpha^m)$ is chosen. Note that if there are two distinct self-justifying optima α_1 and α_2 then we can replace the space \mathcal{S} with a line segment through α_1 and α_2 to generate a new counterexample in which $m = 1$. Thus we may assume that $m = 1$ to begin with.

It is tempting to apply a similar argument to reduce the dimension of the space $\mathcal{X} \subset \mathbf{R}^p$ from which x is drawn. Unfortunately, this works only when $n = 2$. For larger n ,

the attempted reduction would destroy Hypothesis 0.

Proof of Theorem 2. In this case we can reduce to the case $m = 1$, $p = 1$ as in the preceding two paragraphs. Then A , D , E and therefore Z (as defined in (7)) are all negative scalars. Then (5) reveals that $(\partial\beta/\partial\alpha)$ is a scalar with the same sign as $\Theta^T U^T Q^T = \hat{U}^T Q^T$, which is non-positive by Hypothesis 2.

q.e.d.

Proof of Theorem 3. In this case we can reduce to the case $m = 1$ as above (but *not* to the case $p = 1$). Then $A = 0$ and E is a negative scalar so (6) gives $Z = D$ and (5) reveals that $(\partial\beta^i/\partial\alpha^j)$ is a positive scalar multiple of $\Theta^T U^T Q^T = \hat{U}^T Q^T$, which is negative semi-definite by Hypothesis 2.

q.e.d.

5. On the Necessity of Philosophical Disagreements

Consider for simplicity the case where potential social welfare functions are parameterized by a single (one-dimensional) parameter α . Then the first-order condition (4) can be rewritten as

$$\sum_{i=1}^n \mathcal{F}_i \hat{u}^i = 0 \quad (4')$$

where $\hat{u}^i = \partial u^i / \partial \alpha$.

Because the \mathcal{F}_i are all positive, it follows from (4') that at a self-justifying optimum the various \hat{u}^i cannot all be of the same sign (unless all are zero). Although this observation is a mathematical triviality, its interpretation bears comment.

If the planner adjusts the value of α by a small amount $d\alpha$, then the i^{th} consumer's utility changes by the amount

$$\left[\left(\sum_{j=1}^p u_j^i \frac{\partial x^j}{\partial \alpha} \right) + \hat{u}^i \frac{\partial \beta}{\partial \alpha} \right] d\alpha. \quad (8)$$

Within the square brackets, we can interpret the first term as the change in utility resulting from considerations of material self-interest, and the second term as the change in utility resulting from considerations of pure philosophy.

It is clear that at a social optimum, (8) cannot be simultaneously positive (or negative) for all i ; if that were the case, then the planner could make a Pareto improvement by increasing (or decreasing) α .

But the implication of (4') is that it is *also* impossible for just the second term inside the square brackets of (8) to be simultaneously positive (or negative) for all i . This means that in evaluating the direction in which α should be adjusted, people must disagree among themselves for reasons that are rooted in pure philosophy.

To illuminate this further, consider Example 2 of Section 3 above, where classical utilitarians seek high values of α and egalitarians seek low values of α . One can easily imagine the the first agent (Jack) arguing as follows: "On purely philosophical grounds, I am a utilitarian and would like to live in a world where planners maximize sums of utilities. However, a move in that direction would reduce my own consumption. On balance, I would rather we not move in that direction".

At the same time, the second agent (Jill) is gung-ho for utilitarianism on *both* philosophical grounds *and* grounds of material self-interest. However, the planner, taking account of Jack's strong material objection to utilitarianism, considers the situation optimal as it stands. Even though Jack and Jill agree that more utilitarianism would be philosophically desirable, Jack's strong material objections weigh so heavily that the planner decides to stick with the *status quo*.

The point is that this plausible-sounding scenario is impossible. Unless the planner's optimum is at an extreme point (so the first-order conditions are irrelevant) or Jack and Jill are in perfect agreement about the ideal value of α (so that $\partial u/\partial\alpha$ and $\partial v/\partial\alpha$ can be simultaneously zero) the scenario is ruled out by the observation that began this section. If Jack is philosophically committed to more utilitarianism ($\hat{u} = \partial u/\partial\alpha > 0$) then Jill must be philosophically committed to more equity ($\hat{v} = \partial v/\partial\alpha < 0$).

Similar considerations apply when α is multi-dimensional. A change in any direction must be supported by some agents and opposed by others, and that would continue to be the case even if agents could somehow discipline themselves to ignore their material self-

interest (represented by the first bracketed term in (8)). In other words, at a self-justifying optimum, purely philosophical disagreements are inevitable.

6. Concluding Remarks.

In the standard approach to normative economics, the planner's choice of a welfare function F certainly affects individual utility, but it does so indirectly. The decision to maximize one F rather than another leads to a different allocation of resources, and this allocation does enter into the utility functions. By contrast, the models in this paper assume that the people care what the planner is up to, not just through its effect on allocation, but also *directly*; they care not only about the allocation but about how it is reached. Although the associated formalism is simple, its consequences are drastic; the choice of F is largely removed from the planner's discretion.

A natural question to ask at this point is: *What evidence is there that people care about the social planner's objective function except insofar as it affects material allocations?* The most apparent evidence is the energy that economists and others devote to studying such questions, revealing that somebody (either the researchers themselves or those who fund their research) cares about the answers. It would be possible to study different objective functions solely from the viewpoint of their material consequences. But we do not restrict ourselves to this viewpoint. We devote resources to justifying one social structure over another on *moral* grounds.

It is no use to respond that these moral arguments are mere smokescreens that disguise the real material motivations of those who fund our investigations. For why should such smokescreens be at all effective unless there is an audience that cares about moral issues? The "smokescreen" response denies the moral motivations of one segment of society but must attribute similar motivations to a different segment.

Economists' efforts in this direction are only a small part of the story; economists as a class are probably competitive price-takers in the "moral arguments" industry. Philosophers philosophize more than we do (certainly in sheer quantity and probably even after adjusting for quality). Articles in magazines like Harper's or the Atlantic Monthly exhort

us to adopt one system of health care allocation rather than another, and argue on the basis of morality or justice. If readers were not genuinely moved by such arguments, the articles would presumably concentrate on explaining who gains what from each of the alternatives, rather than why some alternatives are right and others wrong.

Imagine a protective tariff whose only material effects are to increase the income of one auto worker by \$30,000 and to decrease the income of one farmer by \$50,000. Economists who revere efficiency will deplore this tariff, and experience small but genuine emotional distress when it is implemented. Would they experience the same emotional distress if they learned that an anonymous auto worker and an anonymous farmer had experienced exogenous gains and losses of the same magnitude?

The rhetorical question can be answered in a variety of ways beginning “No, but...” Such as: “No, but the reason is that the protectionist tariff is a bad omen about future policies.” Partly true, but even if the tariff were known to be a one-shot event containing no information about the future, it would still make me unhappy in a way that good and bad fortunes brought on by the weather would not. Perhaps this is only because the tariff seems more “avoidable”, but I think it goes beyond this. I would hate to base my entire argument on this kind of introspection, but perhaps the reader’s own introspection can buttress the case here.

Having opened the door to introspection, though, let me make one more point before I close it. We appear to have different priorities in our private affairs than we have in the voting booth. One can easily imagine accepting the benefits of a redistributive government program while deploring the program and voting against it. A simple resolution of this “paradox” is that the voter dislikes other aspects of the program more than he likes his personal gains from it. If those despised “other aspects” are purely material, we must conclude that the voter would be willing to pay for the right to transfer income around among a bunch of strangers in a direction opposite to what the government is doing. I doubt if this is often true. A more plausible hypothesis is that voting behavior expresses a preference about the social process itself, not just its outcome.

If such considerations are important, then so is the concept of self-justification. This paper demonstrates that under certain hypotheses there is a unique self-justifying social welfare function. It should be clearly stated that there is no basis for assuming that these hypotheses are in any sense likely to hold. The import of the main theorem is to suggest that multiple self-justifying optima are likely when agents' philosophical and selfish interests reinforce each other, and less likely when the two kinds of interest are at odds.

Even when the uniqueness theorem fails, however, it is clear that in general almost all welfare functions must fail to be self-justifying. This suggests a fundamental reassessment of the traditional approach to normative economics.

References

- K. Arrow, 1950, "A Difficulty in the Concept of Social Welfare", *Journal of Political Economy* 58, 328-346.
- A. Bergson, 1938, "A Reformulation of Certain Aspects of Welfare Economics", *Quarterly Journal of Economics* 52, 310-334.
- F. Browder, 1968, "The Fixed Point Theory of Multi-Valued Mappings in Topological Vector Spaces", *Mathematische Annalen* 177, 283-301.
- F. Hahn, 1982, "On Some Difficulties of the Utilitarian Economist", in *Utilitarianism and Beyond* (A. Sen and B. Williams, editors), Cambridge University Press.
- F. Kant, 1785, *Grundlegung zur Metaphysik der Sitten*.
- F. Kant, 1788, *Kritik der Urteilskraft*.
- E. Klein and A. Thompson, 1984, *Theory of Correspondences*, Wiley and Sons.
- P. Samuelson, 1947, "Foundations of Economic Analysis", Harvard University Press.
- A. Sen, 1977, "On Weights and Measures: Informational Constraints in Social Welfare Analysis", *Econometrica* 45, 1539-1572.
- W. Vickrey, 1960, "Utility, Strategy and Social Decision Rules", *Quarterly Journal of Economics* 74, 507-535.

Acknowledgements

I have had helpful comments from John Boyd, Marvin Goodfriend, James Kahn, Alan Stockman and William Thomson, whom I acknowledge with gratitude.