The Methodology of Normative Economics

Landsburg, Steven E.

Working Paper No. 343
March 1993

The Methodology of Normative Economics

Steven E. Landsburg

Rochester Center for Economic Research
Working Paper No. 343

2/15/93

The Methodology of Normative Economics

by

Steven E. Landsburg

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 affect on consumption.

In Section 1 I will construct a general framework for discussing normative questions. I will introduce the notion of a "self-justifying" welfare function, which is one that avoids the sort of paradox described above. In Section 2 I will present a simple one-dimensional example that illustrates all of the main ideas. In Section 3 I will call attention to a remarkable property of self-justifying optima: At the optimum, agents must disagree about purely philosophical issues, even if they ignore their own self-interest. In Section 4 I will state existence and uniqueness theorems for the one-dimensional model. The import of the uniqueness theorem is that when agents' philosophical predilections are at odds with their own selfish preferences, the self-justifying welfare function tends to be unique. In this circumstance a benevolent planner has essentially no choice about what to maximize. In Section 5, I present a far more general model in which the same sort of conclusion holds. The hypothesis of the theorem is again that agents have inner conflicts between their philosophies and their self-interest. I do not claim that this hypothesis is likely to hold. Rather, the theorem is interesting because it illuminates the fact that if there *are* multiple optima—which there may well be—it tends to be because such conflicts are relatively rare. In Section 6 I will consider and respond to some potential objections to the entire enterprise, and in Section 7 I will state some brief conclusions.

1. Normative Economics

First, we need a framework in which to model normative debates. The framework presented here is not perfectly general; for example it appears not to allow for a doctrine of natural rights. But it is more than adequate for most of the issues (such as equity/efficiency trade-offs) that most economists typically think of when they hear the word *normative*.

There are n individuals, with utility functions $u^1, u^2, \ldots, u^n$. These utility functions might or might not have cardinal significance. They take as argument a vector $x = (x^1, \ldots, x^p)$ which is permitted to range over some 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 according to some algorithm that takes account of individual preferences. For concreteness, I will sacrifice enough generality to assume that the planner maximizes a social welfare function $F(u^1, \ldots, u^n)$. 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

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

The more modern advocate of economic efficiency, rejecting cardinal notions, allows us to insert "weights" α_i and maximize

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

the weights themselves being a matter of indifference. The strong advocate for equity urges us to take

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

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

[1] A possible objection is that preferences regarding F are not primary but derived from preferences regarding outcomes; I will address this objection in Section 6.

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.

I propose to address the paradox by modifying the model. There remain n individuals with utility functions $u^1, \ldots, u^n$. The allocation of resources is represented by a vector x which is allowed to range over some convex subset $X \subset \mathbf{R}^p$. The social welfare function is denoted $F(u^1, \ldots, u^n)$, 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

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

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

The notion of a self-justifying social welfare function has both positive and normative content. Natural arguments suggest both that a benevolent planner *will* choose a self-justifying F and that he *ought* to choose a self-justifying F.

Notice that a self-justifying welfare function corresponds to a fixed point of the process $F \mapsto G(F)$. Any initial welfare function instructs the planner to replace it with one iteration of the process. Conceivably, though not necessarily, the resulting sequence of iterations converges to a fixed point. One alternative is that it settles into a recurrent cycle.

Perhaps the vagaries of the political process, in which the planner's regime evolves over time, can be modeled by this sequence of iterations. Here is a highly stylized example.

Suppose that social equity enters into people's tastes as a normal good. Initially people are poor and direct the planner to adopt policies that will maximize total wealth. Consequently they become wealthier, leading them to demand less emphasis on wealth creation and more on equity; the planner complies; wealth decreases, and the cycle continues.

2. Examples

In the introduction, I gave an example where the planner's initial welfare function calls for a heavy emphasis on equity, to be achieved by improving the welfare of the least well-off member of society. If that least well-off member happens to place a lot of value on living in a world where equity is unimportant to the planner, then the best way ofor the planner to improve his welfare can be to abandon the quest for equity. In this case, the equity- emphasizing welfare function fails to be self-justifying.

Suppose for another example that the planner seeks to maximize the sum of total utility, and that many people get a lot of utility from living in a world where the planner values only equity. Here again, the planner's goal is likely not to be self-justifying.

Here is a less contrived family of examples. Two individuals with differentiable utility functions u and v must divide one unit of a single consumption good. (In the context of the notation from the previous section, $n = 2$, $u^1 = u$, and $u^2 = v$.) 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 differentiable welfare function F_α, where α is a parameter in $[0,1]$ and for each value of α, F_α is a concave function increasing in both variables.

For example, we might have

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

In this example, those with an affinity for classical utilitarianism will prefer α near 1; those with a more egalitarian bent might prefer α near 0.

For another example, we might have

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

Here the choice of α determines how heavily each individual's utility is weighted in the social welfare function.

Returning to generalities on this family of examples, we can write $u = u(x, \alpha)$ and $v = v(x, \alpha)$, where u and v are utility functions assumed to satisfy the usual concavity hypotheses. We assume that $u_x > 0$ and $v_x < 0$. (That is, more is preferred to less for each inidvidual.) Given α, the planner chooses $x = x(\alpha)$ and $\beta = \beta(\alpha)$ to maximize

$$F_\alpha(u(x, \beta), v(x, \beta)). \tag{2}$$

The welfare function F_α is self-justifying if and only if

$$\beta(\alpha) = \alpha. \tag{3}$$

(In general, the functions $x(\alpha)$ and $\beta(\alpha)$ could be multi-valued; that is, there might be more than one argument that solves the maximization problem. In that case, we will say that the welfare function is self-justifying provided (3) holds for some choice of β.

A naive count of equations and variables suggests that in general, (4) should have a finite number of solutions. In Section 4, I will address questions of existence and uniqueness.

3. On the necessity of philosophical disagreements.

In this section I will point out a simple property of the solutions to the examples in Section 2. This property will be used in analyzing the more general model of Section 5.

Continue, then, to refer to the family of examples from Section 2. I will refer to the current value of α as the current *regime*. If the planner adjusts the regime by a small amount $\mathrm{d}\alpha$, the first consumer's utility changes by

$$[u_1 \cdot x'(\alpha) + u_2] \cdot \mathrm{d}\alpha. \tag{4}$$

The two terms in brackets can be understood separately. A change in regime leads to a change in consumption (x) and the first term represents the change in utility due to this change in consumption. Its sign is the sign of x'. The second term is the "purely philosophical" component. It reflects the change in utility resulting from the consumer's preferences about social institutions, without regard to how those institutions affect him directly. Its value could be either positive or negative.

In response to the regime change, the second consumer's utility changes by

$$[v_1 \cdot x'(\alpha) + v_2] \cdot \mathrm{d}\alpha]. \tag{5}$$

The terms can be broken down as for the first consumer. The sign of the first is the sign of $-x'$; the second could be either positive or negative.

It is quite clear that at a self-justifying optimum expressions (4) and (5) cannot be simultaneously either positive or negative. If, for example, both were positive, then an increase in α would be a Pareto improvement and the planner would change the regime.

A far more interesting observation is that the purely philosophical terms u_2 and v_2 also can not be simultaneously positive or negative. This is an easy consequence of the first-order conditions for the maximization problem (1), together with the hypotheses $u_1 > 0$ and $v_1 < 0$.

To illuminate what this means, consider the family of welfare functions (1). One can easily imagine the first consumer (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". Jack is telling us that in equation (4), the second term is positive, the first is negative, and the sum is negative. Such values are perfectly possible.

At the same time, the second consumer (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 u_2 and v_2 can be simultaneously zero) the arithmetic of the first-order conditions for maximizing (2) rules the scenario out. If Jack is philosophically committed to more utilitarianism ($u_2 > 0$) then Jill must be philosophically committed to more equity ($v_2 < 0$).

In some sense, equity (as measured, say, by $1 - \alpha$) can be thought of as a public good that is provided at zero marginal cost. At any interior social optimum, there have to be some citizens who would prefer more of it and others who would prefer less; otherwise there is an obvious Pareto improvement available. But the analogy with public goods is not perfect, since the fixed-point condition that is critical in the present situation plays no role in the traditional public goods analysis. Nevertheless, the intuition leads to a valid conclusion.

The correct intuition is this: At the self-justifying optimum, a change in regime causes a change in each party's consumption. However, because consumption has already been allocated to maximize the welfare function, the resulting change in welfare is of second order. On the other hand, the direct effect on welfare via philosophical preferences is (except in very special circumstances) of first order. Therefore, for small changes in regime, the philosophical term dominates. Consequently, if there were philosophical agreement, the planner would always want to make the corresponding regime change.

4. Existence and uniqueness.

Here we consider existence and uniqueness of a self-justifying optimum for the simple model of Section 2. The results in this section are all special cases of far more general results to be derived in Section 5, and are presented here only for expositional purposes. Therefore proofs will for the most part be omitted in this section.

The existence of a fixed point for the correspondence $\alpha \mapsto \beta(\alpha)$ is a straightforward application of the standard maximum theorems found, for example, in Chapter 9 of (Klein and Thompson, 1984). The interesting question, therefore, is uniqueness.

To get a uniqueness proof, we need some additional assumptions. First, recall that $u(x, \alpha)$ and $v(x, \alpha)$ are concave differentiable functions. We add the following additional hypothesis:

Hypothesis 1. *u and v are additively separable in x and α.*

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

Recall that $F_\alpha(u, v)$ is a parameterized family of differentiable welfare functions from which the planner must choose. These are assumed to be concave and increasing in both u and v. Introduce the function

$$\mathcal{F}(u, v, \alpha) = F_\alpha(u, v)$$

and assume that $\mathcal{F}$ is differentiable. We need the following additional hypothesis:

Hypothesis 2. *The quantity*

$$v_2 \cdot \frac{\partial(\mathcal{F}_2/\mathcal{F}_1)}{\partial \alpha}$$

is everywhere negative.

This hypothesis says that there is a conflict between the second agent's philosophical predilections and his selfish interests. The term v_2 is positive if and only the agent's philosophy argues for an increase in α. The term $\partial(\mathcal{F}_2/\mathcal{F}_1)/\partial\alpha$ is positive if and only if an increase in α would lead the planner to give his utility more relative weight in the social welfare function. The hypothesis says that the 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 $\partial(\mathcal{F}_2/\mathcal{F}_1)/\partial\alpha$ is unambiguously postive, so Hypothesis 2 requires $v_2 < 0$, i.e. the purely philosophical component of the second agent's preferences argues for a decrease in α.

Example 2. Suppose that

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

Then Hypothesis 2 is equivalent to the following statement: $v_2 < 0$ when and only when $v > u$. In view of the necessity of philosophical disagreements, this can be reinterpreted as: The agent with the higher utility always has a philosophical preference for more equity.

Remark 1. Hypothesis 2 appears to be asymmetric in the two agents, but the appearance is deceptive. The necessity of philosophical disagreements implies that Hypothesis 2 is equivalent to the symmetric statement that results when the roles of u and v are reversed.

Remark 2. Suppose that $H(F, \alpha)$ is a differentiable function that is monotonically increasing in F for each α. Then the planner's problem remains unchanged if the function $\mathcal{F}(u,v,\alpha)$ is replaced by $\mathcal{G}(u,v,\alpha)$, defined as

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

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

Remark 3. It is clear that something like Hypothesis 2 is necessary to generate a uniqueness theorem. If everyone's philosophical preferences tended to reinforce their selfish interests, 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 a dictator.

Theorem. *Given Hypotheses 1 and 2, there is a unique self-justifying social welfare function.*

This theorem is a special case of the main theorem to be proved in the next section.

5. A More General Model.

This section presents a more general model of self-justification. There are n agents indexed by $\{1, \ldots, n\}$. Social welfare functions are parameterized by points in a given closed convex subset of $\mathbf{R}^m$, with a typical point being denoted $(\alpha^1, \ldots, \alpha^m)$. Material outcomes (e.g. the allocation of some collection of goods) are parameterized by points in a given closed convex subset of $\mathbf{R}^r$, with a typical point being denoted $(x^1, \ldots, x^r)$. Agent i has the differentiable concave utility function

$$u^i(x^1, \ldots, x^r; \alpha^1, \ldots, \alpha^m).$$

The family of welfare functions is given by a differentiable function

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

For any fixed $\alpha = (\alpha^1, \ldots, \alpha^n)$, the function

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

is concave and increasing in each variable.

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

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

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

Existence of a self-justifying welfare function is a consequence of the maximum theorems cited in the preceding section. We turn to the question of existence, and make three hypotheses.

Hypothesis 0. *The matrix of partials*

$$\left(\frac{\partial u^i}{\partial x^j}\right)$$

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*

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

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

$$M_{ij} = \frac{\partial u^j}{\partial \alpha^i}$$

where i runs from 1 to m and j runs from 2 (not 1!) to n. Let N be the matrix given by

$$N_{ij} = \frac{\partial (F_i/F_1)}{\partial \alpha_j}$$

where i runs from 2 (not 1!) to n and j runs from 1 to m.

Then the matrix

$$M \cdot N$$

is everywhere negative semi-definite.

This generalizes Hypothesis 2 of the preceding section and is a way of formalizing the requirement that there is substantial tension between agents' philosophical predilections and their selfish interests.

Theorem. *Under Hypotheses 0, 1 and 2, the self-justifying welfare function is unique.*

To prove the theorem, we introduce a substantial amount of notation.

Let

$$\mathcal{F}_i = \frac{\partial \mathcal{F}}{\partial u^i} \qquad \hat{\mathcal{F}}_i = \frac{\partial \mathcal{F}}{\partial \alpha^i} \qquad u^i_j = \frac{\partial u^i}{\partial x^j} \qquad \hat{u}^i_j = \frac{\partial u^i}{\partial \alpha^j}$$

$$u^i_{jk} = \frac{\partial u^i}{\partial x^j \partial x^k} \qquad \hat{u}^i_{jk} = \frac{\partial u^i}{\partial \alpha^j \partial \alpha^k}$$

Let U be the matrix defined by $U_{ij} = u^i_j$ and let $\hat{U}$ be the matrix defined by $\hat{U}_{ij} = \hat{u}^i_j$.

The first-order conditions for the planner's problem are

$$\sum_{i=1}^{n} \mathcal{F}_i u_s^i = 0 \quad \text{for all} \quad s \in \{1, \ldots, r\} \tag{6}$$

$$\sum_{i=1}^{n} \mathcal{F}_i \hat{u}_t^i = 0 \quad \text{for all} \quad t \in \{1, \ldots, m\} \tag{7}$$

Equations (6) and (7), together with Hypothesis 0, imply that there is a matrix Θ with

$$U \cdot \Theta = \hat{U}.$$

Now let A, C, D and E be the matrices whose (i,j) components are given by

$$A_{ij} = \sum_{\substack{1 \le k \le n \\ 1 \le \ell \le n}} \mathcal{F}_{k\ell} u_i^k u_j^\ell$$

$$C_{ij} = \sum_{k=1}^{n} \frac{\partial \mathcal{F}_k}{\partial \alpha_j} u_i^k$$

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

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

Set

$$Z = A + D + A\Theta E^{-1} \Theta^T D$$

and note that A, D, E and therefore Z are all negative definite.

Now view x^i and β^j as functions of the α's, chosen to maximize $\mathcal{F}(-,-,\alpha)$.

Differentiating (6) and (7) with respect to each α^j and taking account of Hypothesis 2 gives a matrix equation

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

By inspection, one verifies that the solution to this equation is

$$\left(\frac{\partial x^i}{\partial \alpha^j}\right) = -ZC \qquad \left(\frac{\partial \beta^i}{\partial \alpha^j}\right) = -E^{-1}\Theta^T D Z^{-1} C.$$

To prove the theorem, it suffices to show that $(\partial \beta^i/\partial \alpha^j)$ is negative definite, for which it suffices from the above to show that $\Theta^T C$ is negative definite.

But combining the first-order condition (7) with the quotient rule for differentiation, one easily finds that $\Theta^T C$ differs from MN by a diagonal matrix with non-negative entries, where M and N are the matrices introduced in Hypothesis 3. Hypothesis 3 then completes the proof.

6. Objections and Responses.

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. At this point I will address some natural objections to the entire exercise.

Objection 1: *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 then 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.

Objection 2: *Don't all models leave some things out? Why is it important to include this particular thing?* Let me expand this objection by drawing an analogy: Some people study positive economics because they thirst for knowledge. Yet positive models rarely incorporate a thirst for knowledge in the utility function. This seems like a pretty weak criticism of positive economics. Isn't the present criticism of normative economics perfectly analogous?

Well, no. We leave the thirst for knowledge out of our positive models because we think that it plays a relatively small role in the phenomena we are attempting to explain. There is no reason why the economist studying the model has to represent a large segment of society. But when we do normative theory with the goal of giving policy advice, we set out not to satisfy our own curiosity but to convince the multitudes that we have something interesting to say. This would seem to require a degree of faith that the multitudes actually care about something related to the issues we are addressing.

Objection 3: *The proposal here is to take moral preferences as given. Isn't this inconsistent with the resources that are devoted to trying to sway people's moral preferences?* Yes, in the same sense that the simplest positive models are inconsistent with resources being devoted to advertising. Eventually one hopes for a more sophisticated theory. Perhaps the arguments in the utility function should be *attributes* of the social welfare function rather than the function itself; public debate can then be modelled as an attempt to disseminate scarce information about the attributes of particular F's.

7. Conclusions.

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. I have argued in Section 6 that the issues raised in this paper cannot be subsumed in the traditional framework.

References

E. Klein and A. Thompson, *Theory of Correspondences*, Wiley and Sons (1984).

Acknowledgements

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