J. S. Mill's Liberal Principle and Unanimity

Edward J. Green*

June, 1994

Abstract

The broad concept of an individual's welfare is actually a cluster of related specific concepts that bear a "family resemblance" to one another. One might care about how a policy will affect people both in terms of their subjective preferences and also in terms of some notion of their objective interests. This paper provides a framework for evaluation of policies in terms of welfare criteria that combine these two considerations. Sufficient conditions are provided for such a criterion to imply the same ranking of social states as does Pareto's unanimity criterion. Sufficiency is proved via study of a community of agents with interdependent ordinal preferences.

Keywords: social choice, liberal principle, Pareto relation, unanimity, interdependent preference.

* Department of Economics, University of Minnesota; and Research Department, Federal Reserve Bank of Minneapolis. This paper can be obtained by email or ftp, as ewp-ge/9406004, from the internet archive at econwpa.wustl.edu.

1. Introduction

The broad concept of an individual's welfare is actually a cluster of related specific concepts that bear a "family resemblance" to one another. Economists have tended to focus attention almost exclusively on just one of these specific concepts, though. Essentially, the defining characteristic of this concept is that questions regarding people's welfare have to be reduced to questions regarding their preferences in order to be intelligible.

While agreeing wholeheartedly that this Paretian concept of welfare is worthy of the intensive study that it receives, one can still hope to overcome the exclusive focus of welfare economics on just one among the many possible concepts of welfare. Once it is recognized that welfare is a family of related normative concepts, then there is no inconsistency in desiring to be able to evaluate policies with respect to several of them. One might care about how a policy will affect people both in terms of their subjective preferences and also in terms of some notion of their objective interests. Here I am adopting the term 'objective interest' to denote any specific notion of a person's well-being that might be quantified by an index of measurements of various aspects of the person's situation, in order to reserve the term 'welfare' for use in its broad sense.

Obviously there can be many distinct notions of a person's objective interest. For instance, some notions may be based exclusively on measurements of the bare physical necessities of life while other notions may encompass social and psychological aspects of a person's situation as well. It should be possible to design a satisfactory framework for evaluation of policies in terms of welfare criteria based on notions of people's objective interests, and also in terms of criteria that combine considerations of objective interests and preferences, just as neoclassical economists have succeeded in doing for preference-based notions.

Moreover, once it is granted that there may several alternative concepts of welfare that are all worthy of consideration in policy making, then some interesting general questions arise concerning the relationship between alternative welfare analyses. Are there salient features common to various specific notions of a person's objective interest—analogous to the ordinal features by which economists characterize preferences—that might serve as a basis for economic analysis? If so, can any general comparison be made between the conclusions of welfare analyses based on various notions of objective interest possessing these features, or between the conclusions of such an analysis and a neoclassical analysis based exclusively on preferences? In particular, are there any interesting conditions under which all of these various analyses would reach identical conclusions? This paper is an attempt to study these questions.

The main goal of the paper is to compare the neoclassical welfare criterion of unanimous preference (formulated by Pareto (1909) and emphasized in welfare analysis by Wicksell (1935)) to an alternative criterion that is somewhat in the same spirit, but that relies in an essential way on a notion of each person's objective interest as well as on a notion of persons' preferences. The formulation of this alternative criterion is suggested by John Stuart Mill's discussion of liberalism.[1]

[1] Mill's thinking has also been influential in stimulating the formal investigation (beginning with Sen 1970) of the problem of constraining Paretian welfare analysis to respect persons' rights. Schick (1980) and Riley (1988) have employed formal social-choice theory to interpret Mill's political philosophy as a whole.

Mill emphasizes two connected ideas: that there is a notion of each person's interest that is distinct from that person's preferences, and that a person ought to be permitted to do as he pleases so long as he does not damage the interests (in this non-preferential sense) of others (Mill 1859, Book I and Book IV). He extends this idea of liberal permissibility to coalitions. That is, he proposes that a group of people ought to be permitted to act unanimously as long as they do not damage the interest of anyone outside the group. This proposal would license actions of which the entire community approves, because the group that includes everyone would act unanimously and there would be no outsider whose interest could be damaged. Thus Mill's criterion of liberally permissible action is at least as broad as the Pareto-improvement criterion. Furthermore Mill states explicitly that he interprets the notion of people's interests narrowly enough so that some actions would also be permissible although they were not unanimously approved. That is, Mill takes the position that a person's mere preference that an action should not be taken does not make the action damaging to the person's objective interest. He proposes that someone should be permitted to take an action of which no one else approves, as long as those others' disapproval reflects such mere preferences unconnected to their interests. On this interpretation, Mill's proposed criterion of liberal permissibility is strictly broader than the Pareto-improvement criterion.

The preceding discussion makes it clear that Mill's views on liberally permissible action are based on two separate considerations: a specific notion of what objective interests people possess, and a general characterization of the relationship between people's objective interests and the permissibility of actions that other people might wish to take. Thus someone might endorse Mill's general characterization, but differ with him about concrete cases on the basis of holding a different view of what are people's objective interests. (Indeed, someone might suppose contrary to Mill that a mere preference does create a corresponding objective interest. When Mill's general characterization is interpreted according to that view, the resulting criterion is simply a reformulation of the Pareto-improvement criterion.) The formal theory to be presented here will emphasize the separation between Mill's two considerations. The main result to be proved is that, even in the context of an assumption about the relationship between preferences and interests that is significantly weaker than to posit an identification between them, Mill's general characterization of liberal permissibility continues to coincide exactly with the Pareto-improvement criterion. This result probably would have surprised both Mill and his philosophical critics. It shows that the conflicts between some alternative welfare criteria are much less pervasive than might have been believed. However it does identify some particular situations (namely, those in which the relationship between preferences and welfare assumed in the theorem are implausible) in which the liberal and Paretian criteria are likely to lead to different conclusions.

2. The setting of the problem

In order to examine carefully the logic behind Mill's suggestion that his liberal principle permits a wider class of actions than just those that accomplish unanimously preferred changes, I am going to restate the suggestion in language that is parallel to that of neoclassical welfare economics. There a person's preferences are represented formally by a binary relation that may hold between *social states*, which in turn represent possible situations of the community. A person's objective interest can also be also be represented formally by a binary relation between social states. As in the case of preferences, a distinct relation is identified with each person or *agent* in the community.

In welfare economics, the relation of unanimous preference (often called the "Pareto-improvement relation") is defined from the preference relations of individuals. Specifically, one social state is the *unanimous successor* of another, status-quo social state if every agent prefers the change to it from the status quo, with at least one agent's preference being strict. Analogously, define a social state to be the *liberal successor* of a status-quo social state if all members of some group of agents prefer the change to it from the status quo, with at least one member of the group having a strict preference, and if also the change does not reduce the welfare of any agent outside the group. (In both of these definitions, I refer to one social state as the status quo only to distinguish it from the other social state in the pair. I do not mean to imply that it is historically determined or special in any other way.)

Using these definitions, Mill's suggestion can be restated as the thesis that one social state may be a liberal successor of another without necessarily being its unanimous successor. Now I specify a formal theory in which this thesis can be expressed. The aim of this theory is to characterize the general logic of the thesis in a way that is completely explicit, and that is also general in the sense of being independent of specific proposals regarding what might constitute a person's interest.

Consider a finite set I of agents, and a set X of social states. (i, j, and k will range over I, and w, x, y, and z will range over X.) To each agent j are associated two binary relations on X, W_j and R_j. When xW_jy, this means that state x provides for the objective interest of agent j as well as state y does. I will refer to W_j as the *interest relation,* or simply the *interest,* of agent j. When xR_jy, this means that j weakly prefers x to y.

Define connectedness of a relation to mean that each pair of states is related in at least one direction, and for each agent j, assume that

$$W_j \text{ is transitive and reflexive, and } R_j \text{ is transitive and connected.} \tag{1}$$

V_j and P_j will denote the asymmetric (or strict) parts of W_j and R_j respectively. E_j and I_j will denote the symmetric parts of W_j and R_j respectively.

Within this simple formal language, it is possible to state three substantive conditions that may be placed on agents' preferences. The first condition actually guarantees three things: that each agent cares only about his own interest and possibly the interests of others, that each agent prefers a social state that better provides for his own interest if the interests of others are held

constant, and that each interest relation is connected on any set of states on which all other agents' interests are constant. Formally, for any agent j and for any states x and y,

$$\text{If } \forall i \neq j \; xE_iy, \text{ then } [xW_jy \iff xR_jy]. \tag{2}$$

This will be referred to as the condition that *preferences are based on interests.*

Second, agent j will be called *nonpaternalistic* if his preferences are determined by his own interest and the preferences of others. Formally, j is nonpaternalistic if, for all x and y,[2]

$$\text{If } xW_jy \text{ and } \forall i \neq j \; xR_iy, \text{ then } xR_jy. \tag{3}$$

The third condition that can be stated regarding an agent's preferences is that his preferences about trade-offs between the interests of agents within any one group in the population are independent of the situation regarding interests of other agents outside the group. Formally, agent i has preferences that are *separable in interests* if for any partition of I into groups J and K and for any states w, x, y, and z,

$$\text{If } \forall j \in J \; [wE_jy \text{ and } xE_jz] \text{ and } \forall k \in K \; [wE_kx \text{ and } yE_kz], \text{ then } wR_ix \iff yR_iz. \tag{4}$$

Separability of preferences in interests will play an important technical role in the arguments below. It is a necessary condition for an agent's preferences to be representable by a utility function that is a sum of functions that in turn are closely related to his and other agents' interest relations. Given the other assumptions that have been made here, and a few others that will be introduced shortly, it is also sufficient for the existence of such a representation. It should be noticed, though, that separability is an assumption with a lot of economic content. For example, if each agent's interest is taken to be his wealth, then it rules out the possibility that an agent prefers small increases in wealth for others whenever they are poorer than he is, but that (perhaps because of envy) he prefers that they should suffer small decreases in wealth when they are richer than he is. The example at the end of this paper is of interest principally because it shows that the liberal and unanimity relations may coincide even when agents' preferences are not separable in interests.

Three technical assumptions about the topological structure of the set of social states, and of agents' interest and preference relations, will also be needed. In contrast to the separability assumption, these assumptions do not seem to raise significant issues of interpretation. The meanings

[2] The use of 'nonpaternalism' for this condition is well entrenched in the literature on the interdependent preferences. It is easy to think of examples in which the condition applies in ways that intuitively have nothing to do with nonpaternalism, though. One defense of the terminology would be to point out that the obvious non-intuitive examples involve agents whose preferences are not based on interests. Thus, while the condition fails to capture the meaning of nonpaternalism (which involves concepts that cannot be stated in the parsimonious formal language adopted here), extensionally it defines the subset of interests-based preferences that a nonpaternalistic agent might hold.

of the first two assumptions are immediately obvious. The third assumption states that if E_j is viewed as a correspondence, then it is lower hemicontinuous:

$$X \text{ is a connected, separable topological space.} \tag{5}$$

$$\text{For every agent } j,\ W_j \text{ and } R_j \text{ have closed graph in } X^2. \tag{6}$$

$$\begin{aligned}&\text{For every agent } j, \text{ pair of states } y \text{ and } z, \text{ and neighborhood } Z \text{ of } z, \text{ if } yE_jz,\\ &\text{then there exists a neighborhood } Y \text{ of } y \text{ such that } \forall y' \in Y\ \exists z' \in Z\ y'E_jz'.\end{aligned} \tag{7}$$

The liberal and unanimity relations now have to be defined. State x is a *unanimous successor* to y (and is also described as being *Pareto superior* to y or a *Pareto improvement* of y) if

$$\forall i\ xR_iy \text{ and } \exists j\ xP_jy. \tag{8}$$

Define x to be a *liberal successor* of y if, for some coalition $J \subseteq I$,[3]

$$\forall i \in J\ xR_iy \text{ and } \exists j \in J\ xP_jy \text{ and } \forall k \notin J\ xW_ky. \tag{9}$$

[3] This relation of liberal succession is one of two possible ways (within the present theory) to formalize Mill's criterion for when a change of social state should be countenanced. The other possibility would be to stipulate that a coalition may make any change of state that does not give non-members strictly lower levels of welfare than they previously enjoyed. That is, x could be defined to be a liberal successor of y if, for some coalition $J \subseteq I$,

(9′) $\forall i \in J\ xR_iy$ and $\exists j \in J\ xP_jy$ and not $\exists k \notin J\ yV_kx$.

In general, (9′) defines a more permissive (i.e., set-theoretically larger) relation than does (9). However, (9) and (9′) define the same relation if all welfare relations W_j are connected. While connectedness is not assumed directly, it is implied by other assumptions that will be used here (cf. Lemma 6 below).

3. Mill and Pareto contrasted

At the close of section 1, I provided an informal summary of Mill's liberal principle. I pointed out the principle always countenances Pareto improvements. The formal theory just set forth affords an explicit proof of this. In particular, taking $J = I$ in (9) yields (8), showing that the unanimous-succession relation is a subrelation of the liberal-succession relation. This is exactly the formal statement of the assertion.

I also mentioned Mill's thesis that his principle countenances some changes of social state that are not Pareto improvements. I emphasized that the thesis follows from his view that there is at most a very loose relationship between a person's preferences and his objective interest or even people's objective interests in general. However, in section 2 I have assumed that all agents have nonpaternalistic preferences that are based on interests and separable in interests. These conditions all relate preferences to interests. One might wonder whether these relationships restrict the scope of the liberal principle so tightly that it must coincide exactly with the Pareto principle. Because the conditions in section 2 have been specified in an explicit and formal way, it is possible to construct an example that shows definitely that the conditions *do not imply* that Mill's principle and the Pareto principle coincide. I will provide one such example next. This example should not be interpreted to show that Mill's thesis is necessarily correct, though. Rather it shows that the conditions introduced in section 2 are weaker in some circumstances than they might seem to be. After presenting the example. I will introduce some further assumptions which rule out such degenerate circumstances. Once these assumptions have been set forth, the stage will be set for a serious examination of of the relationship between Mill's liberal principle and the Pareto principle. Now, here is the example of a community in which the liberal-succession relation is much more inclusive than the unanimous-succession relation.

Let $I = \{1, 2, 3, 4\}$, and let X be an interval of the real line. Suppose that xR_1y and xW_2y if $x \geq y$, and that xW_1y and xR_2y if $x \leq y$. Let agent 3 be identical to agent 1, and agent 4 to agent 2. It is easily verified that all of the assumptions (1)–(7) are satisfied. However, no pair of social states is related by the unanimous-succession relation, but every pair of distinct social states is related by the liberal-succession relation.

In this example, the assumptions (2) and (3) of interest-based preference and nonpaternalism have been trivialized. That is, each assumption is an implication with a hypothesis that is never satisfied. Now I will define four conditions which are intended to identify environments that are rich enough for the two assumptions to have nontrivial content. Note that the example above does not satisfy any of these conditions.

First, agents' interests will be said to have *product structure* if [4]

[4] It would be desirable to make only an assumption to the effect that there is some latitude for agents' interests to vary independently. The present, inordinately strong, assumption is made in order to invoke a theorem of conjoint measurement theory below. Krantz, et. al. ((1971), pp. 275-6) discuss the general need to have a theory of measurement based on a weaker assumption. Thus the objectionable feature of the present assumption seems to be a convenient technical fiction that need not be viewed as limiting the validity of the results to be derived here. Of course, social states must be viewed as conceptually possible profiles of interests that are not necessarily technically feasible to satisfy all at once.

$$\forall \{x_i\}_{i \in I}\ \exists y\ \ \forall i\ yE_i x_i. \tag{10}$$

Second, agent i will be said to have *idiosyncratic interest* if

$$\exists x\ \exists y\ [xV_i y \text{ and } \forall j \neq i\ xE_j y]. \tag{11}$$

Third, analogously, agent i will be said to have *idiosyncratic preferences* if

$$\exists x\ \exists y\ [xP_i y \text{ and } \forall j \neq i\ xI_j y]. \tag{12}$$

Fourth, it will be said that an *unambiguous improvement* is possible if

$$\exists x\ \exists y\ \forall i\ [xV_i y \text{ and } xP_i y]. \tag{13}$$

4. Agents' judgments and interpersonal agreement regarding interests

Taken together, assumptions (2) that agents' preferences are based on interests and (4) that agents' preferences are separable in interests have an interpetation that each agent's preferences are based on considerations of trade-offs between his own interest and other agents' interests. (There might well be other possible interpretations of agents' preferences as well.) The idea of making trade-offs among interests implies that the difference between the degree to which an agent's interest is satisfied in two distinct states is conceived as being a cardinal magnitude. Moreover, agents may either agree or disagree with one another about comparisons between these magnitudes. This is a subtle point, and it is also an important one for the welfare analysis which is to follow. Therefore let us consider it carefully now.

Consider two agents i and j, each of whose interest is completely determined by what kind of house he lives in. Suppose that there are four types of house—hovels, cottages, mansions, and palaces—in increasing order of satisfaction of an agent's interest. Suppose that a social state is simply a specification of a type of house for each agent.

The ordinal ranking of types of house determines the answers to some questions regarding cardinal comparisons, such as "Would it make more difference for the satisfaction of agent j's interest if he were to move from a cottage to a palace than if he were to move from a cottage to a mansion?" Clearly the former difference is larger, since the two changes begin at the same point but the latter change accomplishes only part (in terms of the ranking) of what the former accomplishes. However, the ranking does not determine the answers to other questions such as "Would it make a bigger difference for the satisfaction of agent j's interest if he were to move from a mansion to a palace, than if he were to move from a hovel to a cottage?" That is because neither of the ordinal intervals described by these two changes includes the other. If one were to represent

the magnitudes of differences in satisfaction of an agent's interest as distances between points on a line, then pictures like either of the following would be possible.

[hovel] [cottage] [mansion] [palace]

[hovel] [cottage] [mansion] [palace]

Information about agents' preferences can be interpreted to represent the agents as implicitly answering such questions, though. For example, consider the following two pairs of states: x and y, and x' and y'.

State x :	Agent i lives in a mansion;	Agent j lives in a mansion.
State y :	Agent i lives in a cottage;	Agent j lives in a palace.
State x':	Agent i lives in a mansion;	Agent j lives in a hovel.
State y':	Agent i lives in a cottage;	Agent j lives in a cottage.

Suppose that agent i prefers state y to state x and also prefers state x' to state y'. Then I will interpret agent i as judging that agent j's interest is affected more significantly by a move from a mansion to a palace than by a move from a hovel to a cottage, in the following sense. Agent i would be willing to bear the sacrifice of having to move from a mansion to a cottage in order to enable j to move from a mansion to a palace, but he would be unwilling to bear the same sacrifice in order to enable j to move from a hovel to a cottage. This pair of preferences can be interpreted as reflecting a judgment on the part of i that a move from a mansion to a palace would make a larger difference for the satisfaction of j's interest than would a move from a hovel to a cottage.

In an environment with many social states, this judgments-about-interests interpretation of agents' preferences will have restrictive implications. For example, consider another two pairs of social states.

State w :	Agent i lives in a cottage;	Agent j lives in a mansion.
State z :	Agent i lives in a hovel;	Agent j lives in a palace.
State w':	Agent i lives in a cottage;	Agent j lives in a hovel.
State z':	Agent i lives in a hovel;	Agent j lives in a cottage.

Suppose that i were to prefer state w to state z and state z' to state w'. The interpretation of preferences as reflecting judgments about cardinal differences in satisfaction of interests would

suggest that i considers a move from a hovel to a cottage to matter more for the satisfaction of j's interest than would a move from a mansion to a palace. The argument for this implication is the same as before, except now i's contemplated sacrifice is a move from a cottage to a hovel instead of from a palace to a mansion. If i were to hold all of the preferences that have been discussed in this paragraph and in previous one, then the judgments-about-interests interpretation of preferences would impute two inconsistent judgments to him. I will demonstrate later in this paper (in Lemma 10) that the assumptions made in the preceding two sections are sufficient to rule out such imputations of inconsistency. Technically, in addition to the assumptions that have been introduced so far, Lemma 10 also requires an additional *double-cancellation condition* in the case of two-agent communities and of some larger communitites. This condition will be set forth when the lemma is stated formally.

Agent j (or any other agent in the community) might either agree or disagree with agent i's assessments of how much difference various changes of social state would make to the satisfaction of j's interest. The main result to be proved in this paper will be that, in the presence of the assumptions that have already been made (along with the double-cancellation condition), this additional assumption that agents always agree with one another about such assessments is sufficient for the liberal-succession relation and the unanimous-succession relation to coincide.

5. Equivalence of the welfare criteria: a preliminary result

The model of a community set forth in sections 2 and 3 should seem familiar in many respects, especially in relation to work of Debreu. As in Debreu (1960), preferences of an individual agent are described in terms of ordinal and topological assumptions along with an assumption regarding separability. As in Debreu (1959), a welfare proposition regarding a community of agents described in such terms is to be investigated.[5] This investigation divides logically into two parts. First a representation theorem should be proved, asserting the existence of well behaved utility functions which represent agents' preferences and also of analogous functions which represent agents' interests. Then, taking advantage of these numerical representations, the proposition about welfare should be proved.

Since welfare analysis provides the motivation for the representation theorem, and also since the particular set of assumptions studied here in this paper not be the only ones which yield the sort of representation that is derived, I will begin by formulating the representation that is needed and by proving that it implies that the liberal-succession and unanimous-succession relations coincide.

Throughout this section and the remainder of the paper (except where explicitly stated to the contrary), I will assume that conditions (1)–(6) and (10)–(13) hold. The representation theorem that I will prove later from these assumptions (and the two further ones discussed at the end of the preceding section) contains the following three assertions (14)–(16) regarding the possibility of representing agents' preferences and interests by numerical functions.

[5] Debreu (1959) did not impose separability, however.

$$v_j : X \to I\!R \text{ is continuous and } \forall x\ \forall y\ [xW_jy \iff v_j(x) \geq v_j(y)]. \tag{14}$$

$$p_j : X \to I\!R \text{ is continuous and } \forall x\ \forall y\ [xR_jy \iff p_j(x) \geq p_j(y)]. \tag{15}$$

There exist scalars α_{ij} and κ_i such that[6]

$$\forall i\ \forall x\ p_i(x) = \Sigma_{j \leq n} \alpha_{ij} v_j(x) + \kappa_i. \tag{16}$$

The first two of these assertions are just the sort of ordinal representations that one would expect. (Note, however, that (14) implies the transitivity of interest relations which has not been assumed directly in (1).) The real force of the representation is provided by assertion (16). What this assertion obviously accomplishes is to reflect numerically the assumptions (3) and (4) that agents' preferences are based on interests and separable in interests. The assertion states significantly more than that, though. Each agent's preferences might individually be based on interests and separable in interests, and yet the numerical representation of someone's interest (agent i, say) that entered one agent's (agent j) utility function might be an arbitrary continuous, strictly increasing transformation of the numerical representation of i's interests that entered another agent k's utility function. Assertion (16) states that the representation of i's interest that enters j's utility function must be a *strictly increasing affine transformation* of the representation of i's interest that enters k's utility function. This means that when the preferences of agents j and k reflect *cardinal* comparisons of differences in satisfaction of i's interest (such as that it would make a greater difference for i to move from a hovel to a cottage than it would make for him to move from the cottage to a mansion), then j and k must agree with one another. As has already been explained in the preceding section, the existence of such agreement throughout the community regarding cardinal intrapersonal comparisons across social states of each agent's interest is a key assumption of the proof that liberal succession and unanimous succession coincide. This assumption does not follow from the assumptions about the community that have been made so far. An ordinal assumption that does imply assumption (16) (in the presence of the other assumptions) will be introduced below in section 8.

Now all of the assumptions have been introduced that are needed to prove the coincidence of the liberal-succession and unanimous-succession relations. In the proof, it will be convenient to write (16) in matrix form. To do so, define $v : X \to I\!R^n$ and $p : X \to I\!R^n$, where n is the number of agents in i, by

$$\forall x\ \forall j\ [v(x)]_j = v_j(x) \tag{17}$$

and

$$\forall x\ \forall j\ [p(x)]_j = p_j(x). \tag{18}$$

Let A be the matrix with coefficients α_{ij}. Then (16) can be written in matrix notation as:

$$\forall x \; p(x) = Av(x) + \kappa. \tag{19}$$

Now the proof of the welfare theorem will be presented in a series of lemmas.

Lemma 1: $v(X)$ contains a non-empty open subset of $\mathbb{R}^n$.

Proof: Let $x_j V_j y_j$ for each j, as is guaranteed by (11). By continuity of v_j and connectedness of X, the interval $[v_j(y_j), v_j(x_j)]$ is a subset of $v_j(X)$. This interval has nonempty interior, since v_j represents W_j. Since agents' interests have product structure, then,

$$(v_1(y_1), v_1(x_1)) \times \ldots \times (v_n(y_n), v_n(x_n)) \subseteq v(X). \quad \blacksquare$$

Lemma 2: A is nonsingular.

Proof: By idiosyncracy of preferences, for each i there exist x and y; such that xP_iy and $\forall j \neq i \; xI_jy$. Thus $p(x) - p(y) = A[v(x) - v(y)]$ is a scalar multiple of the usual basis vector in dimension i. $\blacksquare$

Let $B = (\beta_{ij}) = A^{-1}$. Then, for every i there exists a scalar λ_i such that

$$\forall x \; p_i(x) = \alpha_{ii} v_i(v) + \textstyle\sum_{j \neq i} \alpha_{ij} [\sum_{k \leq n} \beta_{jk} p_k(x)] + \lambda_i. \tag{20}$$

Equation (20) provides an implicit characterization of the preferences of agent i in terms of i's interest and other agents' preferences. However, p_i occurs on both sides of the equation. In order to have an explicit characterization, p_i should not occur on the right. That is, the term $\sum_{j \neq i} \alpha_{ij} \beta_{ji} p_i(x)$ should be subtracted from both sides of (20), and then both sides of the resulting equation should be divided by $1 - \sum_{j \neq i} \alpha_{ij} \beta_{ji}$. The result of these operations, defining $\gamma_{ik} = \sum_{j \neq i} \alpha_{ij} \beta_{jk}$, $\delta_{ii} = \alpha_{ii} / (1 - \gamma_{ii})$, $\mu_i = \lambda_i / (1 - \gamma_{ii})$, and $\delta_{ik} = \gamma_{ik} / (1 - \gamma_{ii})$ for $k \neq i$, is

$$\forall x \; p_i(x) = \delta_{ii} v_i(x) + \textstyle\sum_{k \neq i} \delta_{ik} p_k(x) + \mu_i. \tag{21}$$

There is one thing that has to be verified for this derivation of (21) to be sound. That is that $\gamma_{ii} \neq 1$. Otherwise, the final step of the derivation would have been division by zero. This verification is now provided.

Recall once again, that by idiosyncracy of i's preference, there exist states x and y such that xP_iy and $\forall k \neq i \; xI_ky$. By nonpaternalism, i's strict preference requires that xV_iy. Equation (15) implies that $p_k(x) = p_k(y)$ for all $k \neq i$, so subtraction of (20) from the corresponding equation for state y yields $[p_i(y) - p_i(x)] =$
$\alpha_{ii}[v_i(y) - v_i(x)] + \gamma_{ii}[p_i(y) - p_i(x)]$. By (14), $[v_i(y) - v_i(x)]$ is strictly positive. Therefore $\gamma_{ii} \neq 1$ if $\alpha_{ii} \neq 0$. However, if $\alpha_{ii} = 0$, then it would follow from (11), (14), (15) and (16) that $y_j R_j x_j$ (where x_j and y_j are the pair of states that (11) guarantees to exist for j), contrary to (2).

Since $\gamma_{ii} \neq 1$, then, preferences representable by (16) are also representable by (21). This latter representation is now studied further.

Lemma 3: In (21), for every i, $\delta_{ii} > 0$.

Proof: By idiosyncracy of preferences, there exist x and y such that xP_iy and $\forall j \neq i\ xI_jy$. Thus, by nonpaternalism of i, xV_iy. Therefore $\delta_{ii} > 0$ follows from (14), (15) and (21). ∎

Lemma 4: In (21), for every i and k, $\delta_{ik} \geq 0$.

Proof: By Lemma 3, it is sufficient to prove this inequality for $i \neq k$. By Lemma 1, there exist a subset $Y \subseteq X$ and a nonempty open set $U \subseteq \mathbb{R}^n$ such that v maps Y onto U. By Lemma 2, $p(Y) = A(U)$ is open in $\mathbb{R}^n$. For in define $q^i : Y \to \mathbb{R}^n$ by $[q^i(x)]_j = \delta_{ii}v_i(x) + \mu_i$ if $i = j$, and $[q^i(x)]_j = p_j(x)$ if $i \neq j$. Note that, by (21), $q^i = Q^ip$, where Q^i is a matrix that will also be denoted by (χ_{jk}). This matrix is the same as the identity matrix except for its ith row, and $\chi_{ii} = 1$ and $\chi_{ik} = -\delta_{ik}$ for $k \neq i$. It is obvious (using row i to expand the determinant) that Q^i is nonsingular, so $q^i(Y) = Q^iA(U)$, which is open in $\mathbb{R}^n$. Thus there exist states x and y such that $q^i(x) - q^i(y)$ is a positive scalar multiple of the usual basis vector in the k dimension. That is, xE_iy, xP_ky, and xI_jy for all other agents j. By nonpaternalism of i, xR_iy, so $\delta_{ik} \geq 0$ follows from (14), (15), and (21). ∎

These results have a useful geometric interpretation. Consider the vector space that is obtained from the space of all continuous, real-valued functions on X when functions that differ by a constant are identified with one another. Let V be the finite-dimensional subspace generated by $\{v_i\}_{i \in I}$. Lemma 1 asserts that V is isomorphic to $\mathbb{R}^n$. For a set F of vectors in V, let $K[F]$ be the convex cone generated by F. The possibility of an unambiguous improvement (i.e., assumption (13)) implies that $K[\{v_i\}_{i \in I} \cup \{p_i\}_{i \in I}]$ does not contain any linear subspace (except for $\{\emptyset\}$) of V. Lemma 2 assets that $\{p_i\}_{i \in I}$ is a basis of V. Lemma 4 asserts that, for each i, $p_i \in K[\{v_i\} \cup \{p_j\}_{j \neq i}]$. These facts are now used to establish the coincidence of the two social-choice relations.

Theorem 1: The additively-separable representation of preferences discussed here is sufficient for the Mill's principle to be equivalent to the Pareto principle. Specifically, suppose that the environment is as described in (1)–(6) and (10)–(13). If agents' interests and preferences can be represented as in (14)—(16), then Pareto superiority and liberal succession coincide.

Proof: Since the Pareto relation is contained in the relation of liberal succession, it is sufficient to prove that x is Pareto superior to y if x is a liberal succession of y. A contradiction will be obtained from the contrary assumption. In particular, assume that $J \subseteq I$ satisfies (9) but that yP_kx. Then $p_k \in K[\{v_i\}_{i \notin J} \cup \{p_j\}_{j \in J}]$. Thus $K[\{v_i\}_{i \notin J} \cup \{p_j\}_{j \in I}]$ is not a subset of $K[\{v_i\}_{i \notin J} \cup \{p_j\}_{j \in J}]$. By (13),
$K[\{v_i\}_{i \notin J} \cup \{p_j\}_{j \in I}]$ contains no linear subspace (except $\{\emptyset\}$) of V. Therefore, by (12) and Corollary 18.5.2 of Rockafellar (1970), there is an agent k such that $p_k \notin K[\{v_i\}_{i \notin J} \cup \{p_j\}_{j \neq k}]$. (Note that (12) rules out that p_j could be a scalar multiple of p_k for any $j \neq k$.) This contradicts Lemma 4, since $k \notin J$. ∎

6. Nonmalevolence

Theorem 1 has antecedents in the work of Bergstrom ((1971), (1989), (1988)), Green ((1979), (1982)), and Pearce (1983), all of which assume some version of (16). In the remaining part of this paper, it will be shown that the assumptions on ordinal preference and interest relations that have been introduced above (along with the double-cancellation condition and an ordinal assumption that will imply agents' agreement about intrapersonal interest comparisons) can replace (16) in the theorem. Before showing this, though, I will now briefly discuss one difference between Theorem 1 and it antecedents. The remainder of the paper is independent of this discussion.

In the antecedent theorems just mentioned, the possibility of an unambiguous improvement (13) has not been assumed but an additional assumption of *nonmalevolence* has been made. Formally, agent j is nonmalevolent if

(13′) $\forall x\ \forall y$ [if $\forall i\ xW_iy$, then xR_iy].

Neither of assumptions (13) and (13′) implies the other. In particular, note that the example in section 3 satisfies (13′) but not (13). To see that (13) does not imply (13′), consider two agents and let X be the Euclidean plane, with $v_1(x) = x_1$ and $v_2(x) = x_2$. Define $p_1 = 2v_1 - v_2$ and $p_2 = 2v_2 - v_1$. Obviously neither preference relation defined by these utility functions satisfies (13′), but (1, 1) is an unambiguous improvement over (0, 0).

Conditional on the other assumptions that have been made here, though, the assumptions (13) and (13′) are equivalent. It is easy to see that the assumptions of this paper (specifically (1), (4), (10), (11) and (13′)) imply (13). The converse implication is proved analogously to Theorem 1. Clearly (13′) implies that all of the coefficients α_{ij} in (16) are nonnegative. Thus, if some agent j does not satisfy (13′), then $p_j \notin K[\{v_i\}_{i\in I}]$. Thus $K[\{v_i\}_{i\in I} \cup \{p_j\}_{j\in I}]$ is not a subset of $K[\{v_i\}_{i\in I}]$. By (13), $K[\{v_i\}_{i\in I} \cup \{p_j\}_{j\in I}]$ contains only the trivial linear subspace $\{\emptyset\}$ of V, so there is an agent k such that $p_k \notin K[\{v_i\}_{i\in I} \cup \{p_j\}_{j\neq k}]$. This contradicts Lemma 4, though, so the failure of (13′) for j implies the failure of (13). That is, if (13) holds, then all agents must satisfy (13′).

7. Additive conjoint measurement of preference

The proof of Theorem 1 has relied heavily on two assumptions that are not obvious consequences of what had previously been assumed. The first of these is assumption (14), that for each agent j there is a continuous function $v_j : X \to \mathbb{R}$ such that $\forall x\ \forall y\ [xW_jy \iff v_j(x) \geq v_j(y)]$. This assumption implies in particular that agents' interest relations are connected on X, which is a stronger connectedness claim than the limited one that follows from assumptions (1) and (2). The other assumption is (16), that the functions v_j can be specified in such a way that there exist scalars α_{ij} and κ_i such that $\forall i\ \forall x\ p_i(x) = \sum_{j\leq n} \alpha_{ij}v_j(x) + \kappa_i$.[7]

[7] (14) resembles (13) and (18) in its form, but it follows routinely from previous assumptions, using a result of Debreu (1959). The crucial difference is that R_i is assumed to be connected, while the connectedness of W_i must be proved.

These assumptions will be derived from ordinal and topological assumptions. The first step to accomplish this is to formulate three conditions (i.e., (21) - (23) below) that jointly imply (14) and (16). The first of these conditions will be established in this section, and the second and third in the next.

Call a function $f : X \to \mathbb{R}$ *compatible with* an equivalence relation E on X if, for all x and y, xEy implies that $f(x) = f(y)$. Say that f *is strictly increasing* with respect to a binary relation V on X if, for all x and y, xVy implies that $f(x) > f(y)$. Note that (14) is equivalent to the statement that, for each j, there is a continuous function v_j that is both compatible with E_j and strictly increasing with respect to V_j. The first of the three conditions incorporates this information from (14) directly in a way that closely resembles (16). For all agents i and j, let $N(i)$ be a set of agents, let ν_i be a scalar, and let p_i and u_{ij} be functions from X to $\mathbb{R}$. Condition (16) will be paraphrased by taking $N(i)$ to be the set of j such that $\alpha_{ij} \neq 0$ and u_{ij} to be $\alpha_{ij} v_j$. Now the first two conditions can be stated:

$$\forall i\ \forall x\ p_i(x) = \textstyle\sum_{j \in N(i)} u_{ij}(x) + \nu_i, \text{ where each } p_i \text{ satisfies (15) and each } u_{ij} \text{ for } j \in N(i) \text{ is continuous, non-constant, and compatible with } E_j, \tag{22}$$

and

$$\text{Each } u_{ij} \text{ for } j \in N(i) \text{ is strictly increasing with respect to } V_j. \tag{23}$$

Conditions (22) and (23) do not quite imply (16), because (16) entails (according to the paraphrase that has just been described) the third condition that[8]

$$\text{If } j \in N(h) \cap N(i), \text{ then there exist scalars } \sigma_{hij} \text{ and } \tau_{hij} \text{ such that } u_{hj} = \sigma_{hij} u_{ij} + \tau_{hij}. \tag{24}$$

Lemma 5: Conditions (22) and (24) imply condition (16).

Proof: Consider a well ordering of I. For every j, define $\iota(j)$ to be the first i such that $j \in N(i)$. Note that $\iota(j) \leq j$ (because $j \in N(j)$ by (2), (11), and (22)). For each h, define $M(h)$ to be the union of the sets $N(i)$ for i preceding h. For each j, let $v_j = u_{\iota(j)j}$. Define $\alpha_{hj} = \sigma_{h\iota(j)j}$ if $j \in N(h)$ and $\iota(j) < h$, $\alpha_{hj} = 1$ if $\iota(j) = h$, and $\alpha_{hj} = 0$ otherwise. Define $\kappa_h = \nu_h + \sum_{j \in M(h) \cap N(h)} \tau_{h\iota(j)j}$. It is routinely verified that these definitions satisfy (16). ∎

The remainder of this section is devoted to deriving (22) as a consequence of ordinal and topological assumptions. Besides the assumptions that have already been stated, only one further assumption is needed. This assumption, called the *double-cancellation condition*, strengthens the separability condition (4) in the case of an agent whose preferences depend only on the interests of himself and one other agent:

---

[8] $\sigma_{hij} = \alpha_{hj} / \alpha_{ij}$ and $\tau_{hij} = 0$.

If $\exists i \neq j\ \forall x\ \forall y\ [xW_iy$ and xW_jy jointly imply that $xR_jy]$, then the following condition holds for all states r, s, t, x, y, and z: if rR_jx and sR_jy and if rW_it, xW_is, yW_iz, rW_jy, sW_jt, and xW_jz, then tR_jz. (25)

Like separability, this condition is necessary (even without any other assumptions) for the existence of an additive conjoint representation. Note that the condition is satisfied trivially if $N(j)$ is not a two-element set.

Now it will be shown that the ordinal and topological assumptions that have been made so far are sufficient to establish (22). The proof that (22) holds for agent i depends on the cardinality of $N(i)$. If $N(i)$ has more than one element, then (22) will follow from a representation theorem for additive conjoint measurement due to Debreu (1959). If $N(i)$ is a singleton (i.e., if i's preferences depend only on his own interest), then it is sufficient to take $p_i = u_{ii} = v_i$ and $\nu_i = 0$, where v_i satisfies (14). The existence of such a function v_i is now established.

Lemma 6: Assumptions (1), (2), (5), (6), and (10) imply assumption (14), i.e., that for each agent j there is a continuous function $v_j : X \to I\!R$ such that $\forall x\ \forall y\ [xW_jy \iff v_j(x) \geq v_j(y)]$.

Proof: By a result of Debreu (1959), it is sufficient under these assumptions to show that each W_j is connected on X. Consider any agent j and any pair of states x and y. By (10), there exists a state z such that xE_jz and $\forall i \neq j\ yE_iz$. Either yR_jz or zR_jy by (1), so either yW_jz or zW_jy respectively by (2). Then by (1), either yW_jx or xW_jy respectively because xE_jz. ∎

In the case general that R_i and W_i may not coincide, the derivation of (22) depends on Debreu's representation theorem, which is now stated as Lemma 7. For a proof, see Debreu (1959), or Krantz, et al. (1971, §6.2, 6.11, 6.12). The continuity of u_{ij} asserted in the lemma is not explicitly mentioned in those sources, but it is clear from a careful inspection of Debreu's proof, in which he has noted on pp. 22-24 that u_{ij} is obtained from a continuous function and that the several steps of the construction of u_{ij} preserve continuity.[9]

Lemma 7 (Representation theorem for additive conjoint measurement): Suppose that $R_i \neq W_i$ and that assumptions (1) - (6), (11) and (25) are satisfied. Suppose that $X = \Pi_{j \in I}\ X_j$, where each X_j is a topological space and X has the product topology, and that, for each j, $\forall x\ \forall y\ [x_j = y_j$ implies $xE_jy]$. Then $\forall x\ p_i(x) = \sum_{j \in N(i)} u_{ij}(x) + \nu_i$, where $i \in N(i) \subseteq I$, p_i satisfies (15) and each u_{ij} for $j \in N(i)$ is continuous and non-constant and factors through X_j (i.e., $U_{ij}(x)$ depends only on x_j).

Lemma 7 falls short of the objective of this section in two respects. First, even given the hypotheses of the theorem, the conclusion that each function u_{ij} factors through X_j is weaker than the desired conclusion that each u_{ij} is compatible with W_j. Second, the hypothesis that X is a cartesian product of spaces that individually determine agents' interest levels is stronger than the

[9] Unlike Debreu's proof, the proof given by Krantz, et al. does not assume the separability of X, and it does not assert the continuity of the functions u_{ij}. I do not know whether the functions constructed in that proof actually are continuous. Since continuity of the functions representing agents' interest relations was used in Lemma 1 and will be used again in the next section, I have asserted the separability of X in (6).

product-structure assumption (10). The cartesian-product representation restricts the economic applicability of the lemma to private-goods environments, since the level of provision of a public good would be a coordinate of the social state that would affect the interests of all agents. The conclusion of Lemma 7 can actually be strengthened to include compatibility of each u_{ij} with E_j, and the hypothesis of a cartesian-product representation can be weakened to assumption (10) if (7) is also assumed. These amendments are now made in Lemma 8 and Lemma 10.

Lemma 8: Suppose that assumptions (1) - (6), (11) and (25) are satisfied. Suppose that $X = \Pi_{j \in I} X_j$, where each X_j is a topological space and X has the product topology, and that, for each j, $\forall x\ \forall y\ [x_j = y_j$ implies $xE_jy]$. Then $\forall x\ p_i(x) = \sum_{j \in N(i)} u_{ij}(x) + \nu_i$, where $i \in N(i) \subseteq I$, p_i satisfies (15) and each u_{ij} for $j \in N(i)$ is continuous, non-constant, and compatible with E_j.

Proof: Lemma 7 establishes everything except for the compatibility of u_{ij} with E_j, and it establishes that each u_{ij} factors through X_j. A contradiction will be derived from the assumption that, for some $j \in N(i)$, u_{ij} is not compatible with E_j. Specifically, suppose that xE_jy but that $u_{ij}(x) < u_{ij}(y)$. Let $z_j = x_j$ but let $z_h = y_h$ for all $h \neq j$. Then $\forall h\ zE_hy$, so zR_iy by (2). However, $p_i(z) < p_i(y)$, contradicting (15). ∎

A technical result is needed in order to relax the cartesian-product hypothesis to (10). Define a set Y of states to be *saturated* with respect to an equivalence relation E if $x \in Y$ whenever, for some y, $y \in Y$ and xEy. Note that a function is compatible with E if and only if its level sets are saturated with respect to E. It will be required that the property of saturation with respect to the relations E_j should be preserved when topological closures are taken. Lemma 9 asserts that this requirement can be met.

Lemma 9: If (1), (6), and (7) hold and Y is a set of states that is saturated with respect to E_j, then $c\ell(Y)$ is saturated with respect to E_j.

Proof: (1) entails that E_j is an equivalence relation. By (6), the graph of E_j is closed. Suppose that $c\ell(Y)$ is not saturated with respect to E_j. Then there exist states $y \in c\ell(Y)$ and $x \notin c\ell(Y)$ such that xE_jy. Since $x \notin c\ell(Y)$, there is an open set U containing x and disjoint from Y. By (7), there is an open set V containing y and satisfying $\forall v \in V\ \exists u \in U\ uE_jv$. Thus Y cannot be saturated with respect to E_j, because $Y \cap V \neq \emptyset$. ∎

Recall that if E is an equivalence relation on X, then the *quotient space* X / E is the set of E-equivalence classes of elements of X, with the finest topology that makes the mapping from elements to their equivalence classes continuous.

Lemma 10: Suppose that assumptions (1)–(7), (10), (11) and (25) are satisfied. Then $\forall y\ p_i(x) = \sum_{j \in N(i)} u_{ij}(x) + \nu_i$, where $i \in N(i) \subseteq I$, p_i satisfies (15) and each u_{ij} for $j \in N(i)$ is continuous, non-constant, and compatible with E_j. That is, (22) is satisfied.

Proof: Begin by defining xE_Iy if and only if $\forall i\ xE_iy$. Obviously E_i is an equivalence relation and each W_i induces a relation on X / E_I. By (2), each R_i also induces a relation on X / E_I. ('W_i' and 'R_i' will denote the relations on X / E_I, as well as the relations on X). There are functions v_i and p_i satisfying (14) and (15) by Lemma 6 and (Debreu (1959)), respectively, and these induce functions from X / E_I to $\mathbb{R}$ that are compatible with E_i (resp. I_i) and strictly increasing with respect to

V_i (resp. P_i). The induced functions are continuous by Bourbaki (1965, I.3.4, Prop. 6). Thus the induced functions satisfy (14) and (15) relative to the induced relations on X / E_I. Therefore the relations R_i and W_i on X / E_i have closed graph. Now, for each i, define $X_i = X / E_i$. There is a canonical bijection between X / E_I and $\Pi_{j \in I} X_j$. By Lemma 9 and Bourbaki (1965, I.5.4, Prop 6 and corollary to Prop. 7), this bijection is a homeomorphism. Thus R_i and W_i induce relations with closed graph on $\Pi_{j \in I} X_j$. These induced relations also satisfy the other hypotheses of Lemma 8. The conclusion of this lemma is derived by composing the canonical surjection of X onto $\Pi_{j \in I} X_j$ with the functions from $\Pi_{j \in I} X_j$ to $\mathbb{R}$ that Lemma 8 asserts to exist. ∎

8. Publicity of cardinal intrapersonal interest comparisons

Conditions (23) and (24) still have to derived. In combination with (22), condition (23) states that, when one agent cares about the interest of another at all, then he strictly prefers an increase in the interest of that agent if the interests of others are held constant. Condition (24) states that agents' preferences reflect agreement about the cardinal magnitudes of intrapersonal interest differences between states. (Note that no form of interpersonal comparability is being assumed.) Thus, both of these conditions are economically significant ones.

The assumptions that have been made so far are insufficient to guarantee either of these conditions. However, both conditions can be derived if the assumption (26) that will be stated below is added to those already made. The assumption states (in the context of (4)) that agents' preference relations reflect agreement about the directions and relative magnitudes of pairs of interest differences, so evidently it is necessary for (22), (23), and (24) to hold. The statement of (26) will be simplified by the following lemma.

Lemma 11: If assumptions (2), (10), (11), and (22) hold, then $\forall i \; i \in N(i)$ and

$$j \in N(i) \iff \exists x \; \exists y \; [x P_i y \text{ and } \forall h \neq j \; x E_h y]. \tag{26}$$

Proof: If $j \notin N(i)$ and $\forall h \neq j \; x E_h y$, then (22) implies that $x I_i y$. Thus, if $x P_i y$ and $\forall h \neq j \; x E_h y$, then $j \in N(i)$. By (2) and (11), this implication entails that $i \in N(i)$. The converse implication follows from (15) and the nonconstancy and compatibility assertions of (22). If $j \in N(i)$, then there exist states w and z such that $u_{ij}(w) > u_{ij}(z)$. By (10), there are states x and y such that $x E_j w$, $y E_j z$, and $\forall h \neq j \; x E_h y$. These states therefore satisfy $x P_i y$, as well. ∎

Now the ordinal assumption regarding publicity of cardinal intrapersonal interest comparisons can be stated. For brevity, this assumption will be called *interest cardinality.*

> Let $N(i)$ be defined by (26). Suppose that $j \in N(h) \cap N(i)$. Then the following implication holds for all states w_h, w_i, x_h, x_i, y_h, y_i, z_h, and z_i satisfying $w_h E_j w_i$, $x_j E_j x_i$, $y_h E_j y_i$, $z_h E_j z_i$, and $\forall g \neq j \; [w_h E_g x_h, \; w_i E_g x_i, \; y_h E_g z_h$ and $y_i E_g z_i]$: if $w_h R_h y_h$, $z_h R_h x_h$, and $y_i R_i w_i$, then $z_i R_i x_i$. (27)

To understand this assumption, think of agents h and i as forming their preferences by weighing interest gains to agent j against interest losses to other agents. Agent h judges that j's gain moving from w_h to y_h would not outweigh others' losses but that j's gain moving from x_h to z_h would at least balance others' losses. Since others' interest is the same in w_h as in x_h and also in y_j as in z_h, h must judge that j gains at least as much by moving from x_h to z_h as by moving from w_h to y_h. If i shares this judgment, then he should also judge that j gains at least as much by moving from x_i to z_i as by moving from w_i to y_i, since j has the same interest in each of the 'i' states as in the corresponding 'h' state. If i also judges that j's gain moving from w_i to y_i is sufficient to balance others' losses, then he should judge as well that j's gain in moving from x_i to z_i is sufficient to balance others' losses, as the conclusion of (27) reflects.

Lemma 12: If (2), (10), (11), (22) and (27) hold and $j \in N(i)$, then u_{ij} is strictly increasing in V_j. That is, under these hypotheses (23) holds and $v_j = u_{ij}$ satisfies (14) for j.

Proof: First, in (27) substitute w for w_h and w_i, x for x_h and x_i, y for y_h and y_i, and z for z_h and z_i. Next substitute j for i, i for h, and x for w and y. The resulting statement is the implication that, if zR_ix and $\forall g \neq j\ xE_gz$, then zR_jx. Now, consider any states x and w such that xV_jw. By (10) there exists z such that wE_jz and $\forall g \neq j\ xE_gz$. Note that xV_jz, so that xP_jz by (2) and therefore xP_iz by (27) (using the substitutions that have just been made). Therefore by (22), $u_{ij}(x) > u_{ij}(z) = u_{ij}(w)$. ∎

Now the affine-relation condition (24) can be studied. The first step is to derive from (24) a condition that resembles (27) more closely.

Lemma 13: If (2), (10), (11), (22) and (27) hold and $j \in N(h) \cap N(i)$, but (24) does not hold, then there exist states x, y, and z such that xV_jy, yV_jz, and

$$[u_{hj}(y) - y_{hj}(z)] \,/\, [u_{hj}(x) - u_{hj}(z)] \neq [u_{ij}(y) - u_{ij}(z)] \,/\, [u_{ij}(x) - u_{ij}(z)]. \tag{28}$$

Proof: Consider any states r and s such that rV_js. Define a function $f : X \to I\!R$ by

$$f(t) = u_{ij}(s) + \{[u_{ij}(r) - u_{ij}(s)] \,/\, [u_{hj}(r) - u_{hj}(s)]\}\ [u_{hj}(t) - u_{hj}(s)].$$

Note that the denominator of the fraction in the expression is nonzero by Lemma 12, and that $f(t) = u_{ij}(t)$ for $t = r, s$. If (24) does not hold, then there must be some other state t such that $f(t) \neq u_{ij}(t)$, and it is impossible that tE_jr or tE_js. Let x, y, and z be r, s, and t arranged in order of decreasing interest afforded to j. The conclusion of the lemma is obtained by noting that

$$[u_{hj}(y) - u_{hj}(z)] \,/\, [u_{hj}(x) - u_{hj}(z)] = [f(y) - f(z)] \,/\, [f(x) - f(z)],$$

that $u_{ij}\ (t) \neq f(t)$ for exactly one of the values $t = x, y, z$, and that $[a - c] \,/\, [b - c]$ is strictly monotone in each of its variables on $\{(a, b, c) \,|\, a > b > c\}$. ∎

Lemma 14: Suppose that (2), (6), (10), (11) and (27) hold that $j \in N(h) \cap N(i)$, where h and i are distinct agents. Suppose that each of $N(h)$ and $N(i)$ contains an element distinct from j. Then (24) holds with respect to j.

Proof: Consider any three states x, y, and z such that xV_jy and yV_jz. To simplify notation, it will be assumed without loss of generality that j is neither h nor i.[10] Let $r_hV_hs_h$ and $r_iV_is_i$, as guaranteed by (11). Choose any natural number n sufficiently large so that

$$u_{hj}(x) - u_{hj}(z) < n\,[u_{hh}(r_h) - u_{hh}(s_h)] \text{ and } u_{ij}(x) - u_{ij}(z) < n\,[u_{ii}(r_i) - u_{ii}(s_i)]. \tag{29}$$

Because the functions u_{hh} and u_{ii} are continuous and X is connected, there exist t_h and t_i such that $u_{hj}(x) - u_{hj}(z) = n\,[u_{hh}(t_h) - u_{hh}(s_h)]$ and $u_{ij}(x) - u_{ij}(z) = n\,[u_{ii}(t_i) - u_{ii}(s_i)]$. Now, because u_{hj} is continuous and X is connected, there exists a function $\eta_n : \{0, 1, \ldots, n\} \to X$ such that, for $m \leq n$, $u_{hj}(\eta_n(m)) = u_{hj}(z) + m\,[u_{hj}(t_h) - u_{hj}(s_h)]$. (In particular, this implies that $\eta_n(0)E_jz$ and $\eta_n(n)E_jx$.) Assumption (27) will now be used to show that, for all $m \leq n$, $u_{ij}(\eta_n(m)) = u_{ij}(z) + m\,[u_{ij}(t_i) - u_{ij}(s_i)]$. Since $\eta_n(0)E_jz$ and $\eta_n(n)E_jx$, this is equivalent to
$\forall m < n\; u_{ij}(\eta_n(m+1)) - u_{ij}(\eta_n(m)) = u_{ij}(t_i) - u_{ij}(s_i)$. Suppose, to the contrary, that $u_{ij}(\eta_n(m+1)) - u_{ij}(\eta_n(m)) \neq u_{ij}(t_i) - u_{ij}(s_i)$ for some $m < n$. Without loss of generality, suppose that $u_{ij}(\eta_n(m+1)) - u_{ij}(\eta_n(m)) < u_{ij}(t_i) - u_{ij}(s_i)$. Then, for some p, $u_{ij}(\eta_n(p+1)) - u_{ij}(\eta_n(p)) > u_{ij}(t_i) - u_{ij}(s_i)$. (Otherwise $u_{ij}(z) - u_{ij}(x) = \sum_{k<n} [u_{ij}(\eta_n(k+1)) - u_{ij}(\eta_n(k))] < n\,[u_{ij}(t_i - u_{ij}(s_i)]$.) Now, using (10), states w_h, w_i, x_h, x_i, y_h, y_i, z_h, and z_i will be specified that satisfy the hypotheses of (27). Here, when a state is subscripted by 'g', the subscript takes both values h and i. The subscript 'f' takes all values except j and the value taken by g in the same expression. (That is, 'f' can always take any value except h, i, and j, and it can take whichever of $\{h, i\}$ is not taken by g.) Let q be a fixed, arbitrary state.

Suppose that $w_gE_j\eta_n(p)$, $x_gE_j\eta_n(m)$, $y_gE_j\eta_n(p+1)$, and $z_gE_j\eta_n(m+1)$; that $w_gE_gt_g$, $x_gE_gt_g$, $y_gE_gs_g$, and $z_gE_gs_g$; and that w_gE_fq, x_gE_fq, y_gE_fq, and z_gE_fq.

With respect to the states so specified, (27), fails to hold if (22) holds. Since this contradicts the hypothesis of the lemma, $\forall m \leq n\; u_{ij}(\eta_n(m)) = u_{ij}(z) + (m\,/\,n)\,[u_{ij}(x) - u_{ij}(z)]$.

For any n large enough to satisfy (29), there is a natural number $\mu_n < n$ satisfying $\eta_n(\mu_n+1)V_jy$ and $yW_j\eta_n(\mu_n)$. Thus both $[u_{hj}(y) - u_{hj}(z)]\,/\,[u_{hj}(x) - u_{hj}(z)]$ and $[u_{ij}(y) - u_{ij}(z)]\,/\,[u_{ij}(x) - u_{ij}(z)]$ are in the interval $[\mu_n\,/\,n,\ (u_n + 1)\,/\,n)$. Since n can be arbitrarily large, $[u_{hj}(y) - u_{hj}(z)]\,/\,[u_{hj}(x) - u_{hj}(z)] = [u_{ij}(y) - u_{ij}(z)]\,/\,[u_{ij}(x) - u_{ij}(z)]$. Therefore (24) holds, by Lemma 13. ∎

Lemma 15: If conditions (2), (6), (10), (11), (22) and (27) hold, then condition (24) holds for all agents h, i, and j.

Proof: In view of Lemma 14, (24) only has to be established now in the case that $h = j$, $N(j) = \{j\}$ and $j \in N(i)$ for some $i \neq j$. In that case, simply take $u_{hj} = u_{ij}$ for some i satisfying $j \in N(i)$. ∎

The foregoing lemmas immediately establish the following theorem.

---

[10] This entails that h and i are elements of $N(h)$ and $N(i)$, respectively, that are distinct from j. In general, agents $h' \in N(h)$ and $i' \in N(i)$ will have to be specified that are distinct from j.

Theorem 2: Suppose that assumptions (1)–(6), (10)–(13), (11), (25) and (27) are satisfied, and that either (7) is satisfied or else X possesses the cartesian-product structure described in Lemma 7. Then Pareto superiority and liberal succession coincide.

9. An alternative cardinality condition on preferences

Within many communities there is wide, if approximate agreement, regarding cardinal intrapersonal comparisons of interest. It is uncontroversial that someone's interest would be more significantly advanced if he were to move from a hovel to a decent house than if he were to move from the house to a mansion, for example. The prevalence of this kind of shared intuition about welfare contributes to the plausibility of condition (27).

However, some people may find it less plausible that there is public agreement about intrapersonal cardinal interest comparisons, than that there is public agreement about intrapersonal cardinal preference comparisons.[11] For one thing, it is often more obvious what people prefer than what is objectively good for them. For another, people who endorse expected-utility theory are already committed to accept one form of the cardinality of persons' preferences, even if they do not accept the cardinality of persons' interests. People who hold these views may prefer to assume the publicity of intrapersonal cardinal preference comparisons rather than to assume (27). It will be shown here that such an assumption can be formulated in a way similar to (27), and that the formal assumption (together with the other axioms) implies the existence of common additive interest factors.

Consider how to express the agreement of two agents, j and k, about intrapersonal cardinal preference comparisons concerning agent i. Suppose that $\forall h \notin \{i, j\}$ $[wI_hy$ and $xI_hz]$, and that wE_jy and xE_jz. Make an analogous assumption with k substituted for j, but with respect to a different set of social states. Specifically, suppose that $\forall h \notin \{i, k\}$ $[w'I_hy'$ and $x'I_hz']$, and that $w'E_jy'$ and $x'E_jz'$. Let i be indifferent between the corresponding social states mentioned in these two assumptions. That is, let wI_iw', xI_ix', yI_iy', and zI_iz'.

Now suppose that xR_jw, yR_jz, and $w'R_kx'$. By reasoning analogous to that regarding (27), the two preferences of j imply that j must regard i's interest gain from being in x rather than w as being at least as great as i's interest gain from being in z rather than y would be. The preference of k implies that k must regard i's interest gain from being in x' rather than w' as being insufficient to outweigh strictly the net disadvantages of x' relative to w' for everyone else. (Note that these will be in terms of preference for everyone except k, and in terms of interest for k.) Thus, if k agrees with j about intrapersonal cardinal preference comparisons regarding i, then k should also regard i's interest gain from being in z' rather than y' as being insufficient to outweigh strictly the net disadvantages of z' relative to y' for everyone else. That is, it should be the case that $y'R_kz'$.

The implication from these various assumptions to their conclusion jointly constitute the assumption that *intrapersonal cardinal preference comparisons are public:*

[11] I am indebted to Arthur Robson for stating the two considerations that I discuss now.

Define $j \in S(i)$ if and only if, for some states x and y, xE_iy, $\forall h \notin \{i, j\}$ xI_hy, and xP_hy. If $j \in S(h) \cap S(i)$, then following implications holds for all states w_h, w_i, x_h, x_i, y_h, y_i, z_h, and z_i satisfying $w_hI_jw_i$, $x_hI_jx_i$, $y_hI_jy_i$, $z_hI_jz_i$, $\forall g \notin \{h, j\}$ $[w_hI_gx_h$ and $y_hI_gz_h]$, $\forall g \notin \{i, j\}$ $[w_iI_gx_i$ and $y_iI_gz_i]$, $w_hE_hx_h$, $y_hE_hz_h$, $w_iE_ix_i$, and $y_iE_iz_i$: if $w_hR_hy_h$, $z_hR_hx_h$, and $y_iR_iw_i$, then $z_iR_ix_i$. (30)

It seems very plausible that this condition can play an analogous role to (27) in guaranteeing the coincidence of Pareto superiority and liberal succession on the basis of ordinal and topological assumptions.

10. Coincidence without separability or publicity

Theorem 2 and the alternative just suggested each hypothesize (a) product structure of interests and separability of preferences in interests, and (b) public agreement regarding some form of intrapersonal cardinal comparison. The standard construction of nonpaternalistic preferences, by beginning with cardinal interest-representation functions as in Theorem 1, yields a family of preferences that satisfy all of these qualitative hypotheses. Although the example provided in section 3 has shown that the hypotheses of interest-determination and nonpaternalism alone are too broad to characterize the class of nonunanimous preference profiles on which the liberal principle and unanimity coincide, it might well be suspected that the additional qualitative hypotheses would prove to be necessary as well as sufficient for this coincidence. A counterexample that disproves this conjecture is now provided.[12]

Let there be three agents, $I = \{1, 2, 3\}$, and let $X = \mathbb{R}^3_+$. Define

$$v_i(x) = x_i, \tag{31}$$

and define

$$p_i(x) = \min\{x_i/2,\ x_j,\ x_k\}, \text{ where } I = \{i, j, k\}. \tag{32}$$

Then define W_i and R_i as in (14) and (15).

It is evident from (31) and (32) that agents' preferences are determined by interests, and that the agents are not unanimous. It will now be shown that their preferences are also nonpaternalistic and that the relation of liberal succession coincides with that of Pareto improvement. Since the latter claim implies the former, only the coincidence of the two welfare orderings needs to be

[12] This example will involve preferences that are represented by utility functions defined by taking the minimum of several numbers. Such utility functions are pointwise limits of sequences of CES utility functions. Every CES utility function represents a preference relation satisfying the separability assumption. Thus, despite the example there may still be a close connection between separability and the coincidence of the two welfare relations.

shown. Recall that every Pareto improvement is automatically a liberal successor, so it is sufficient to establish the converse inclusion.

Suppose, then, that y is a liberal successor of x. Specifically, suppose that

$$\forall i \in C \; p_i(y) \geq p_i(x), \tag{33}$$

that

$$\forall j \notin C \; y \geq x, \tag{34}$$

and that

$$p_k(x) > p_k(y). \tag{35}$$

By (35), y is not a Pareto improvement over x. This will be shown to be a contradiction.

By (32) and (35),

$$\min \{y_i, y_j, y_{k/2}\} < \min \{x_i, x_j, x_{k/2}\}. \tag{36}$$

By (32) and (35),$k \notin C$, so $y_k \geq x_k$ by (34). Therefore (36) implies that, for one of the other agents $h \in \{i, j\}$,

$$y_h = \min \{y_i, y_j, y_{k/2}\} < \min \{x_i, x_j, x_{k/2}\} \leq x_h. \tag{37}$$

Therefore $h \in C$ by (34). But, by (32) and (37),

$$p_h(y) \leq y_{h/2} < \min \{x_{h/2}, x_i, x_j, x_k\} = p_h(x). \tag{38}$$

Therefore $h \notin C$ by (33), so a contradiction has been reached. This establishes that liberal succession coincides with Pareto improvement, and consequently that all agents have nonpaternalistic preferences.

However, agents' preferences are not separable in interests, and cardinal interpersonal comparisons of neither preferences nor interests are public. To see that the preferences of agent 1 are inseparable, let $J = \{1\}$ and $K = \{2, 3\}$, and define $w = (1, 2, 2)$, $x = (2, 2, 2)$, $y = (1, 0, 0)$, and $z = (2, 0, 0)$. Then (4) fails to hold because, although its hypothesis holds, xP_1w while yR_1z. By the symmetry of the example, the preferences of the other two agents are not separable in interests either.

To show that intrapersonal cardinal interest comparisons are not public, let $i = 1$ and $j = 2$ in (27). Define $w = z = (3, 2, 2)$ and $x = y = (2, 2, 2)$. Then the hypothesis of (27) holds, but wR_1x

and not yR_1z, so the conclusion of (27) fails to hold. Therefore (27) (which is an implication) fails to hold.

Intrapersonal cardinal preference comparisons can similarly be shown not to be public. To do so, let $i = 1$, $j = 2$, and $k = 3$ in (30), and define:

$$w = z = (3, 3, 3), \qquad x = y = (2, 3, 3),$$
$$w' = z' = (3, 3, 7), \qquad x' = y' = (2, 3, 7).$$

These social states have the feature that agents 1 and 2 always have "selfish" preferences among them (that is, the preference and interest relations of each agent among these states coincide), and that agent 3 has "selfish" preferences among $\{w, x, y, z\}$ but has totally "altruistic" preferences for the interest of agent 1 among $\{w', x', y', z'\}$. From these considerations, it can easily be seen that (30) fails to hold.

11. Examples and conclusion

Strictly speaking, Mill's principle concerns only the permissibility of private activities. Contemporary libertarians advocate that public authorities should not have the power to take any action that Mill's principle would prohibit to the coalition of persons who prefer the action. Such a position obviously has strong distributive implications. An alternative view would hold that public authorities can legitimately take some actions that would be prohibited to private persons and coalitions, but would also recognize that public actions should be structured in a way that will minimize "transactions costs." On this view, the public authorities should not take an action if some person or coalition can propose an alternative action that the liberal principle would endorse as a replacement for the initially contemplated action. Such a view implicitly defines a notion of "liberal efficiency" that is analogous to the familiar concept of Pareto efficiency.

On this latter normative view about government action, there is a broad scope for welfare analysis but the familiar Paretian analysis is foundationally inadequate. There are several issues in public finance, with respect to which such a view seems to be widely held. One of these issues is whether redistributive policy ought to be implemented by cash transfers or by transfers in kind (e.g., provision of public housing). One of the common arguments in favor of cash transfers is that it is demeaning to the recipients of transfers not to be granted autonomy over the use of the resources that society is willing to transfer.[13] This argument makes an implicit appeal to the liberal principle. It is clear that some of the proponents of the argument regard it as being conceptually distinct from the argument that transfers in kind are inefficient, although those proponents (especially when they are economists) typically produce the latter argument as a corollary. It might be thought that there could be transfer programs with respect to which the argument about recipients' autonomy could be raised although cash transfers would not be Pareto superior to transfers in kind. The results of this paper indicate the special features that such a program would have to possess. In cases where

[13] Conversely, one of the arguments sometimes made in favor of transfers in kind is that some classes of recipients (e.g., addicts) are incapable of exercising such autonomy.

such features are absent, an efficient transfer never violates the preferences of recipients about how the resources dedicated to them should be used.

It would be widely agreed that the assumptions about interests and preferences that have been discussed here are reasonable ones to make in the context of an evaluation of transfers to competent adults. The one assumption from which there would possibly be substantial dissent is the additive separability of preferences in interests. As was shown in the preceding section, this assumption will not be satisfied if people's preferences reflect maximin considerations regarding the satisfaction of interests. However, the example studied in that section makes it plausible that the results proved here are robust to this specific kind of failure of the separability assumption. This is a matter for further research. If these results are indeed robust, then it will be fair to conclude that the welfare evaluation of transfers to competent adults can safely focus on efficiency questions to the exclusion of questions about liberal choice.[14]

Another issue concerns the regulation of markets for services such as education or medical care. Because these services are differentiated products, it is conceivably efficient to limit the number of product varieties in order to take advantage of economies of scale. Suppose, for simplicity, that any feasible allocation involving production of more than one product variety is Pareto dominated by some feasible allocation in which only a single variety is produced. Thus, according to either the efficiency criterion or the liberal criterion, only an allocation with a single variety can be optimal. The question is, which varieties may be produced in optimal allocations? Parallel to the case of transfers just discussed, the answer to this question is the same for both criteria if the assumptions studied in this paper hold. Also parallel to that case, the appropriateness of separability and cardinality assumptions is open to question on account of features of the situation to which the results of this paper plausibly are robust. However, in the case of differentiated-product allocation, there are also intrinsic difficulties about the publicity of intrapersonal cardinal interest comparisons. That is, the failure of the cardinality assumption to hold may not be simply a by-product of a violation of the separability assumption.

To understand the nature of these possible intrinsic difficulties, consider some of the troublesome and divisive questions about resource allocation in education and medicine. How much emphasis should schools give to the development of students' skills in reading and arithmetic, versus the development of critical thinking and appreciation of the arts? How much of medical research funding should be spent on finding cures for rare diseases that strike people early in life, versus common diseases of older people? Debates about these questions have to do partly with distributional issues, but they also reflect pronounced differences in the participants' views about the relative importance of disparate elements of the good life. That is, they reflect the fact that participants are (at least implicitly) expressing cardinal judgments about the components of persons' interests, and that they disagree about those judgments although their ordinal judgments may be unanimous. It

[14] It might still be argued that there is something ethically preferable about letting recipients make their own choices, rather than making choices for them that are consistent with their preferences. Nevertheless, this argument does not seem as compelling as an argument that people's preferences about matters that are their own business are actually being thwarted. For example, most people would judge that rather small reductions of administrative cost constitute sufficient grounds to justify centralized allocation if the result is consistent with recipients' preferences.

is not at all clear whether the results of this paper can be extended to cover situations that possess this complication.

These examples illuminate the philosophical significance of the technical results that have been derived here. These results constitute a limited defense of the study of the unanimity relation in welfare analysis to address concerns that may fundamentally be about respect for persons' objective interests. This defense acknowledges the legitimacy of these concerns, but argues that they actually are addressed, even though implicitly, by the technically simpler efficiency analysis. Following Mill, a theoretical concept of a person's non-preferential interest has been introduced in order to relate the specification of rights and liberties to the specification of preferences. While there may be many possible ways to give a substantive definition of this concept, only a few qualitative assumptions about the concept are needed in order to make the defense. For the most part, these assumptions seem to be appropriate for the discussion of distributional issues and of other issues with which applied welfare analysis typically deals. A formal example has shown that the assumption that is likely to be regarded as the most restrictive one, separability, may be stronger than is actually needed. The two economic examples that have just been discussed show that, even if this conjecture is true, the limited defense of the unanimity criterion falls short of showing that explicit consideration of the liberal criterion would be completely dispensable for applied welfare analysis.

References

Bergstrom, T., (1971), "Interrelated Consumer Preferences and Voluntary Exchange," *Papers in Quantitative Economics*, vol. I, A. Zarley, ed., University Press of Kansas: Lawrence, KS.

Bergstrom, T., (1988), "Systems of Benevolent Utility Interdependence," University of Michigan, working paper.

Bergstrom, T., (1989), "Puzzles: Love and Spaghetti, the Opportunity Cost of Virtue," *Journal of Economic Perspectives,* **3(2)**, 165-173.

Bourbaki, N., (1965), *Topologie Generale*, Hermann: Paris, France.

Buchanan, J., and G. Tullock, (1962), *The Calculus of Consent,* University of Michigan Press: Ann Arbor, MI.

Debreu, G., (1959), *Theory of Value,* Yale University Press: New Haven, CT.

Debreu, G., (1960), "Topological Methods in Cardinal Utility Theory," in *Symposium on Mathematical Methods in the Social Sciences*, Arrow and Suppes, eds., Stanford University Press: Stanford, CA.

Green, E., (1979), "A Welfare Theory for Nonmalevolent Agents," Princeton University, working paper.

Green, E., (1982), "Equilibrium and Efficiency Under Pure Entitlement Systems," *Public Choice,* **39**, 185-212.

Krantz, D., R.D. Luce, P. Suppes, and A. Tversky, (1971), *Foundations of Measurement,* Academic Press: New York, NY.

Mill, J.S., (1859), *On Liberty,* Parker: London, England.

Pareto, V., (1909), *Manual of Political Economy*, Second Edition, translated by A. Schwier and edited by A. Schwier and A. Page, reprinted by Augustus M. Kelley: New York, NY, 1971. (The translation is of the fifth edition (1927) which is substantially identical to the second edition.)

Pearce, D., (1983), "Two Essays in the Theory of Strategic Behaviour," Princeton University, Ph.D. dissertation.

Riley, J., (1988), *Liberal Utilitarianism: Social Choice Theory and J.S. Mill's Philosophy,* Cambridge University Press: New York, NY.

Rockafellar, R.T., (1970), *Convex Analysis,* Princeton University Press: Princeton, NJ.

Schick, F., (1980), "Towards a Logic of Liberalism," *Journal of Philosophy,* **77**, 80-98.

Sen, A.K., (1970), "The Impossibility of a Paretian Liberal," *Journal of Political Economy,* **78**, 152-157.

Wicksell, K., (1935), *Lectures on Political Economy,* Routledge and Sons: London, England.