9. Deducing interpersonal comparisons from local expertise

IGNACIO ORTUÑO-ORTIN AND JOHN E. ROEMER

1. Introduction

Economists accept the idea that a person can have a coherent ordering over the states of the world; yet it is commonplace to balk at the notion that there exists a coherent interpersonal ordering, which would give sense to statements of the form 'person i is better off in state x than person j is in state y.' The reason for such skepticism is that whereas in the first case one mind is making judgments about states of the world, there is no universal mind that can make interpersonal judgments. Nevertheless, most of us feel capable of making some interpersonal comparisons, perhaps by virtue of the limited empathy we feel, because we believe at some level all people are relevantly similar. We will argue that it may be quite reasonable to suppose the existence of an interpersonal ordering of the states of the world, based on a kind of empathy that a person can legitimately feel, because he has, during his life, indeed been a person of various different types.

Interpersonally comparable utility has had a checkered history. In the nineteenth century (see Cooter and Rappoport (1986)), the possibility of interpersonal comparisons was taken for granted by many social theorists. The ordinalist revolution dissolved this innocent presumption; its supporters claimed that interpersonal comparisons were necessarily normative, hence not within the purview of positive economics (see also Sen (1979) for a discussion).[1] There are, it seems, two different bases for the current agnosticism, or rather nihilism,

This research was partially supported by a grant from the National Science Foundation. We are also indebted to David Donaldson for his comments.

[1] Although interpersonal comparisons may be made for normative *reasons*, the comparisons themselves may be matters of fact. We wish to argue that this may be so, quite generally.

with respect to the existence of an interpersonal ordering. First, positive economics (including general equilibrium theory) does not require interpersonal comparability. Hence, parsimony suggests that such information not be assumed. Second, it is widely believed that to assume interpersonal comparability presupposes some kind of supraperson authority who makes the decisions that no individual can make. Such authoritarian decisions would have either an objective or a dictatorial quality that would cut against the grain of the twentieth-century subjective approach to preference.

But the necessity of establishing foundations for interpersonal comparability need hardly be mentioned if one is interested in social choice and distributive justice. Without interpersonal comparability, one can hardly move beyond Pareto optimality as a social criterion for evaluating alternative states, but with it social choices can be made (see, for example, Blackorby, Donaldson, and Weymark (1984)). In the late 1960s and throughout the 1970s, attempts were made to move beyond Pareto optimality without imposing an assumption of interpersonal comparability with the development of the notion of fairness (envy-free, Pareto efficient allocations) and the related notions of egalitarian equivalent allocations and fair net trades[2] (see Foley (1967), Kolm (1972), and Thomson and Varian (1986) for a survey of this literature). Fairness, so defined, however, does not reach very far in resolving questions of distributive justice. Fair allocations do not always exist; but more importantly, when the distribution of internal traits of persons becomes a topic for distributive justice – and this is central to the contemporary theories of Rawls (1971), Dworkin (1981), and Sen (1981) – then fairness becomes an almost useless concept (see Roemer (1985) for an explanation).

The tension between the necessity for positing interpersonal comparability in order to make progress on questions of social choice and distributive justice, and the agnosticism with respect to the possibility of making interpersonal judgments in an objective way is seen, for example, in the following quotation from Arrow:

> In a way that I cannot articulate and am not too sure about defending, the autonomy of individuals, an element of mutual incommen-

[2] See, however, the introduction of Kolm (1972), in which the author writes that the concept of fairness was discussed by J. Tinbergen in 1953.

> surability among people, seems denied by the possibility of interpersonal comparisons. No doubt it is some such feeling as this that has made me so reluctant to shift from pure ordinalism, despite my desire to seek a basis for a theory of justice (Arrow (1977, p. 225)).

We hope to chip away at this incommensurability; in particular, to show that a supra-personal authority is not necessary, but that people can be expected to make interpersonal comparisons themselves by combining their individual judgments, based on local expertise.

2. Local expertise

Let X denote the set of states of the world over which a person has a preference ordering. If we had to design an experiment to deduce the person's preferences, we would probably ask him to rank the alternatives in small subsets of X. In carrying out this revealed preference experiment, it is likely that inconsistent answers will be given. Most persons, when confronted with an agenda of such requests, will produce intransitivities, if X is large enough. (We assert this as a piece of conventional wisdom.) But faced with such demonstrated inconsistencies, we do not declare the incoherence of the notion of intrapersonal comparability. We are prone to say, instead, that the person has made a mistake or that he has bounded rationality; the ideal of an intrapersonal ordering remains acceptable.

We suggest that such errors are made in the revealed preference experiment because the person does not have sufficient remembered experience of all the states he has been asked to rank.[3] Some of the states are distant from his personal experience, so he does not have a good basis on which to rank them against some other states. Perhaps it would be appropriate, if this were the case, to say that he really has incomplete preferences over X.

We propose to extend the charity that we show in assuming people have coherent intrapersonal orderings, despite evidence to the contrary, to the ideal of interpersonal comparability.[4] Let T be the set of

[3] This differs from May's (1954) proposal. If a person has n orderings of the states, each according to one of n criteria that are important to him, there will not in general exist an overall ordering that satisfactorily aggregates the n criteria-specific orderings. This is an application of Arrow's impossibility theorem.

[4] But see Gibbard (1986) for a discussion of the difficulties in constructing a coherent intrapersonal ordering.

types of persons. The information summarizing a person's type is sufficient to determine his ordinal preferences. A type is a long vector, some components of which describe salient aspects of a person's history and, perhaps, his biochemical and genetic makeup. We will assume that a distinction can be made between the characteristics that determine type, and the attributes of the social alternatives, or states, that comprise X – an important and perhaps controversial assumption. An interpersonal ordering of X is an ordering of the set $X \times T$, interpreted as follows: $(x, t) \gtrsim (y, s)$ means a person of type t is at least as well off in state x as a person of type s is in state y.

Suppose we conduct an experiment in which we ask different people to rank subsets of $X \times T$. If everyone agreed on these rankings, we might feel confident in asserting the coherence of an interpersonal ordering. But in all likelihood there will be disagreements among people, even when they are posed the problem of comparing some pairs of states in $X \times T$. We suggest that two people, say of types i and j, disagree about the ranking of (x, s) and (y, t) for essentially the same reason that one person commits inconsistencies in his intrapersonal ordering of X. At least one of the *positions* (x, s) or (y, t) is too distant from the personal experience of i or j. This might be so either because, say, x is too far from the states in X that the type i person has experienced, or because a person of type s is too far from a person of type i. Thus, the person in question cannot be considered a competent judge of interpersonal comparisons when his experience – say (z, i) – is 'too far' from (x, s) or (y, t).

Just as each person has had experience with different states in X, it is the case that each person has had experience with different types in T. He himself has been different types. Some of the characteristics that define a type vary with personal experience – age, health, and wealth. Thus each person has traveled through some, perhaps small, subset of the set of types T. If we are willing to assume that a person has a coherent intrapersonal ordering on X, then we should be willing to assume he is capable of providing an *interpersonal* ordering on $X \times T_i$, where T_i is the subset of types in his remembered experience. Not only can the person who is currently of type i report the intrapersonal orderings (of X) for types in the set T_i; he can, as well, make intertype comparisons.

(We could say that a particular person has experienced types i, j, k and that he experienced subsets of X denoted X_i, X_j, X_k, respectively, while of those types. He would then be capable of providing an interpersonal ordering only on the set $(X_i \times \{i\} \cup (X_j \times \{j\}) \cup (X_k \times \{k\})$. But our goal is to establish the cogency of interpersonal comparability given the cogency of intrapersonal comparability. Therefore we assume that a person has a complete order of X for each type he has experienced.)

That is, a person is assumed to have accurate memory, or one mind, that can unify the perceptions he has had as his type has varied, and so he is capable not only of rendering accurate intrapersonal orderings for each type he has experienced, but of recalling the interpersonal ordering among these types: "I was happier living in that dump as a student than I am in this palace in middle age." These comparisons are not ones made from the point of view of his current type, but are of his actual experienced welfare levels at the two positions in question. In fact, we will not require people to remember how they felt many years ago; it will be sufficient if they can remember their feelings only for types very 'close' to their present type, which is to say, types they have experienced in the recent past.

The assumption suggested by this discussion is that each person is competent to make interpersonal comparisons – or, more accurately, intertype comparisons – on some neighborhood in state-type space of the point at which he is currently situated. It follows that if the neighborhoods of competence of two people intersect, then the interpersonal orderings on the intersection must agree, because they have both experienced the positions in the intersection.[5] This is so, in particular, because those comparisons are not made from the point of view of one's current position.

In the next section we present a simple model in which this condition suffices to determine a partial ordering on the space $X \times T$. We use the opinion of local experts to piece together a consistent (but perhaps incomplete) interpersonal ordering of $X \times T$.

[5] What if two people, each of whom is putatively a local expert in regard to two positions (x, s) and (y, t) disagree on their ordering? Then we must say that the space of types has not been sufficiently disaggregated to distinguish properly between these two persons and between types s and t.

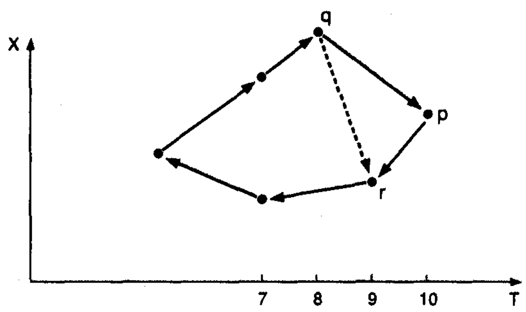

Figure 9.1

3. A simple model

In this model, it will be assumed that the set of types can be represented by a discrete one-dimensional set. We identify the types with integers. The set X is an abstract set. A position in $X \times T$ is schematically represented by a point in the plane in Figure 9.1 whose first coordinate is an integer. It is assumed that each type i has experienced one neighboring type on each side. Thus, the set $T_i = \{i - 1, i, i + 1\}$. We assume that the person of type i has an ordering on the set $X \times T_i$. Agents of types i and j agree on the interpersonal ordering on the intersection $X \times (T_i \cap T_j)$.

By linking together the judgments of individuals of neighboring types, we induce a partial binary relation on $X \times T$. The question is whether this procedure will be consistent, or whether it will generate intransitivities. Under the earlier assumptions, this procedure leads to no intransitivities.

Denote i's preference ordering on $X \times T$ by $>_i$. We only respect his ordering on his domain of competence, $X \times T_i$. Let $p, q \in X \times T_i$, and suppose $p <_i q$. If $q, r \in X \times T_j$ and $q <_j r$, then we define $p < r$, where $<$ represents the interpersonal ordering under construction. Suppose there is a cycle under this procedure, as illustrated in Figure 9.6. We draw an arrow $p \rightarrow q$ to indicate that $p <_i q$ for some i.[6]

If there is a cycle, there is one involving a smallest number of types. Consider such a minimal cycle. We derive a contradiction by showing

[6] $p <_i q$ means $p \lesssim_i q$ and p and q are not indifferent. (Indifference is the conjunction of $p \lesssim_i q$ and $q \lesssim_i p$.)

that a cycle can be constructed that does not involve the agent on the right-hand extreme, position p in Figure 9.1. There are only several possibilities for what the cycle looks like near p. One is illustrated in Figure 9.1, where the position q immediately inferior to p is a position of type 8, and the position r, which is immediately superior to p, is a position of type 9. But then $r, q \in X \times T_8$, and so it must be that $r >_8 q$, or else there would be smaller cycle created among r, p, q, which is impossible, for all three positions lie within the local ordering of 8. But if $r >_8 q$, then agent 10 can be removed from the cycle, as the dotted line indicates, which completes the argument. There are several other possible configurations for p, q, and r, but the same argument works. Hence the procedure for aggregating the opinion of local experts works.

Note, first, that this procedure does not necessarily lead to a complete order on $X \times T$. Second, this piecing-together procedure for deducing interpersonal comparability does not verify the conventional wisdom that if everyone agrees on the order of two positions, then that must be the correct interpersonal order. Although it may be the case that everybody believes that (x, i) is better than (y, j), we might deduce that $(x, i) \lesssim (y, j)$. No one person may be competent to make judgments comparing these two positions, and when the opinions of local experts are linked together, the opposite conclusion may hold. Thus, we do not concur with the conventional wisdom that universal agreement about the ordering of two positions is sufficient grounds for concluding that that is the correct interpersonal ordering. We trust only the opinions of people who are competent to judge.

We proceed to show how a complete interpersonal ordering can be deduced from the opinions of local experts when the set of types is a discrete, n-dimensional set.

4. Interpersonal comparability on a lattice

We suppose now that the set of types T can be represented as the points in an n-dimensional rectangular lattice. Each dimension is interpreted as one of a set of traits, which together characterize a type. The set X, as before, is any abstract set.

We work with a two-dimensional lattice T, although the definitions and theorems are general for a lattice of any finite dimension. A *type* in T is denoted (i, j), after its integer coordinates. We postulate that a person located at any type has experienced, as well, the four closest types in the lattice.[7] Thus, his neighborhood of competence, in type space, is:

$$T_{i,j} = \{(i, j), (i - 1, j), (i + 1, j), (i, j - 1), (i, j + 1)\}$$

It is postulated that (i, j) has an ordering on $X \times T_{i,j}$, which will be denoted $\gtrsim_{i,j}$. Furthermore, it is postulated that:
Axiom of Coincidence The orderings $\gtrsim_{i,j}$ and $\gtrsim_{k,l}$ agree on $X \times (T_{i,j} \cap T_{k,l})$.

We furthermore postulate:
Axiom of Continuity. If $(x; k, l) \epsilon X \times T_{i,j}$, then there exists $(y; i, j) \epsilon X \times \{(i, j)\}$ such that $(x; k, l)\ _{i,j}\ (y; i, j)$.

We call this an axiom of continuity because it is plausible if the types that are neighbors in the lattice are 'close' to each other, in a psychological sense, and if the set X is sufficiently dense that this kind of indifference curve can be drawn as the types vary. The Axiom of Continuity is perhaps only plausible if X is a continuum, such as the set of all possible allocations of some continuously divisible set of commodities, although the theory does not require X be such a set.
Theorem 1. Let T be an n-dimensional square lattice and let $t = (i_1, i_2, \ldots, i_n)$ represent an arbitrary type in the lattice. Let the neighborhood of t, denoted T_t, consist of the $2n$ points of unit distance from t, plus t itself. Let X be any set, and let $\gtrsim_t$ be a complete order of $X \times T_t$, for every t. Suppose the axioms of coincidence and continuity hold. Then the local orderings $\gtrsim_t$ can be extended to a complete ordering of $X \times T$ in a unique way.

Proof: Section 6.

One might object that it is unreasonable to suppose that there is an order $\gtrsim_t$ associated with every point in the lattice. Perhaps there are persons associated only with some proper subset of types on the

[7] This condition may seem too strong. Our age, for example, is now the maximum we have ever had, so we should not postulate that we have experienced ages older than we are. This difficulty is easily resolved: The type where a person is located in the lattice does not need to be his present type, only a type he has experienced in the recent past.

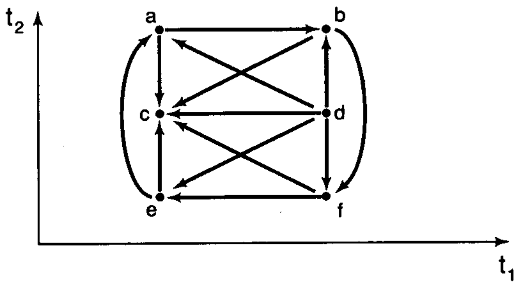

Figure 9.2

lattice. In this case, it is still possible to recover a partial ordering of $X \times T$.

Without the axiom of continuity it is not difficult to find examples of intransitivities. Let a, b, c, d, e, f be six different types in a two-dimensional lattice $T_{i,j}$ (see Figure 9.2). Suppose that the set of states of the world contains just one element: $X = \{x\}$. Call $A = (x, a) \in X \times T$, $B = (x, b) \in X \times T$, and so on. We now write $A \gtrsim_k B$ for $(x, a) \gtrsim_k (x, b)$ for $k \in T$. Let the orderings for the six types in each of their neighborhoods of competence be as follows:

$C >_a B >_a A$
$B >_b A >_b D$
$C >_c A >_c E >_c D$
$C >_d F >_d B >_d D$
$C >_e E >_e F$
$E >_f F >_f D$

It is easy to check that the axiom of coincidence holds, but we can form the intransitivity $A <_b B <_d F <_f E <_c A$. This is illustrated in Figure 9.2, where an arrow from a to b means $B >_a A$, and so on.

As a corollary to Theorem 1, a similar result follows if $T = R^n$.

The theorem remains true for a continuum of types. For this we need:

Generalized Axiom of Continuity. Let T and X be sets, and for each $p \in T$, let $T_p \subset T$, and let $\gtrsim_p$ be an order on $X \times T_p$. $\gtrsim_p$ satisfies the axiom of continuity if, for all $q \in T_p$, $x \in X$ *there exists* $y \in X$ *such that* $(x, p) \sim_p (y, q)$.

Theorem 2. Let $T = R^n$, X be any set, and ε be a positive number. Let T_p be an arc-connected neighborhood of p containing a ball of radius at least ϵ about p. For all $p \in T$, let $\gtrsim_p$ be an order on $X \times T_p$. Suppose the axioms of coincidence and continuity hold. Then the local orders extend uniquely to a complete order of $X \times T$.

Proof: Available from authors.

Remark. An alternative model for our problem does not distinguish between the states X and the types T but postulates a set Y – say, a rectangular lattice of many dimensions – whose members are identified with positions (in some state-type space). Each point in Y specifies everything about a person, where his type is not distinguished from the state. Associated with each $y \in Y$ is an ordering $>_y$ of a small neighborhood of y. The axiom of coincidence is postulated. The axiom of continuity, however, no longer makes sense because type and state cannot be distinguished. What other conditions on the local orders are sufficient to guarantee that the induced binary relation on Y is an order? The problem is trivial if $n = 1$; there is a strong condition that suffices for $n = 2$; but the problem becomes very difficult at $n = 3$. The advantage of the approach we have taken – of distinguishing states from types – is that the problem becomes tractable because dimension no longer plays a critical role. Hence the assumption that type can be distinguished from state is perhaps the most important, and contentious, assumption of the model.

5. Conclusion

Our theorems provide a basis for legitimating a belief in interpersonal comparability, if the idea of local expertise is accepted. If each person is competent to make interpersonal judgments locally, among positions occupied by types close to his own, then a unique complete order on state-type space exists, which is the transitive extension of the local orders. We interpret this global ordering as *the* interpersonal ordering. More accurately, we can say that *if* an interpersonal ordering exists, this must be it, for it is the unique order that coincides with all the orderings of local experts. It may still be objected that no interpersonal ordering exists, and so the order whose existence we have proved has no significance – other than being the transitive extension of local

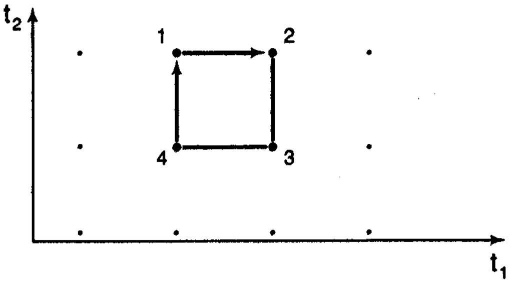

Figure 9.3

orders. But we find this nihilism unconvincing. It would be more convincing to argue that we have merely reduced the problem of interpersonal comparability to one of communicating one's type. As Arrow[8] has said, that may be the insurmountable problem.

6. Proofs

Proof of Theorem 1:

We prove the theorem for the two-dimensional lattice $T = N \times N$, where N is the non-negative integers. T_{ij} is the five-point neighborhood of (i, j). Let $I(y^1)$ be an intransitivity $y^1 \gtrsim y^2 \gtrsim \ldots \gtrsim y^n$, $I(y^1) \subset$ X × T and for all $m \in \{1, 2, \ldots, n\}$, y^m and y^{m+1} belong to $X \times T_{ij}$ for some $(i, j) \in N \times N$. We write the mth term in $I(y^1)$ as y^m or $(x^m; i^m, j^m)$.

The proofs of this theorem and Theorem 2 are easier if we work with the projection of $I(y^1)$ on T: We associate each point in $X \times T$ with its coordinate in T and we relate these points in T in such a way that $(i, j) \gtrsim (i', j')$ if there exists a pair $(y^p, y^{p+1}) \subset I(y^1)$ such that $y^p = (x, i, j) \gtrsim (x', i', j') = y^{p+1}$. This projection of $I(y^1)$ on T produces an intransitivity on T.

Lemma 1 There cannot be an intransitivity $I(y^1)$ such that its projection on T, $I_T(y^1)$, is a square of area equal to one.

Proof. See Figure 9.3, where an arrow (or undirected segment) from a to b means $a < b$ (or $a \sim b$). Note that $\{4, 2\} \in T_1$; therefore type 1

Figure 9.1

Figure 9.2

Figure 9.3

says that $2 > 4$ – that is, that the states y^2, y^4 in $X \times T$ associated with 2 and 4, respectively, are such that $y^2 > y^4$. But $\{4, 2\} \in T_3$ and type 3 says that $4 \sim 2$. This contradicts the postulated agreement of agents of types 1 and 3 about types 2 and 4 that lie in $T_1 \cap T_3$.

Now we assign a utility function over the set X to each type in T. Choose the type (0, 0) and assign a utility function u_{00} to his ordering over the set X – that is, a function u_{00}: $X \rightarrow R$ such that $u_{00}(x) \geq u_{00}(y)$ $\Leftrightarrow (x; 0, 0) \gtrsim (y; 0, 0)$. To simplify the notation, let $(0, 0) = a$, $(1, 0) = b$, $(0, 1) = c$, and $(1, 1) = d$. Type a has experienced types b and c, so given u_a he can construct utility functions for type b and c, as follows: Define the function u_b: $X \rightarrow R$ by $u_b(x) = u_a(y)$, where y is such that $(x; b) \sim_a (y; a)$. This is well defined by the axiom of continuity. We define u_c in a similar way. These utility functions can then be used to define utility functions for types to the right and above, and so on. The utility functions so defined will provide us with a complete ordering over the set $X \times T$. We first have to show that the utility function that type d receives from type b coincides with the one that type c gives to type d.

Lemma 2 Let u_{dc} (u_{db}) be the utility function assigned to type d by type c (type b). Then $u_{dc}(x) = u_{db}(x) \quad \forall\, x \in X$.

Proof. Suppose not – that is, $\exists\, x \in X$ such that $u_{dc}(x) \neq u_{db}(x)$. Then by the axiom of continuity and the definition of u_{dc}, u_c, and u_a there exists x', $x'' \in X$ such that $u_{dc}(x) = u_c(x') = u_a(x'')$, and this is equivalent to

$$(x; d) \sim_c (x'; c) \sim_a (x''; a) \tag{i}$$

For the same reason there exists y, $y' \in X$ *such that* $u_{db}(x) = u_b(y) = u_a(y')$, which is equivalent to

$$(x; d) \sim_b (y; b) \sim_a (y'; a) \tag{ii}$$

Clearly $u_a(x'') \neq u_a(y')$. Assume that $u_a(x'') > u_a(y')$, then (i) and (ii) imply the intransitivity

$$(x''; a) \sim_a (x'; c) \sim_c (x; d) \sim_b (y; b) \sim_a (y'; a) < (x''; a)$$

which clearly contradicts Lemma 1.

Now we will show that if we repeat the same process – that is, where each type receiving a utility function endows utility functions to the types immediately to the right and above him, we can construct a complete ordering on $X \times T$. After that we will prove that the ordering is the unique one that respects all of the local orderings given by the local experts.

Let $c = \langle l^0, l^1, l^2, \ldots, l^k \rangle$ be a finite chain of elements of T. Let C be the set of all chains such that $\forall c \in C$, $c = \langle l^0, l^1, \ldots, l^k \rangle$, $l^0 = (0, 0)$, and $l^s \in T_{l^{s-1}}$, ($s = 1$ to k).

Definition: A finite chain $c = \langle l^0, l^1, l^2, \ldots, l^k \rangle$ has the northeast (NE) property if

$$(i_s, j_s) \in \{(i_{s-1} + 1, j_{s-1}), (i_{s-1}, j_{s-1} + 1)\} \qquad s = 1, \ldots, k$$

Let C^{NE} be the family of all chains with the NE property such that $l^0 =$ (0, 0). Clearly $\forall c \in C^{NE}$ and $l^s \in c$, $l^s \in T_{l^{s-1}}$. Therefore we can assign utility functions to all types in c in the way described here – that is, given the chain $c \in C^{NE}$, $c = <l^0, l^1, \ldots, l^k>$, we assign a utility function over the set $X \times l^0$ to the first type in the chain, denoted by u^{c_0}; then type l^0 endows a utility function u^{c_1} over the set $X \times l^1$ to type l^1, and so on.

For each $(m, n) \in T$ we can find a chain $c \in C^{NE}$, $c = <l^0, l^1, \ldots, l^l>$ such that $l^l = (m, n)$. Therefore we can assign utility functions to all elements of T. We have to show that the utility function assigned to any type in T does not depend on the chain $c \in C^{NE}$ we choose to connect that point with the origin.

Lemma 3 For all $c, c' \in C^{NE}$, $c = <l^0, l^1, \ldots, l^k>$, $c' = <l^0, l^1, \ldots, l^l$ such that $l^k = l^l$ we have $u^{c_k} = u^{c'}_l$.

Proof. Let $C_n \in C^{NE}$ be the set of all C^{NE} chains such that $\forall c \in C_n$, $\forall l^s \in c$, $l^s \in \{(i_s, 0), (i_s, 1), (i_s, 2), \ldots, (i_s, n)\}$.

(a) It is evident that the lemma is true for all members of C_0.

(b) Now we consider elements that can be reached by C_1 chains. If $c \in C_1$ has endpoint $l^k = (i_k, 0)$ we are in case (a). The lemma holds trivially for the type (0, 1) in T because there exists only one $c \in C^{NE}$ with end point (0, 1). Now take the type (1, 1). There are two chains $c, c' \in C_1$ with endpoint (1, 1). By Lemma 2 we know that $u^c_{(1,1)} = u^{c'}_{(1,1)}$. There are three chains c_1, c_2, $c_3 \in C_1$ with endpoints equal to (2, 1):

$c_1 = <(0,0), (0,1), (1,1), (2,1)>, c_2 = <(0,0), (1,0), (1,1), (2,1)>$

and

$c_3 = <(0, 0), (1, 0), (2, 0), (2, 11)>$.

As shown, $u^{c_1}_{(1,1)} = u^{c_2}_{(1,1)}$, and therefore $u^{c_1}_{(2,1)} = u^{c_2}_{(2,1)}$, so it remains to be proved that $u^{c_2}_{(2,1)} = u^{c_3}_{(2,1)}$: Clearly this is true because we can apply Lemma 2 by rewriting types in this way: $(1, 0) = a$, $(1, 1) = b$, $(2, 0) = c$, and $(2, 1) = d$. In the same way we can show that the four chains with endpoint (3, 1) will assign the same utility function to type (3, 1). We do the same for the rest of the types that can be reached by chains in C_1. This proves that the utility function assigned to any element that belongs to a C_1 chain is well-defined – that is, it does not depend on the chain we choose.

(c) For types that belong to C_2 chains, we can prove the lemma by the argument we used in (a) and (b) because the utility functions assigned to types with the second coordinate equal to 1 are unique; therefore we can view a chain $c \in C_2$ with endpoint $(i_k, 2)$ as a C_1 chain after taking type (0, 1) as the origin (instead of type (0, 0)). If we do the same for all chains C_n, $n \in N$, we prove the lemma.

Lemma 3 provides us with a complete ordering on the set $X \times T$ that respects the local orderings of all local experts. The next step is to prove that it is unique.

Denote by u^t the utility function that type t receives when we allow only C^{NE} chains. Let u^{c_t} be the utility function assigned to type $t \in T$ using the chain $c \in C$. There are two cases to consider:

(a) Let C^a be the set of chains of C such that

$\forall c \in C^a, c = <t^0, t^1, \ldots, t^k>, t^s \in \{(i_{s-1} + 1, j_{s-1}), (i_{s-1}, j_{s-1} + 1), (i_{s-1}-1, j_{s-1}), (j_{s-1}, j_{s-1}-1)\}, s = 1 \ldots k.$

Suppose that t is the last type in a chain $c \in C^a$. We want to show that $u^t = u^{c_t}$. If $c \in C^{NE}$, there is nothing to prove; otherwise, assume $u^t \neq u^{c_t}$.

Let $l^p = t'$ be the first term in c such that $u^{t'} \neq u^{c_{t'}}$; without loss of generality assume that $l^p = (i_p, j_p) = (i_{p-1}-1, j_{p-1})$. Now rename types in the following way: $t' = b$, $(i_{p-1}, j_{p-1}) = d$, $(i_{p-1}, j_{p-1}-1) = c$, $(i_{p-1}-1, j_{p-1}-1) = a$. The assumption says that $u^b \neq u^{c_b}$. Using the axiom of continuity in the same way we did for Lemma 2, we can show that there exists an intransitivity in the square a, c, d, b, which is impossible by Lemma 1. Therefore $u^{t'} = u^{c_{t'}}$. We have proved that for any type, the utility function assigned to it using chains in C^a coincides with the one assigned using chains in C^{NE}.

(b) There is a different way to assign utility functions to types in T. Given the utility function for type (i, j), we allow for the possibility that type $(i + 1, j)$ assigns a utility function to type $(i + 2, j)$ (and the same for j) – that is, we can have "jumps". Evidently this is not a problem because type (i, j) can assign a utility function to type $(i + 1, j)$, and given this utility function, type $(i + 1, j)$ assigns one to type $(i + 2, j)$, which must coincide with the one he assigned first, for otherwise we would have an intransitivity within $T_{(i+1, j)}$.

All chains of c are studied in cases (a) and (b). It therefore follows that the complete ordering is unique.

The proof for the n-dimensional case – that is, when $T = Nx \ldots xN$, parallels the proof earlier. First, we prove the equivalent of Lemma 1 for the n-cube with sides of area equal to one. Next, we assign a utility function over the set X to the type $(0, 0, \ldots, 0)$. In turn this type will assign utility functions to the n types having just one coordinate equal to one and the rest zeros, and we continue in the same way as before, with the natural modifications for the definitions of the different types of chains.

Theorem 2 is a corollary of Theorem 1. Its proof is available from the authors.

References

Arrow, Kenneth. 1977. "Extended Sympathy and the Possibility of Social Choice," *American Economic Review* 67, 219–225.

Blackorby, C., Donaldson, D., and Weymark, J. 1984. "Social Choice with Interpersonal Utility Comparisons: A Diagrammatic Introduction," *International Economic Review* 25, 327–356.

Cooter, Robert, and Peter Rappoport. 1986. "Were the Ordinalists Wrong About Welfare Economics?" *Journal of Economic Literature* 22, 507–530.

Dworkin, Ronald. 1981. "What is Equality? Part 1: Equality of Welfare," *Philosophy and Public Affairs* 10, 185–246.

Foley, Duncan. 1967. "Resource Allocation and the Public Sector," *Yale Economic Essays* 7, 45–98.

Gibbard, Allan, 1986. "Interpersonal comparisons: Preference, good, and the intrinsic reward of a life," in Elster, J., and A. Hylland, *Foundations of Social Choice Theory,* Cambridge: Cambridge University Press.

Kolm, Serge-Christophe. 1972. *Justice et Equité,* Paris: Editions du Centre National de la Recherche Scientifique.

May, Kenneth. 1954. "Intransitivity, Utility, and the Aggregation of Preference Patterns," *Econometrica* 22, 1–13.

Rawls, John. 1971 *A Theory of Justice.* Cambridge, Mass.: Harvard University Press.

Roemer, John. 1985. "A Note on Interpersonal Comparability and the Theory of Fairness," Dept. of Economics, University of California, Davis: Working Paper No. 261.

Sen, Amartya. 1979. "Interpersonal Comparisons of Welfare," in M. Boskin (ed.), *Economics and Human Welfare: Essays in Honor of Tibor Scitovsky,* New York: Academic Press.

Sen, Amartya. 1980. "Equality of What?" In *The Tanner Lectures on Human Values,* Volume I. Salt Lake City and New York: University of Utah Press and Cambridge University Press.

Thomson, William, and Hal Varian. 1986. "Theories of Justice Based on Symmetry," in Leonid Hurwicz et al. (eds.), *Social Goals and Social Organization: Essays in Memory of Elisha Pazner,* New York: Cambridge University Press.