Soc Choice Welfare (2001) 18: 211

Social Choice and Welfare

Editorial

The seven papers published in this issue were presented at conferences. The papers by Peter Fristrup and Hans Keiding, Charles Horvath, and Matjaž Omladič and Vesna Omladič were presented at the Toronto Fields Institute Conference "*New Directions in the Theory of Markets and Games*" organized by Myrna H. Wooders in honor of Robert J. Aumann in October 1995. The four other papers were presented at the Caen Conference "*Rationality in Utility Theory and Game Theory*" organized by John H. Nachbar, Maurice Salles and Norman Schofield in June 1997. All papers have been refereed according to the usual procedure followed by **Social Choice and Welfare**.

The Editors

Soc Choice Welfare (2001) 18: 213–226

Social Choice and Welfare

Strongly implementable social choice correspondences and the supernucleus

Peter Fristrup[1], **Hans Keiding**[2]

[1] Systems Analysis Department, RISØ National Laboratory, DK-4000 Roskilde, Denmark (e-mail: peter.fristrup@risoe.dk)
[2] University of Copenhagen, Institute of Economics, Studisetraede 6 DK-1455 Copenhagen K, Danemark (e-mail: Hans.keiding@econ.ku.dk)

Received: 2 February 1996/Accepted: 2 February 1998

Abstract. In this paper, we present a characterization of social choice correspondences which can be implemented in strong Nash equilibrium, stated in terms of the power structure implicit in the social choice rule. We extend the notion of an effectivity function to allow for simultaneous vetoing by several coalitions. This leads to the concept of a domination structure as a generalized effectivity function.

Using this concept and a solution known from the theory of effectivity functions, the supernucleus, we give a characterization of strongly implementable social choice correspondences as supernucleus correspondence relative to an appropriate domination structure.

1 Introduction

Implementation theory is concerned with the construction of mechanisms or game forms such that a particular behavior of individuals constrained by the mechanism will result in outcomes prescribed by a given social choice function or correspondence.

In this paper, we shall be interested in cooperative implementation, where it is assumed that individuals cooperate in their strategy choices. Following the tradition in this field, we consider strategy choices which are strong Nash

Paper presented at the conference "New Directions in the Theory of Markets and Games, Toronto, October 1995. The authors stand in debt to a referee who pointed out a flaw in our proof of the main result and suggested a correction which also considerably improved the formulation of the result.

equilibria: No coalition of agents can find a strategy array which will give a better outcome for all its members given the choices of the agents outside the coalition. Social choice correspondences which may be implemented in strong Nash equilibrium are called strongly implementable.

A useful tool in the study of strong implementability is the effectivity function, introduced by Moulin and Peleg [9], which describes the inherent power structure of a social choice correspondence: For each coalition it specifies subsets of alternatives such that the coalition by a suitable specification of the members' preferences can guarantee that output will be something from this subset.

The information provided by the effectivity function is exactly what is needed in order to determine whether a given social choice correspondence is partially implemented in strong Nash equilibrium, that is whether there is a strongly implementable social choice correspondence containing its values at each preference profile. However, it does not carry enough information to permit a full reconstruction of the social choice correspondence; several different social choice correspondences may have the same effectivity function. In this paper, we propose a refinement called a domination structure. The essential new feature of the domination structure as compared with the effectivity function is that it describes also simultaneous effectivity, cases where two or more non-disjoint coalitions may be simultaneously effective for particular subsets of alternatives.

This additional information provides a link between power structure and social choice: We show that strongly implementable social choice correspondences are characterized by their associated domination structures.

In the course of proving this result, we make use of a particular solution concept for cooperative games, namely the supernucleus, introduced in Fristrup and Keiding [3], cf. also Abdou and Keiding [1]. It turns out that every social choice correspondence which is strongly consistent may be represented as a supernucleus correspondence.

The paper is organized as follows: We start with a short description of the background and provide the necessary definitions in Sect. 2. In Sect. 3, we introduce domination structures and some of their properties, and in Sect. 4, we define the supernucleus with respect to domination structures. Section 5 contains the main result characterizing strongly implementable social choice correspondences.

2 Definitions

In this section, we introduce the notational conventions and the basic concepts to be used in the sequel. We are concerned with the standard situation of social choice where a finite set of individuals have to make a common choice from a finite set of alternatives. Individuals are endowed with preferences over the set of alternatives, and society's choice of alternative must be based on the individual preferences.

Let N be the set of individuals and A the set of alternatives, both assumed to be finite and nonempty. The set of all subsets of N is denoted by $P(N)$, and $2^N = P(N)\setminus\{\emptyset\}$ is the set of all the nonempty subsets of N; elements of 2^N are called coalitions. The set of all subsets of A is denoted by $P(A)$, and we let $P^2(A) = P(P(A))$. Thus, $P^2(A)$ consists of all families (including the empty family) of subsets of A.

By $\mathscr{L}$ we denote the set of all linear orders (complete, transitive, and antisymmetric binary relations) on A. Elements of $\mathscr{L}$ are written as R, Q, T, etc., and for $x, y \in A$, $x\,R\,y$ means that x is at least as good as y (and actually preferred to y if $x \neq y$) in the relation R. If B is a nonempty subset of A and $R \in \mathscr{L}$, then $(R|B)$ denotes the restriction of R to B, and $\max R$ is the unique element x^0 in A such that $x^0\,R\,y$ for all $y \in A$.

Let $S \in 2^N$ be a coalition. A map from S to $\mathscr{L}$ is called an S-profile and written $R^S = (R^i)_{i \in S}$, $R^i \in \mathscr{L}$. If $S = N$, then R^N is called a profile (without explicit mention of the coalition). We identify the profile $(R^S, R^{N\setminus S})$ with R^N for any coalition $S \in 2^N$. For nonempty subsets B and C of A and $S \in 2^N$ we write $B\,R^S\,C$ if $x\,R^i\,y$ for all $x \in B$, $y \in C$, $i \in S$. For simplicity, we write $x\,R^S\,C$ if $B = \{x\}$ and $B\,R^S\,y$ if $C = \{y\}$.

A social choice correspondence is a map $H : \mathscr{L}^N \to 2^A$. We shall be interested in social choice correspondences which are implementable in strong Nash equilibrium. By this we understand that there is some mechanism for decentralized choice so that when each individual chooses according to his own self-interest, the final outcome will be as prescribed by the social choice correspondence H. In order to state this formally, we need some more concepts:

An n-person game form is an array $G = (\Sigma^1, \ldots, \Sigma^n, \pi)$, where for each i, Σ^i is a nonempty set of strategies for player i, and where $\pi : \Sigma^N = \Sigma^1 \times \cdots \times \Sigma^n \to A$ is a surjective (outcome) function assigning to each array $\sigma^N = (\sigma^1, \ldots, \sigma^n)$ of strategies an element $\pi(\sigma^N)$ of the set A of alternatives. A game is defined as a pair (G, R^N), where G is a game form and $R^N \in \mathscr{L}^N$ is a profile. For a game $\Gamma = (G, R^N)$ we say that $\sigma^N \in \Sigma^N$ is a strong Nash equilibrium if there is no coalition $S \in 2^N$ and S-strategy $\tau^S = (\tau^i)_{i \in S}$ such that

$$\pi(\tau^S, \sigma^{N\setminus S})\, R^S\, \pi(\sigma^N), \quad \pi(\tau^S, \sigma^{N\setminus S}) \neq \pi(\sigma^N).$$

We say that the social choice correspondence H is implemented by the game form G in strong Nash equilibrium, or strongly implemented by G, if for all profiles R^N, the set of strong Nash equilibrium outcomes of the game (G, R^N) coincides with $H(R^N)$. A social choice correspondence H is strongly implementable if there exists a game form G so that H is implemented by G.

If the social choice correspondence $H : \mathscr{L}^N \to 2^A$ is strongly implementable, then it satisfies the following condition called strong positive association: For all $R^N, Q^N \in \mathscr{L}^N$ and $x \in A$, if $x \in H(R^N)$ and $\{i \mid x\,R^i\,y\} \subset \{i \mid x\,Q^i\,y\}$, all $y \in A\setminus\{x\}$, then $x \in H(Q^N)$ (see e.g. Peleg [11], lemma 6.5.1).

A useful tool in the study of strongly implementable social choice correspondences is its associated effectivity function. An effectivity function is a map $E : P(N) \to P^2(A)$ (where $E(S)$ is interpreted as the collection of subsets B of A for which S is effective, i.e. S can restrict society's choice to something from B) satisfying the following four conditions: (i) $\forall S \in 2^N$, $A \in E(S)$, (ii) $\forall S \in P(N)$, $\emptyset \notin E(S)$, (iii) $\forall B \in 2^A$, $B \in E(N)$, and (iv) $\forall B \in P(A)$, $B \notin E(\emptyset)$. An effectivity function $E : P(N) \to P^2(A)$ is maximal if for all $B \in 2^A$, $S \in 2^N$, if $B \notin E(S)$, then $A \backslash B \in E(N \backslash S)$ (if a coalition is not effective for B, then the complement of S must be effective for the set of alternatives not in B), and E is A-monotonic if for all $B, B' \in 2^A$ with $B \subset B'$ and all $S \in 2^N$, $B \in E(S)$ implies $B' \in E(S)$.

Effectivity functions can be constructed from either game forms or social choice correspondences. Let $G = (\Sigma^1, \dots, \Sigma^n, \pi)$ be a game form. Then the α-associated effectivity function E_α^G is given by

$$E_\alpha^G(S) = \{B \in 2^A \mid \exists \sigma^S \; \forall \tau^{N \backslash S} : \pi(\sigma^S, \tau^{N \backslash S}) \in B\}$$

for $S \in 2^N$, and $E_\alpha^G(\emptyset) = \emptyset$. Similarly, if $H : \mathscr{L}^N \to 2^A$ is a social choice correspondence which is non-imposed in the sense that for each $x \in A$, there is $R^N \in \mathscr{L}^N$ such that $H(R^N) = \{x\}$, then the α-effectivity function associated with H, E_α^H, is defined by

$$E_\alpha^H(S) = \{B \in P(A) \mid \exists Q^S \; \forall R^{N \backslash S} : H(Q^S, R^{N \backslash S}) \subset B\}$$

for $S \in 2^N$, $E_\alpha^H(\emptyset) = \emptyset$. Other constructions than the one given here are possible, but in the situations to be considered in this paper, where H is strongly implemented by G, most of them coincide. For further material on this, the reader is referred to e.g. Peleg [11], Moulin [8], Abdou and Keiding [1].

We shall need the notion of the core: Let $E : P(N) \to P^2(A)$ be an effectivity function, and let $R^N \in \mathscr{L}^N$ be a profile. We say that $B \in 2^A$ dominates $x \in A$ via $S \in 2^N$ in the profile R^N if $B \in E(S)$, $x \notin B$, and $B \, R^S \, x$. The core of E w.r.t. the profile R^N, written as $\mathscr{C}(E, R^N)$, is the set of alternatives $x \in A$ which are not dominated in the profile R^N. The effectivity function E is stable if $\mathscr{C}(E, R^N) \neq \emptyset$ for all profiles $R^N \in \mathscr{L}^N$.

The fundamental result of Moulin and Peleg [9] links all the concepts introduced above together: If H is strongly implementable by the game form G, then

$$E_\alpha^H = E_\alpha^G = E,$$

E is maximal and stable, and $H(R^N) \subset \mathscr{C}(E, R^N)$ for all profiles R^N. Conversely, if E is a maximal and stable effectivity function, then $\mathscr{C}(E, \cdot)$ is strongly implementable, and

$$E_\alpha^{\mathscr{C}(E, \cdot)} = E.$$

Thus, there is a canonical effectivity function associated with a strongly implementable social choice correspondence, and conversely, if we are given a maximal and stable effectivity function, then there is a particular strongly implementable social choice correspondence, namely the core correspondence, which has exactly this effectivity function associated with it.

As it was shown by Holzman [4], there may be several strongly implementable social choice correspondences associated with the same effectivity function. In order to be able to distinguish social choice correspondences with the same associated effectivity function, we need further information about the implicit power structure, and in the following section, we introduce a way of formalizing this information.

3 Domination structures

In this section, we introduce the notion of domination structures. As before, N is a set of individuals and A a set of alternatives. A domination structure $\mathcal{D}$ on (N, A) is a nonempty set of domination patterns, where a domination pattern is a finite (possibly empty) subset

$$D = \{(S_x, x) \mid x \in B\}$$

of $2^N \times A$, with B a proper subset of A. The domination structure is assumed to satisfy the following two consistency conditions:

(i) $\{(N, x) \mid x \in B\} \in \mathcal{D}$ for any proper subset B of A,
(ii) if $D \in \mathcal{D}$ and $D' \subset D$, then $D' \in \mathcal{D}$.

A domination pattern may be interpreted as an agreement between coalitions on coordinated vetoing: The coalition S_x can exclude x from being chosen by society, and it might find this action worthwhile if it knows that S_y excludes y for $y \in B \backslash \{x\}$, so that the exclusion of x will not entail the choice by society of any of those alternatives.

For $D = \{(S_x, x) \mid x \in B\}$ a domination pattern, the coalition $\mathrm{supp}(D) = \bigcup_{x \in B} S_x$ is called the support of D, and the set B is called the scope of D and denoted $\mathrm{sc}(D)$. If D is the empty domination pattern, then it has empty support and scope.

A domination structure $\mathcal{D}$ is superadditive if for each pair (D^1, D^2) of domination patterns from $\mathcal{D}$ with $\mathrm{supp}(D^1) \cap \mathrm{supp}(D^2) = \emptyset$ and $\mathrm{sc}(D^1) \cap \mathrm{sc}(D^2) = \emptyset$, we have that

$$D^1 \cup D^2 \in \mathcal{D}.$$

$\mathcal{D}$ is monotonic if for any $D \in \mathcal{D}$, $(S, y) \in D$, and $S' \in 2^N$ with $S \subset S'$, the domination pattern $D' = (D \cup \{(S', y)\}) \backslash \{(S, y)\}$ belongs to $\mathcal{D}$.

If D is a domination pattern and $S \in 2^N$ a coalition, then the restriction of D to S, denoted $(D|S)$, is the largest subset D' of D such that the $\mathrm{supp}(D') \subset S$. At a later stage, we shall need another operation on domination patterns: Let $\{S_1, \ldots, S_m\}$ be a partition of N, and for $k = 1, \ldots, m$, let $D^k = \{(S_y^k, y) \mid y \in B^k\} \in \mathcal{D}$ be a domination pattern with $\mathrm{supp}(D^k) \subset S_k$. Define

$$\bigoplus_{k=1}^{m} D^k = \left\{ (S_y, y) \mid y \in \bigcup_{k=1}^{m} B^k \right\}$$

where for each $y \in \bigcup_{k=1}^{m} B^k$,

$$S_y = \bigcup_{k: y \in \mathrm{sc}(D^k)} S_y^k,$$

i.e. S_y is the union of all sets S_y^k with $(S_y^k, y) \in D^k$ for some k. This family is not necessarily a domination pattern from $\mathscr{D}$; however, if $\mathscr{D}$ is superadditive and monotonic, it does belong to $\mathscr{D}$.

If $E : P(N) \to P^2(A)$ is an effectivity function which is A-monotonic, then E induces a family $\mathscr{D}_E$ of domination patterns

$$D_{(S,B)} = \{(S, x) \mid x \in A \backslash B\}$$

for $B \in E(S)$, $B \neq A$. The family $\mathscr{D}_E$ clearly satisfies condition (i) for domination structures, and (ii) follows from A-monotonicity. We say that the domination structure $\mathscr{D}$ refines E if $\mathscr{D}_E \subset \mathscr{D}$. We shall be interested in domination structures refining a given effectivity function which is associated with a strongly implementable social choice correspondence; such effectivity functions are A-monotonic.

Just as it was the case with effectivity functions, domination structures arise in a natural way from game forms or from social choice correspondences.

Lemma 1. *Let $G = (\Sigma^1, \ldots, \Sigma^n, \pi)$ be a game form, and let $\mathscr{D}^G$ be the set of families $\{(S_x, x) \mid x \in B\}$ for $B \subset A$, such that there is $\sigma^N \in \Sigma^N$ with the property*

$$\pi(\sigma^{S_x}, \tau^{N \backslash S_x}) \in A \backslash \{x\}$$

for any $\tau^{N \backslash S_x} \in \Sigma^{N \backslash S_x}$. Then $\mathscr{D}^G$ is a domination structure refining E_α^G, and $\mathscr{D}^G$ is monotonic and superadditive. Furthermore, if $D^k = \{(S_z^k, z) \mid z \in B^k\}$, $k = 1, \ldots, m$, belong to $\mathscr{D}^G$ and have mutually disjoint supports, then there is $\sigma^N \in \Sigma^N$ such that $\pi(\sigma^{S_z^k}, \tau^{N \backslash S_z^k}) \in A \backslash \{z\}$ for all $\tau^{N \backslash S_z^k} \in \Sigma^{N \backslash S_z^k}$, each $z \in B^k$ and $k \in \{1, \ldots, m\}$.

Proof. $\mathscr{D}^G$ is a domination structure: First of all we note that if $\{(S_x, x) \mid x \in B\}$ belongs to $\mathscr{D}^G$, then $B \neq A$, since otherwise there would be $\sigma^N \in \Sigma^N$ with $\pi(\sigma^N) \neq x$ for all $x \in A$, a contradiction.

Further, $\mathscr{D}^G$ satisfies (i) and (ii) for domination structures: (i) Choose $w \in A \backslash B$; by the surjectivity of π, there is $\sigma^N \in \Sigma^N$ with $\pi(\sigma^N) = w$, and using the definition of $\mathscr{D}^G$, we have trivially $\{(N, x) \mid x \in B\} \in \mathscr{D}^G$. (ii) Let $D \in \mathscr{D}^G$ and let σ^N be a strategy array associated with D according to the definition. Then any nonempty subset D' of D belongs to $\mathscr{D}^G$ with the same associated strategy array σ^N.

$\mathscr{D}^G$ refines E_α^G: For each (S, B) with $B \in E_\alpha^G(S)$, $B \neq A$, if $\sigma^N \in \Sigma^N$ is such that $\pi(\sigma^S, \tau^{N \backslash S}) \in B$ for all $\tau^{N \backslash S} \in \Sigma^{N \backslash S}$, then $\{(S, y) \mid y \in A \backslash B\}$ belongs to $\mathscr{D}^G$.

Monotonicity of $\mathscr{D}^G$ is an easy consequence of its definition; for $S, S' \in 2^N$ with $S \subset S'$ and any strategy array σ^N such that $\pi(\sigma^S, \cdot)$ does not contain y in its image, we have a fortiori that $\pi(\sigma^{S'}, \cdot)$ never attains y.

Superadditivity will follow immediately from the final statement of the lemma, to which we turn: Suppose that $D^k = \{(S_z^k, z) \mid z \in B^k\}$, $k = 1, \ldots, m$, belong to $\mathscr{D}^G$ and have disjoint supports; then for each k there is σ_k^N in Σ^N such that $\pi(\sigma^{S_z^k}, \tau^{N \setminus S_z^k}) \in A \setminus \{z\}$, all $\tau^{N \setminus S_z^k}$, for each $z \in B^k$. Choose a strategy array σ^N agreeing with σ_k^N on $\mathrm{supp}(D^k)$, $k = 1, \ldots, m$; then σ^N has the desired property. □

The domination structure $\mathscr{D}^G$ is defined by the property that there is a strategy array σ^N such that for each x, the coalition S_x can avoid x no matter what the complementary coalition chooses. Thus, it resembles the traditional α-construction used in the definition of characteristic functions or effectivity functions.

Our next step is to associate a domination structure with a social choice correspondence. Let $H : \mathscr{L}^N \to 2^A$ be a social choice correspondence, and let $\mathscr{D}^H$ be the family of all subsets of domination patterns $D = \{(S_y, y) \mid y \neq x\}$ for which there is a profile $R^N \in \mathscr{L}^N$ with $x \in H(R^N)$ and $S_y = \{i \mid x R^i y\}$. The family $\mathscr{D}^H$ may not be a domination structure (since (i) is not necessarily fulfilled) for an arbitrary social choice correspondence H. However, in the case where H is strongly implementable, we show below that $\mathscr{D}^H$ is indeed a domination structure.

A domination structure $\mathscr{D}$ is called regular if for each $D \in \mathscr{D}$, there is $D' \in \mathscr{D}$ with $D \subset D'$ such that the scope of D contains exactly $|A| - 1$ elements.

Theorem 1. *Let $H : \mathscr{L}^N \to 2^A$ be a social choice correspondence which is strongly implemented by the game form $G = (\Sigma^1, \ldots, \Sigma^n, \pi)$. Then*

$$\mathscr{D}^H = \mathscr{D}^G = \mathscr{D},$$

and $\mathscr{D}$ is a regular domination structure satisfying monotonicity and superadditivity.

Proof. Let $D' \in \mathscr{D}^H$; by definition, D' may be extended to a domination pattern $D = \{(S_y, y) \mid y \neq x\}$ such that for some profile $R^N \in \mathscr{L}^N$, $x \in H(R^N)$ and $S_y = \{i \mid x R^i y\}$, each $y \neq x$. Since H is strongly implemented by G, there is a strategy array $\sigma^N \in \Sigma^N$ with $\pi(\sigma^N) = x$ which is a strong Nash equilibrium in the game (G, R^N). In particular, for each $y \neq x$ the coalition $N \setminus S_y$ consisting of all individuals i with $y R^i x$ cannot achieve y. But this means that σ^N has the property that for each y, $\pi(\sigma^{S_y}, \cdot)$ does not contain y in its image, thus $D \in \mathscr{D}^G$. By property (ii) of domination structures, we conclude that $D' \in \mathscr{D}^G$.

Next, we show that $\mathscr{D}^G$ is regular. Let $D = \{(S_x, x) \mid x \in B\} \in \mathscr{D}^G$ be arbitrary, and let $\sigma^N \in \Sigma^N$ be such that

$$\pi(\sigma^{S_x}, \tau^{N \setminus S_x}) \in A \setminus \{x\}$$

for all $\tau^{N \setminus S_x} \in \Sigma^{N \setminus S_x}$, $x \in B$; further, let $w = \pi(\sigma^N)$. Then the domination pattern $D \cup \{(N, x) \mid x \in A \setminus [B \cup \{w\}]\}$ trivially belongs to $\mathscr{D}^G$, which therefore is regular.

Now, let $D' \in \mathscr{D}^G$; by regularity of $\mathscr{D}^G$, we may extend D' to a domination pattern of the form $D = \{(S_y, y) \mid y \neq x\}$ for some $x \in A$. Using the definition of $\mathscr{D}^G$, we know that there is a strategy array σ^N such that $\pi(\sigma^N) = x$ and $\pi(\sigma^{S_y}, \cdot)$ does not contain y in its image, each $y \neq x$.

For each $i \in N$, let $A_i = \{y \mid i \in S_y, y \neq x\}$, and let $R^N \in \mathscr{L}^N$ be a profile such that $A \backslash A_i \, R^i \, x \, R^i \, A_i$ for each $i \in N$. By our construction, σ^N is a strong Nash equilibrium of (G, R^N), and since G strongly implements H, we have $x \in H(R^N)$. But this means exactly that D, and consequently D', belongs to $\mathscr{D}^H$.

Since $\mathscr{D}^H = \mathscr{D}^G$, monotonicity and superadditivity follow from Lemma 1. □

4 The supernucleus of a domination structure

In analogy with the classical approach to partial implementability, where the implementable correspondences were characterized as core correspondences with respect to the associated effectivity function, we now introduce a solution concept for games defined by domination structures. However, we shall use not the core but the supernucleus, a concept introduced for effectivity functions by Fristrup and Keiding [3], where an alternative x belongs to the supernucleus at a given profile R^N if there exists another ("associated") profile Q^N such that y is dominated at Q^N via $\{i \mid x \, R^i \, y\}$, each $y \neq x$.

In the extension of the supernucleus to domination structures, we have no need for the associated profile but work directly with domination patterns: Let $R^N \in \mathscr{L}^N$; the supernucleus of $\mathscr{D}$ at R^N, denoted $\mathrm{sNuc}(\mathscr{D}, R^N)$, is the set of alternatives x for which there is a domination pattern $D = \{(S_y, y) \mid y \neq x\}$, said to be associated with x at R^N, such that $x \, R^{S_y} \, y$. The domination structure $\mathscr{D}$ is said to be supernucleus-stable if $\mathrm{sNuc}(\mathscr{D}, R^N) \neq \varnothing$ for each profile $R^N \in \mathscr{L}^N$.

Intuitively, the supernucleus at R^N is obtained by vetoing under a suitable misrepresentation of preferences, so that each alternative is vetoed by a coalition sincerely preferring the chosen alternative to the vetoed alternative. In the context of domination structures, the associated domination pattern represents a joint vetoing strategy rather than a profile, so that the interpretation as misrepresentation is less direct.

The next lemma gives us a convenient property of a suitably chosen associated profile:

Lemma 2. *Let $\mathscr{D}$ be a domination structure satisfying monotonicity, and let $x \in \mathrm{sNuc}(\mathscr{D}, R^N)$. Then there is a domination pattern $D = \{(S_y, y) \mid y \neq x\}$ associated with x at R^N such that*

$$S_y = \{i \mid x \, R^i \, y\}$$

for all $y \in A$ and $i \in N$. In particular, if $\mathscr{D}$ is regular and supernucleus-stable, then

$$\mathscr{D}^{\mathrm{sNuc}(\mathscr{D}, \cdot)} = \mathscr{D}.$$

Proof. For $x \in \mathrm{sNuc}(\mathscr{D}, R^N)$, choose a domination pattern $D' = \{(S'_y, y) \mid y \neq x\}$ associated with x at R^N. Choose $y \neq x$ arbitrarily; we have $S'_y \subset \{i \mid x\, R^i\, y\}$ by the definition of the supernucleus. By monotonicity, for each $y \neq x$ the domination pattern

$$[D' \cup \{(\{i \mid x\, R^i\, y\}, y)\}] \backslash \{(S'_y, y)\}$$

belongs to $\mathscr{D}$. Repeating the argument for each $y \neq x$ we get that the domination pattern $\{(\{i \mid x\, R^i\, y\}, y) \mid y \neq x\}$ is associated with x at R^N.

From the first part of the lemma, we get that $\mathscr{D}^{\mathrm{sNuc}(\mathscr{D}, \cdot)} \subset \mathscr{D}$. For the converse inclusion, let $D = \{(S_y, y) \mid y \neq x\} \in \mathscr{D}$ and choose a profile $R^N \in \mathscr{L}^N$ such that $S_y = \{i \mid x\, R^i\, y\}$, each $y \neq x$. Then $x \in \mathrm{sNuc}(\mathscr{D}, R^N)$ trivially, and $D \in \mathscr{D}^{\mathrm{sNuc}(\mathscr{D}, \cdot)}$. □

Lemma 3. *Let $H : \mathscr{L}^N \to 2^A$ be a social choice correspondence such that $\mathscr{D}^H$ is a domination structure. Then $H(R^N) \subset \mathrm{sNuc}(\mathscr{D}^H, R^N)$ for each profile $R^N \in \mathscr{L}^N$; in particular, $\mathscr{D}^H$ is supernucleus-stable. If H satisfies strong positive association, then $H(R^N) = \mathrm{sNuc}(\mathscr{D}^H, R^N)$ for all $R^N \in \mathscr{L}^N$.*

Proof. Let $R^N \in \mathscr{L}^N$ be arbitrary, $x \in H(R^N)$, and let $D = \{(S_y, y) \mid y \neq x\}$, where $S_y = \{i \mid x\, R^i\, y\}$. Then $D \in \mathscr{D}^H$, and x belongs to the supernucleus of $\mathscr{D}^H$ at R^N with associated domination pattern D.

Suppose that H satisfies strong positive association, and let $x \in \mathrm{sNuc}(\mathscr{D}^H, R^N)$. Then there is a domination pattern $D = \{(S_y, y) \mid y \neq x\} \in \mathscr{D}^H$ such that $S_y \subset \{i \mid x\, R^i\, y\}$, each $y \neq x$. From the definition of $\mathscr{D}^H$ we get that there is $Q^N \in \mathscr{L}^N$ with $S_y = \{i \mid x\, Q^i\, y\}$, all $y \neq x$, and $x \in H(Q^N)$. Using strong positive association we conclude that $x \in H(R^N)$. □

5 A characterization of strongly implementable social choice correspondences

We have seen in the previous section that a strongly implementable social choice correspondence is a supernucleus correspondence. In order to apply this in the characterization of strongly implementable correspondence, we need a converse saying that the supernucleus correspondence of a domination structure satisfying suitable conditions is indeed strongly implementable.

If $\mathscr{D}$ is a domination structure and $D = \{(S_y, y) \mid y \in B\}$ a domination pattern (not necessarily belonging to $\mathscr{D}$), then we denote by $\mathrm{sNuc}(\mathscr{D}, D)$ the set of supernucleus elements obtained from profiles compatible with D, that is

$$\mathrm{sNuc}(\mathscr{D}, D) = \{x \in A \backslash B \mid \exists Q^N \in \mathscr{L}^N : \{i \mid x\, Q^i\, y\} = S_y, y \in B, x \in \mathrm{sNuc}(\mathscr{D}, Q^N)\}.$$

If $\mathscr{D}$ is regular, then $\mathrm{sNuc}(\mathscr{D}, D)$ is nonempty for each $D \in \mathscr{D}$: D may be extended to a domination pattern of the form $\{(S_y, y) \mid y \neq x\}$, and $x \in \mathrm{sNuc}(\mathscr{D}, R^N)$ for each R^N with $S_y = \{i \mid x\, R^i\, y\}$, $y \neq x$.

If an alternative z does not belong to $\mathrm{sNuc}(\mathscr{D}, D)$ for some $D \in \mathscr{D}$, then z will not be chosen by the supernucleus correspondence at the profile Q^N even if $z\,Q^{S_y}\,y$ for each $(S_y, y) \in D$. Thus, the domination pattern D establishes an indirect domination of the alternative z.

We shall need a property of domination structures assuring that indirect domination translates to the existence of a suitable domination pattern in $\mathscr{D}$: A domination structure $\mathscr{D}$ is said to satisfy property (P) if the following holds: Let $\{S_1, \dots, S_m\}$ be a partition of N, $D^1, \dots, D^m \in \mathscr{D}$ domination patterns, B a subset of A, and for each $y \in B$, let S^y be a subset of N. If for all $y \in B$ we have that

$$y \notin \mathrm{sNuc}\left(\mathscr{D}, \bigoplus_{k=1}^{m} (D^k | S_k \backslash S^y)\right)$$

then

$$\{(N \backslash S^y, y) \mid y \in B\} \in \mathscr{D}.$$

To see what property (P) is about, it might be helpful to consider a simple case: Let the partition be $\{N\}$ (the trivial one), and let D be a domination pattern from $\mathscr{D}$. Suppose that $z \notin \mathrm{sNuc}(\mathscr{D}, D)$ for some z not in the scope of D. Let $B = \mathrm{sc}(D) \cup \{z\}$, and choose sets $S^y = N \backslash S_y$ for $y \in \mathrm{sc}(D)$, $S^z = N \backslash \mathrm{supp}(D)$. Then $y \notin \mathrm{sNuc}(\mathscr{D}, (D|N\backslash S^y))$ since $(S_y, y) \in (D|N\backslash S^y)$, $y \in B$, and also $z \notin \mathrm{sNuc}(\mathscr{D}, (D|N\backslash S^z))$. Property (P) now tells us that $(\mathrm{supp}(D), z)$ may be added to the domination pattern D to give a new domination pattern belonging to $\mathscr{D}$; the alternative z which was indirectly dominated by D is actually dominated by an extension of D. Property (P) in its general formulation covers also the case of several alternatives being indirectly dominated and indeed gives exactly what we shall need later on.

Domination structures which are associated with strongly implementable social choice correspondences have the property (P):

Lemma 4. *Let $H : \mathscr{L}^N \to 2^A$ be strongly implementable. Then $\mathscr{D}^H$ has property (P).*

Proof. Let $\{S_1, \dots, S_m\}$ be a partition of N, let $D^1, \dots, D^m \in \mathscr{D}^H$, and let $B \subset A$ and $S^y \subset N$, $y \in B$, be such that

$$y \notin \mathrm{sNuc}\left(\mathscr{D}^H, \bigoplus_{k=1}^{m} (D^k | S_k \backslash S^y)\right)$$

for each $y \in B$.

Let $G = (\Sigma^1, \dots, \Sigma^n, \pi)$ be a game form implementing H. Since $\mathscr{D}^H$ is monotonic and superadditive, the domination pattern $\bigoplus_{k=1}^m (D^k|S_k)$ as well as each of the domination patterns $\bigoplus_{k=1}^m (D^k|S_k \backslash S^y)$ for $y \in B$ belong to $\mathscr{D}^H$. By Theorem 1, $\mathscr{D}^H = \mathscr{D}^G$, and using the last statement in Lemma 1 we get the existence of a strategy array $\tau^N \in \Sigma^N$ such that $\pi(\tau^{S_z^k}, \cdot)$ does not attain z, for each z with $(S_z^k, z) \in (D^k|S_k)$, $k = 1, \dots, m$.

Choose $y \in B$ arbitrarily. Suppose that there were $\sigma^N \in \Sigma^N$ with $\pi(\sigma^N) = y$ and such that $\pi(\sigma^{S_z}, \cdot)$ takes values in $A \backslash \{z\}$, each $(S_z, z) \in \bigoplus_{k=1}^m (D^k|S_k \backslash S^y)$. Let the profile $R^N \in \mathscr{L}^N$ be such that $\{i \mid y\,R^i\,z\} = S_z$ for all such z and

$y\,R^N\,x$ for all other alternatives x in A. Then σ^N is a strong equilibrium of (G, R^N); indeed, for $z \in B$ with $(S_z, z) \in \bigoplus_{k=1}^m (D^k|S_k \backslash S^y)$, the coalition $\{i \,|\, z\,R^i\,y\} = N \backslash S_z$ cannot obtain z, and $\{i \,|\, x\,R^i\,y\} = \emptyset$ for all remaining $x \in A$, $x \neq y$. By strong implementability and Lemma 3, we get that $y \in H(R^N) = \mathrm{sNuc}(\mathscr{D}^H, R^N)$. However, this contradicts the fact that

$$y \notin \mathrm{sNuc}\left(\mathscr{D}^H, \bigoplus_{k=1}^{m} (D^k|S_k \backslash S^y)\right).$$

Thus, every σ^N such that $\pi(\sigma^{S_z}, \cdot)$ takes values in $A \backslash \{z\}$, all z with $(S_z, z) \in \bigoplus_{k=1}^m (D^k|S_k \backslash S^y)$, has the property that $\pi(\sigma^{N \backslash S^y}, \cdot)$ never attains y. Since τ^N is such a strategy array, and $y \in B$ was chosen arbitrarily, we have that $\{(N \backslash S^y, y) \,|\, y \in B\}$ belongs to $\mathscr{D}^G = \mathscr{D}^H$, so that $\mathscr{D}^H$ satisfies property (P). □

Several of the properties considered previously follow from property (P):

Lemma 5. *Let $\mathscr{D}$ be a domination structure which satisfies property (P). Then* $\mathrm{sNuc}(\mathscr{D}, D) \neq \emptyset$ *for each $D \in \mathscr{D}$, and $\mathscr{D}$ is monotonic, regular, and superadditive.*

Proof. Assume that $\mathrm{sNuc}(\mathscr{D}, D) = \emptyset$ for some $D = \{(S_x, x) \,|\, x \in B\} \in \mathscr{D}$; consider the (trivial) partition $\{N\}$ and the family $\{S^x \,|\, x \in A\}$, where $S^x = N \backslash S_x$ if $x \in B$ and $S^x = \emptyset$ otherwise. Then $(D|N \backslash S^x)$ contains (S_x, x) if $x \in B$, so that $x \notin \mathrm{sNuc}(\mathscr{D}, (D|N \backslash S^x))$, and for $x \in A \backslash B$ we have $(D|N \backslash S^x) = D$ and $x \notin \mathrm{sNuc}(\mathscr{D}, (D|N \backslash S^x))$ by our assumption. By property (P),

$$\{(N \backslash S^x, x) \,|\, x \in A\} \in \mathscr{D},$$

and we have exhibited a domination pattern from $\mathscr{D}$ with scope A, a contradiction. We conclude that $\mathrm{sNuc}(\mathscr{D}, D) \neq \emptyset$ for each $D \in \mathscr{D}$.

To show monotonicity, let $D = \{(S_x, x) \,|\, x \in B\} \in \mathscr{D}$, and choose $(S_y, y) \in D$ and $S' \in 2^N$ with $S_y \subset S'$. Let $\{N\}$ be the trivial partition of N and define the sets $S^x = N \backslash S_x$ for $x \in B \backslash \{y\}$ and $S^y = N \backslash S'$. Then $x \notin \mathrm{sNuc}(\mathscr{D}, (D|N \backslash S^x))$ follows from $(S_x, x) \in (D|N \backslash S^x)$, all $x \in B$. By property (P), the domination pattern

$$\{(N \backslash S^x, x) \,|\, x \in B\} = (D \cup \{(S', y)\}) \backslash \{(S_y, y)\}$$

belongs to $\mathscr{D}$.

For regularity, let $D = \{(S_y, y) \,|\, y \in B\} \in \mathscr{D}$. Choose $x \in \mathrm{sNuc}(\mathscr{D}, D)$ (which is nonempty according to the first part of the lemma); then there is a profile $R^N \in \mathscr{L}^N$ with $x \in \mathrm{sNuc}(\mathscr{D}, R^N)$ and $\{i \,|\, x\,R^i\,y\} = S_y$ for each $y \in B$. Using the definition of the supernucleus we get that there is a domination pattern $D' = \{(S'_y, y) \,|\, y \neq x\} \in \mathscr{D}$ such that $S'_y \subset S_y$ for each $y \in B$. By monotonicity,

$$\{(S_y, y) \,|\, y \in B\} \cup \{(S'_y, y) \,|\, y \notin B, y \neq x\}$$

is a domination pattern from $\mathscr{D}$; it extends D and its scope is $A \backslash \{x\}$.

To show that $\mathscr{D}$ satisfies superadditivity, let $D^1, D^2 \in \mathscr{D}$ be domination patterns with disjoint supports and scopes. Denoting the latter by B^1, B^2, we choose a partition $\{S^1, S^2\}$ of N such that $\mathrm{supp}(D^i) \subset S^i$, $i = 1, 2$, and for each $z \in B = B^1 \cup B^2$, we let $S^z = N \setminus S_z$. Clearly,

$$z \notin \mathrm{sNuc}(\mathscr{D}, (D^1|S^1\setminus S^z) \oplus (D^2|S^2\setminus S^z))$$

for $z \in B^i$, $i = 1, 2$, and by property (P), $\{(N\setminus S^z, z) \,|\, z \in B\} = D^1 \cup D^2$ belongs to $\mathscr{D}$. □

By now we have shown that if the social choice correspondence H is strongly implementable, then $\mathscr{D}^H$ is supernucleus-stable (Lemma 3), monotonic (Lemma 1), regular (Theorem 1), and satisfies property (P) (Lemma 4). As it can be seen from Lemma 5, property (P) has a key role, and this impression is confirmed by the following result, which together with the previous results gives a full characterization of strongly implementable social choice correspondences.

Theorem 2. *Let $\mathscr{D}$ be a domination structure on (N, A) which is supernucleus-stable and satisfies property (P). Then the social choice correspondence* $\mathrm{sNuc}(\mathscr{D}, \cdot) : \mathscr{L}^N \to 2^A$ *is strongly implementable, and* $\mathscr{D}^{\mathrm{sNuc}(\mathscr{D},\cdot)} = \mathscr{D}$.

Proof. Define the game form $G = (\Sigma^1, \ldots, \Sigma^n, \pi)$ as follows: For each i, Σ^i consists of all triples (D^i, Q^i, t), where D^i is a domination pattern from $\mathscr{D}$, $Q^i \in \mathscr{L}$ is a preference relation, and $t \in \mathbf{N}$ is a natural number. The outcome function π sends strategy n-tuples

$$\sigma^N = ((D^1, Q^1, t^1), \ldots, (D^n, Q^n, t^n))$$

to alternatives; the definition of π is as follows:

Let $\{(S_1, \ldots, S_m)\}$ be the coarsest partition of N into disjoint subsets such that for each k

$$i, j \in S_k \Rightarrow \sigma_1^i = \sigma_1^j = D^{S_k},$$

(where σ_1^i denotes the first component of the strategy σ^i, $i \in N$), so that members of S_k agree on the first component of their strategies; $\{S_1, \ldots, S_m\}$ is called the partition induced by σ^N, and

$$\tilde{D} = \bigoplus_{k=1}^{m} (D^{S_k}|S_k)$$

is called the domination pattern induced by the strategy array σ^N; by Lemma 5, $\mathscr{D}$ is superadditive and monotonic, so that $\tilde{D} \in \mathscr{D}$, and $\mathrm{sNuc}(\mathscr{D}, \tilde{D}) \neq \emptyset$. Now define the outcome as

$$\pi(\sigma^N) = \max(Q^{i_0}|\mathrm{sNuc}(\mathscr{D}, \tilde{D})),$$

where i^0 is the smallest index i of an individual with $t^i = \max_{j \in N} t^j$.

We show that the supernucleus at any profile may occur as strong Nash equilibrium outcome in the game defined by G: Let $R^N \in \mathscr{L}^N$ be arbitrary; since $\mathscr{D}$ is supernucleus-stable, there is $x \in \mathrm{sNuc}(\mathscr{D}, R^N)$ with an associated domination pattern $D = \{(S_y, y) \mid y \neq x\} \in \mathscr{D}$. Choosing strategies $\sigma^i = (D, R^i, 1)$, $i = 1, \ldots, n$, we get that the partition of N induced by this strategy array is $\{N\}$, the induced domination pattern $\tilde{D}$ is D itself, and $\text{sNuc}(\mathscr{D}, D) = \{x\}$. It follows that $\pi(\sigma^N) = x$.

Suppose that $y \neq x$ and that S is a coalition with $y\, R^S\, x$. If S were to obtain y, it must choose an S-strategy τ^S with first component of τ^j different from D for some $j \in S$, giving rise to a new induced domination pattern $\hat{D} \neq D$. However, the partition of N induced by $(\tau^S, \sigma^{N\backslash S})$ contains a superset S' of $N\backslash S$, and since $S_y \subset N\backslash S$, we have that $y \notin \text{sNuc}(\mathscr{D}, \hat{D})$, i.e. $y \neq \pi(\tau^S, \sigma^{N\backslash S})$. Thus, σ^N is a strong Nash equilibrium of (G, R^N).

Next, we show that all strong Nash equilibrium outcomes are supernucleus elements: Let $R^N \in \mathscr{L}^N$ be an arbitrary profile and let σ^N be a strong Nash equilibrium of the game (G, R^N) with outcome $x = \pi(\sigma^N)$. Let $\{S_1, \ldots, S_m\}$ be the partition of N induced by σ^N, and let $D^{S_1}, \ldots, D^{S_m}$ be the corresponding domination patterns. For each $y \in A\backslash\{x\}$, let $S^y = \{i \mid y\, R^i\, x\}$. If for some $y \neq x$ we have

$$y \in \text{sNuc}\left(\mathscr{D}, \bigoplus_{k=1}^{m} (D^{S_k} | S_k \backslash S^y)\right),$$

then S^y would be able to deviate from σ^N so as to obtain y, in contradiction to σ^N being a strong Nash equilibrium. It now follows by property (P) that $\{(N\backslash S^y, y) \mid y \neq x\} \in \mathscr{D}$, and therefore $x \in \text{sNuc}(\mathscr{D}, R^N)$, as required. $\square$

In order to use the characterization result for checking whether a given correspondence H is strongly implementable, one has first of all to check whether it satisfies strong positive association. Next one has to exhibit its domination structure $\mathscr{D}^H$ (checking that it is actually a domination structure), and then see whether it is supernucleus-stable and satisfies property (P). The first task is in principle a simple one, since $\mathscr{D}^H$ can be obtained in a rather mechanical way from H; supernucleus-stability follows from Lemma 3 once it is known that $\mathscr{D}^H$ is a domination structure. However, showing that property (P) is fulfilled may not be quite straightforward in applications.

Comparing the present characterization of strongly consistent social choice correspondences using the supernucleus with respect to a domination structure with the characterization by Dutta and Sen [2], or with the recent version of it given by Suh [12], we employ a somewhat simpler family of implementing game forms, but otherwise the two approaches are related, as it might well be expected. In the present approach, the check for strong consistency consists in building the associated domination structure and checking it for particular properties; compared with the Dutta-Sen conditions, which are concerned with existence of certain auxiliary sets, our proposed procedure looks simpler. However, it amounts essentially to giving a more explicit description of the abstract auxiliary correspondences which play an important role in their characterization. Consequently, the present approach may be considered as complementing the work of Dutta and Sen rather than as an alternative to their result.

References

[1] Abdou J, Keiding H (1991) Effectivity functions in social choice. Kluwer Academic Publishers, Dordrecht
[2] Dutta B, Sen A (1991) Implementation under strong equilibrium: A complete characterization. J Math Econ 20: 49–67
[3] Fristrup P, Keiding H (1988) Strongly implementable social choice functions. Discussion Paper 88-14, Institute of Economics, University of Copenhagen
[4] Holzman R (1987) Sub-Core solutions of the problem of strong implementation. Int J Game Theory 16: 263–289
[5] Keiding H (1985) Necessary and sufficient conditions for stability of effectivity functions. Int J Game Theory 14: 93–101
[6] Maskin ES (1977) Nash equilibrium and welfare optimality. Mimeo MIT
[7] Maskin ES (1979) Implementation and strong nash equilibrium. In: Laffont J-J (ed.) Aggregation and revelation of preferences. North-Holland, Amsterdam pp. 433–439
[8] Moulin H (1983) The strategy of social choice. North Holland, Amsterdam
[9] Moulin H, Peleg B (1982) Cores of effectivity functions and implementation theory. J Math Econ 10: 115–145
[10] Peleg B (1982) Convex effectivity functions. Discussion Papers, The Hebrew University, Jerusalem
[11] Peleg B (1984) Game theoretic analysis of voting in committees. Cambridge University Press, Cambridge
[12] Suh S-C (1996) An algorithm for checking strong Nash implementability. J Math Econ 25: 109–122

Soc Choice Welfare (2001) 18: 227–250

Social Choice and Welfare

On the topological social choice problem

Charles D. Horvath

Département de Mathématiques, Université de Perpignan, 66860 Perpignan cedex, France

Received: 31 January 1996/Accepted: 12 March 1997

1 Introduction

In this paper we consider the social choice problem as presented by Chichilnisky and Heal [13]. We refer to their work for the motivation and the interpretation and to [8], [23] for further discussion.

The fundamental work of Chichilnisky and Heal has been developed further by Baryshnikov in [3].

We are given a topological space P, the preference space, and the number of individuals n. A *social choice rule* is a continuous function $\Phi : P^n \to P$ for which the following two properties are verified:

(a) For any permutation $(m_1, \ldots, m_n)$ of $(1, \ldots, n)$ one has $\Phi(p_1, \ldots, p_n) = \Phi(p_{m_1}, \ldots, p_{m_n})$.
(b) For any $p \in P$ one has $\Phi(p, \ldots, p) = p$.

Condition (a) is called *anonymity* and condition (b) *unanimity*.

An element of P^n is called a society's profile.

Here we assume that P is given with its topology. We are not concerned with what the elements of P are, normalized gradients, complete continuous monotonic preorders or other classes of preorders, and we do not ask how natural the topology of P is.

The problem is to find conditions on the topological space P from which one can prove that social choice rules exist.

Solutions have been given by Le Breton and Uriarte [23], and Allen [1].

The main theorem of [13], Theorem 1 on page 82, is stated as follows:

Theorem 1 (Chichilnisky-Heal). *A necessary and sufficient condition for the existence of a social choice rule for each $n \geq 2$ is that the space of preferences P be contractible.*

Let us recall in which context it is stated.

Y is the unit sphere in $\mathbb{R}^n$, $V(Y)$ is the space of all C^1 vector fields on Y and P is the space of all locally integrable elements of $V(Y)$ which have norm one.

To prove their theorem, Chichilnisky and Heal assume that P is a parafinite CW-complex whose convex hull in $V(Y)$, which they note $k(P)$, is also a parafinite CW-complex. Furthermore, in the proof itself they use the fact that the pair $(k(P), P)$ is a relative CW-complex. That is not so obvious. We restate the theorem in the following form, which is what their proof shows:

Theorem 2 (Chichilnisky-Heal). *Assume that the preference space P is a connected parafinite CW-complex. Then a necessary and sufficient condition for the existence of a social choice rule is that the preference space be contractible.*

The original proof by Chichilnisky and Heal uses tools from algebraic topology. We will prove that contractibility is a sufficient condition for any CW-complex, in fact for a much larger class. We will also prove that for a large subclass contractibility is a necessary condition. Furthermore, we will see that on a zero dimensional metrizable and separable space there is always a social choice rule. Once we notice that the class of spaces for which there is for each n a social choice rule is closed under arbitrary products and under continuous retractions, we see that such spaces are many and that they might have little to do with contractible CW-complexes.

Showing that a given space admits a CW-structure might not be easy, and even more so if one has to exhibit a parafinite CW-structure. There is therefore some motivation for extending the result to a larger class.

There is another limitation to the result given by Chichilnisky and Heal. Parafinite CW-complexes are countable dimensional, they can be expressed as a countable union of finite dimensional spaces, (a finite CW-complex of dimension n can be embedded in $\mathbb{R}^{2n+1}$ and by definition a parafinite CW-complex is a countable union of finite CW-complexes, its skeletons), therefore, the theorem can only be applied to such spaces. Furthermore a parafinite finite dimensional CW-complex is compact.

Consequently, $\mathbb{R}^n$ is not a parafinite CW-complex.

In their paper they give as an example of a space for which there is a social rule a Hilbert space or the unit sphere in a Hilbert space. It is known that in a Banach space which is not finite dimentional the unit sphere is homeomorphic to the whole space. The existence of a social rule follows from the existence of the homeomorphism but it does not follow from their theorem since a Hilbert space which is not finite dimensional is not countable dimensional and therefore it cannot admit a parafinite CW-stucture.

There is also a problem with their characterization of contractible spaces of preferences. The result is right but their proof does not work. The crunch of their argument is that a subspace X_1 of a space X which is homeomorphic to a deformation retract X_2 of X is also a deformation retract of X. This is unfortunately false. We will provide a counterexample in the appendix.

There are other spaces which admit an obvious social rule without being countable dimensional, for example a convex subspace of a topological vector space, the Hilbert cube $[0,1]^{\mathbb{N}}$, or $[0,1]^E$ where E is an arbitrary infinite set. In this last case the existence of social choice rules is a consequence of the fact that the space is the product of spaces on which there are social choice rules. Also $[0,1]^{\mathbb{N}}$ does not have a CW-structure since a compact space with a CW-structure has to be a finite CW-space and therefore be finite dimensional.

The existence of a social rule is limited neither to infinite dimensional spaces nor to contractible spaces.

Consider the following two spaces: (a) the space of rational numbers with the topology induced by the topology of the real line, (b) the space of irrational numbers with the topology induced by the topology of the real line.

Neither of these spaces is contractible, both are finite dimensional and have nontrivial topologies and both admit social choice rules. In the first case because the average of n rational numbers is again a rational number and in the second case because the maximum value of n irrational numbers is again an irrational number.

Social choice rules on a topological space were studied maybe for the first time in 1935 by G. Auman. They are known in topology as *topological means*, or *n-means* if n is fixed. Aumann proved, [2], that there is no topological mean on the sphere, that any retract of a space with an n-mean also has an n-mean, and that each connected component of a space with an n-mean is also a space with an n-mean. The fundamental result on the homotopy and homology groups of topological spaces for which there is a social rule was established by B. Eckman in 1954 [15]. It is this result that was rediscovered by Chichilnisky and Heal in their proof of the necessity of contractibility.

Theorem 3 (Eckman). *(a) Let n be a fixed integer and let G be a group. Then, there is a group homomorphism $\Phi : G^n \to G$ which respects the conditions of anonymity and unanimity if and only if G is an abelian group and the function $x \mapsto n.x$ is a group isomorphism. In this case there is only one such function:*

$$(x_1, \ldots, x_n) \mapsto \frac{x_1 + \cdots + x_n}{n}.$$

(b) Let n be a fixed integer and X a topological space for which there is a social choice rule on X^n. Then, the first homotopy group of X is abelian, and on all the homotopy and homology groups of X, multiplication by n is a group isomorphism.

Corollary 1. *If for each positive integer n there is for the topological space X a social choice rule and if the homotopy groups of X are finitely generated then they are all trivial. The same holds for the singular homology groups.*

That can be proved as in Chichilnisky and Heal, or as in Eckman.

Eckman uses these results to show that a connected polyhedron admits for each n, an n-mean, if and only if it is contractible.

He also shows that contractibility is not a necessary condition for the existence of an n-mean for a given n.

Let us recall that a contractible space has trivial homotopy groups, but there are spaces which are not contractible and have trivial homotopy groups.

In this work we will generalize the theorem of Chichilnisky and Heal to a very large class of topological spaces and we will also generalize the notion of social rule that they consider.

Let us be more explicit on the second point. We view a social rule as a continuous function $\Phi : X^n \to X$ which is subject to certain constraints, for exemple unanimity and anonymity.

Each permutation σ of the first n positive integers gives rise to a continuous function $g_\sigma : X^n \to X^n$, and the condition of anonymity simply says that $\Phi(g_\sigma(x^n)) = \Phi(x^n)$ for any such function; here x^n denotes a typical element of X^n.

More generally we could consider constraints given by a set of continuous functions $g_i : X^n \to X^n$, $i \in I$. A function $\Phi : X^n \to X$ would then fulfil the constraints if for each index $i \in I$ one has $\Phi(g_i(x^n)) = \Phi(x^n)$.

If the set of functions $\{g_i; i \in I\}$ is a finite group of homeomorphisms of X^n or if it is a compact group acting continuously on X^n then the binary relation R defined on X^n by $(x^n, y^n) \in R$ if there exists $i \in I$ such that $g_i(x^n) = y^n$ is an equivalence relation with closed graph, such that for any closed subspace F of X^n the set

$$\bigcup\{R(x^n) : x^n \in F\}$$

is closed, where $R(x^n)$ is the equivalence class of x^n. An equivalence relation on X^n having the previous property will be called an *upper semicontinuous decomposition* of the space X^n. One can also formulate that property as follows: for any $x^n \in X^n$ and any open set $U \subseteq X^n$ such that $R(x^n) \subseteq U$ there is an open set W such that $x^n \in W$ and for any $y^n \in W$ $R(y^n) \subseteq U$. Such a relation could be of the following form: $(x^n, y^n) \in R$ if $p_i(x^n) = p_i(y^n)$ for each $i \in I$, where $p_i : X^n \to Z_i$ are continuous functions sending closed sets to closed sets. As a matter of fact, it is not hard to see that any upper semicontinuous decomposition is of this form.

Now we state the *generalized social choice problem*:

(a): we are given an integer n, the number of individuals, and a topological space P, the space of preferences.

(b): we are given an upper semicontinuous decomposition $D \subseteq P^n \times P^n$ such that for any $x \in P$, $D(x, \ldots, x) = \{(x, \ldots, x)\}$. D is interpreted as a family of constraints. Referring to the last property we will say that D is *unanimous*.

A continuous function $\Phi : P^n \to P$ which is unanimous and such that for any pair $(x^n, y^n) \in D$ one has $\Phi(x^n) = \Phi(y^n)$ will be called a *D-compatible social rule*.

The question is then the following: given a unanimous upper semicontinuous decomposition D of P^n, is there on P^n a D-compatible social rule?

This paper is organized as follows:

(a) In Sect. 2 we introduce the class of topological spaces in which the social choice problem can be treated with great generality. That class contains arbitrary CW-complexes and also arbitrary metric spaces. Theorem 4 shows that the existence of a social rule is, when the space is contractible, a matter of what the space is like locally. But even in the absence of contractibility the local structure of the space might be nice enough to allow for the existence of partial social choice functions. Throughout that section, contractibility is assumed and one could ask to what extent it is a necessary condition for spaces of a given class. We know, for example, from Chichilnisky and Heal, that is necessary in the class of parafinite CW-complexes. We show that the necessity of contractibility can be extended to a larger class. Majority rules are a special class of social choice rules. It turns out that the existence of a continuous majority function on the preference space when the number of individuals is three and when the space is path connected imposes very strong topological properties. We show also that when the preference space is contractible any two social choice functions can be continuously deformed one into the other while at each stage of the deformation the function remains a social choice function. The generalized social choice problem is treated in this section only. We leave it to the reader to see how the results of the other sections can be modified to deal with the generalized problem.

(b) In Sect. 3 we look at noncontractible spaces and at the existence of social choice rules which pick out one of the individual choices as the social choice.

(c) The results of Sect. 4 are of a more tentative kind. We address there the following points:

(i) How to extend the model to an infinite space of individuals?

(ii) Given that the space of individuals is finite but that their number can be arbitrary, is there a way to derive from a single function a social choice function Φ_n for each n?

(iii) Is it possible to introduce some dynamic in the model, that would take into account the change in the number of individuals over different periods and the change in their preferences?

2 Social choice rules on contractible topological spaces

The class of CW-complexes does not behave very well when it comes to basic topological constructions. For example a product of CW-complexes does not have to be a CW-complex or a subspace of a CW-complex does not have to be a CW-complex.

We introduce next a class of topological spaces which is, in that respect, much better behaved.

For the terms from general topology which are not defined here the book of Engelking is a good reference, for the more technical results one should look at the references.

All the topological spaces under consideration are at least Hausdorff. For a given topological space X let $\Gamma(X)$ be the family of its closed subspaces, and let $\Omega(X)$ be the family of its open subspaces.

The topological space X is *stratifiable*, see for example [5], if for each $n \in \mathbb{N}$ there is a function $G_n : \Gamma(X) \to \Omega(X)$ such that:

(i) if $F_1 \subseteq F_2$ then $G_n(F_1) \subseteq G_n(F_2)$.
(ii) $F = \bigcap_{n \in \mathbb{N}} \mathrm{cl}_X G_n(F)$.

The definition will not be used, what will be of importance to us are the basic properties of the class of stratifiable spaces listed below.

(S_1): A metric space is stratifiable.
(S_2): An arbitrary subspace of a stratifiable space is stratifiable.
(S_3): The topological product of a countable familly of stratifiable spaces is stratifiable.
(S_4): If $f : X \to Y$ is a continuous surjective map sending closed sets to closed sets and if X is stratifiable then Y is also stratifiable.
(S_5): Any CW-complex is stratifiable.
(S_6): A statifiable space is paracompact, and therefore normal.

All of these results can be found for example in [5] or in [26].

Let us now introduce some useful concepts from topology. Some more information can be found in the appendix and in the papers of Borges, Cauty, and Hanner listed in the bibliography, or in the book of Hu, the terminology of which we follow.

A topological space X is an **ANE**(*Stratifiable*), (*Absolute Neighbourhood Extensor for the class of stratifiable spaces*), if for any stratifiable space Y, any closed subspace A of Y, and any continuous function $g : A \to X$, there exists a neighbourhood U of A in X and a continuous function $f : U \to X$ whose restriction to A is g. If one can always take $U = Y$ then the topological space X is an **AE**(*Stratifiable*), (*Absolute Extensor for the class of stratifiable spaces*).

If in the preceding definitions we replace the word stratifiable by metrizable, we get the classes of **ANE**(*Metrizable*) and **AE**(*Metrizable*) spaces.

As examples of spaces which are **ANE**(*Stratifiable*) we have:

(1) Any CW complex.
(2) Any convex subset of a locally convex topological vector space.

The first proposition is due to Cauty [10] Corollary 2.4, and the second is due to Borges [5] Theorem 4.3.

From these two examples one can get a lot more from the following properties of the class of **ANE**(*Stratifiable*):

(P_1) Any open subspace of an **ANE**(*Stratifiable*) is also an **ANE**(*Stratifiable*).
(P_2) If each point of X has a neighborhood which is an **ANE**(*Stratifiable*) then X is an **ANE**(*Stratifiable*).
(P_3) Any finite product of **ANE**(*Stratifiable*) is an **ANE**(*Stratifiable*), and an arbitrary product of **AE**(*Stratifiable*) is an **AE**(*Stratifiable*).
(P_4) An **ANE**(*Stratifiable*) is an **AE**(*Stratifiable*) if and only if it is contractible.

Property (P_2) is a consequence of Theorem 19.2 in Hanner [19], and the fact that stratifiable spaces are paracompact. Property (P_4) follows from Theorem 12.3 of the same paper and the fact that paracompact spaces are normal.

The same properties hold for **ANE**(*Metrizable*).

A stratifiable space which is also an **ANE**(*Stratifiable*) will be called an **ANR**(*Stratifiable*), (*Absolute Neighbourhood Retract for the class of stratifiable spaces*).

A stratifiable space which is also an **AE**(*Stratifiable*) will be called an **AR**(*Stratifiable*), (*Absolute Retract for the class of stratifiable spaces*).

A stratifiable space will be an **AR**(*Stratifiable*) if and only if it is a contractible **ANR**(*Stratifiable*). Replacing stratifiable by metrizable we get **ANR**(*Metrizable*), and **AR**(*Metrizable*) spaces.

When the class of spaces is not specified it is understood that one is referring to the class of metrizable spaces, **ANR** stands for **ANR**(*Metrizable*), and **AR** stands for **AR**(*Metrizable*).

In [1], Allen showed that the space of complete preorders on $\mathbb{R}^n_+$ which are continuous and monotonic is, for the topology of closed convergence, a metrizable space homeomorphic to a convex subset of a locally convex topological vector space, namely the space of continuous functions from $\mathbb{R}^n_+$ into $\mathbb{R}$ with the compact open topology. She also shows that the same holds for the set of continuous preorders on $\mathbb{R}^n_+$ which are locally nonsatiated.

Consequently the space of complete preorders on $\mathbb{R}^n_+$ which are continuous and monotonic, or the space of continuous preorders on $\mathbb{R}^n_+$ which are locally nonsatiated, are **AR** for the topology of closed convergence.

The books of Borsuk [6], and Hu [20], are standard references for **ANR** and **AR** theory. The book of Van Mill [28], is more recent but deals only with separable metric spaces. See also the book by Bessaga and Pelczinski [4].

We come now to the main theorem of this section.

Theorem 4. *If the preference space P is an* **AR**(*Stratifiable*)*, then, for any* $n \in \mathbb{N}$ *and any unanimous upper semicontinuous decomposition D of* P^n*, there is a social rule on* P^n *which is D-compatible.*

Proof. First, let P^n/D be the quotient space of P^n with respect to the equivalence relation D, it has a natural topology, the quotient topology. Since D is an upper semicontinuous decomposition of P^n, the projection onto the quotient space sends closed subspaces to closed subspaces and P^n/D is also Hausdorff, (this is a consequence of the normality of P^n and of the upper semicontinuity of the decomposition D). The space P^n/D is therefore stratifiable.

Since D is the identity on the diagonal, the space P is homeomorphic to a subspace of P^n/D. The homeomorphism is the composition of the projection onto the quotient space with the function which identifies each point of P with the corresponding point of the diagonal of P^n. We will use the same symbol for the space P and its homeomorphic image in P^n/D.

If $\Psi : P^n/D \to P$ is a continuous function which coincides with the identity on the subspace P, then, by taking the composition with the projection from P^n onto P^n/D, we get a social rule which is D-compatible and unanimous.

But P is a closed subspace of P^n/D and the function $P \to P$ sending a point to itself is obviously continuous. Since, by hypothesis, P is an absolute extensor for the class of stratifiable spaces this function can be extended to a continuous function $\Psi : P^n/D \to P$. □

Corollary 2. *Assume that P is a connected CW-complex. Then:*

(A) If P is contractible, then, for any $n \in \mathbb{N}$ and any unanimous upper semicontinuous decomposition D of P^n, there is a social rule on P^n which is D-compatible.

(B) If the homotopy groups of P are finitely generated, or if P is simply connected and the homology groups are finitely generated, then contractibility is a necessary condition for the existence of a social choice rule on each P^n.

Proof.

(A) Borges has shown that CW-complexes are stratifiable, and Cauty proved that they are **ANE**(*Stratifiable*), a contractible CW-complex is therefore an **AR**(*Stratifiable*).

(B) If for each $n \in \mathbb{N}$ there is a social choice rule on P^n, then we know from Eckman's theorem that all the homotopy groups are commutative, and that on each of them multiplication by n is a group isomorphism. A finitely generated commutative group in which, for each $n \in \mathbb{N}$, multiplication by n is an isomorphism has to be the trivial group $\{0\}$. This follows from the structure theorem of finitely generated abelian groups, in [18], page 79, for example.

From a theorem of Whitehead [27] page 89, it follows that a connected CW-complex with trivial homotopy groups is contractible.

If the homology groups are finitely generated the same argument implies that they are all trivial. By Hurewicz isomorphism theorem, [27] page 185, all the homotopy groups are also trivial. □

Parafinite CW-complexes, they are also known as CW-complexes of *finite type*, have finitely generated homology groups. Indeed a finite CW-complex has finitely generated homology groups, and by definitinion the n-skeleton of a parafinite CW-complex is finite. Furthermore, for any CW-complex, the n-th homology group is completely determined by its $(n+1)$-skeleton. A finite complex also has finitely generated first homotopy groups. If there is a social choice rule, that first homotopy group is abelian and finitely generated. Therefore, is there is a social choice rule for each n, the first homotopy group is trivial, by the argument used in the previous theorem. The space is therefore simply connected.

Corollary 3. *Assume that P is a connected* **ANR**. *Then:*

(A) If P is contractible, then, for any $n \in \mathbb{N}$ and any unanimous upper semicontinuous decomposition D of P^n, there is a social rule on P^n which is D-compatible.

(B) If the homotopy groups of P are finitely generated, or if P is simply

connected and the homology groups are finitely generated, then contractibility is a necessary condition for the existence of a social choice rule on each P^n.

Proof. For the first part of the proof we proceed as previously, with the same notation, but now the quotient space P^n/D might not be metrizable so we cannot use right away the fact that P is an **AR**.

Recall that a topological space is *perfectly normal* if any closed subset is the intersection of countably many open sets.

Michael [25], proved that any **AR** is also an asolute extensor for the class of paracompact and perfectly normal spaces.

Stratifiable spaces are paracompact and perfectly normal, Ceder [12], or Borges [5].

But P^n/D is the image of the metric space P by a closed map, the projection. P^n/D is therefore stratifiable, and the identity can consequently be continuously extended to a map from P^n/D onto P.

For the second part we proceed as before taking into account that a connected **ANR** is contractible if and only if its homotopy groups are trivial, Hu [20] Corollary 8.5, or Van Mill [28], page 212. □

If we only want to consider social choice rules then the proof is simpler since the quotient space will be metrizable, and we do not have to use Michael's theorem.

Let us now give some examples of spaces to which the previous results could apply.

Assume that P is paracompact and each $x \in P$ has a neighborhood U which is homeomorphic to a retract of one of the following:

(a) an open subspace of a CW-complex.
(b) an open subspace of a normed space.
(c) a metrizable and locally contractible space of finite covering dimension.
(d) an open subspace of a topological cube $[0, 1]^E$ where the set E is at most countable.

In (a) we use the facts that a retract of an **ANE**(*Stratifiable*) is an **ANE**(*Stratifiable*) and that a paracompact space which is locally an **ANE**(*Stratifiable*) is also an **ANE**(*Stratifiable*), to conclude that P must be an **ANE**(*Stratifiable*).

In (b) to (d) we use the same facts for **ANR**, and also that a paracompact locally metrizable space is metrizable.

Topological manifolds, spaces in which each point has a neighbourhood homeomorphic to an open subset of some $\mathbb{R}^n$, or Banach manifolds, spaces in which each point has a neighbourhood homeomorphic to an open subset of a Banach space, fulfil one of the conditions (b) to (c).

Any locally contractible subset of $\mathbb{R}^n$ fulfils condition (c).

It is known that the set of unit vectors in a Banach space which is not finite dimensional is an **AR**, more generally, if from a closed ball in an infinite dimensional normed space one removes a family of disjoint open balls then the resulting space is an **AR** [14], page 94.

As a consequence of the previous theorems it follows that the unit sphere in $\mathbb{R}^n$, or more generally, any noncontractible topological manifold cannot have for any $n > 1$, a social choice rule.

Let us look a little bit more closely at metrizable spaces. Recall that contractible absolute neighbourhood retracts are absolute retracts. At this point one might wonder if the absolute neighbourhood retracts which have, for each positive integer, a social choice rule are not exactly the absolute retracts. We know that the answer is positive if the homotopy groups are finitely generated. The next result, which is due to Van Mill and Van de Vel [29], implies that connected metric spaces having a special kind of social choice rule have trivial homotopy groups.

First, let us say that a continuous function $\mu : X^3 \to X$ is a majority function if it is invariant under circular permutation of its arguments, and if for any $x, y \in X$ one has $\mu(x,x,y) = \mu(x,y,x) = \mu(y,x,x) = x$.

We will also need the following definition:

A metrizable space X is $\mathbf{C^n}$, provided that for every $0 \le m \le n$ every continuous function $f : S^m \to X$ from the m-dimensional sphere into X extends to a continuous function $g : B^{m+1} \to X$ on the unit ball. The space X is $\mathbf{C}^\infty$, provided it is $\mathbf{C^n}$ for every n.

It is known that an **ANR** is contractible if and only if it is $\mathbf{C}^\infty$.

Theorem 5 (Van Mill-Van de Vel). *Let X be a space which is compact, metrizable and has an open cover by path connected sets. If there is a majority function on X then X is* $\mathbf{C}^\infty$.

Furthermore, if X is finite dimensional and path connected, then there is a majority function on X if and only if X is an **AR**.

The result of Van Mill and Van de Vel implies that for some preference spaces having a majority rule when the number of individuals is exactly three is equivalent to having a social rule for any number of individuals. Before stating this formally we recall that the class of absolute neighbourhood retracts, **ANR**, is quite large. It contains open subspaces of normed spaces, manifolds, locally compact CW-complexes, any open subspace of an **ANR**, and is closed under finite products and continuous retractions.

Corollary 4. *Assume that P is a pathconnected compact* **ANR**. *Then if P has a majority function there is for each* $n \in \mathbb{N}$ *and any unanimous upper semicontinuous decomposition D of* P^n *a social rule on* P^n *which is D-compatible.*

Proof. By the theorem of Van Mill and Van de Vell the space P is $\mathbf{C}^\infty$. An **ANR** which is also $\mathbf{C}^\infty$ is an absolute retract.

The conclusion follows from Corollary 3. □

Let us remark that proving the existence of a social choice rule on an **AR** is straightforward:

Any metric space can be embedded as a closed subspace of a convex subset in a normed space. On a convex set there is for each n a social rule, for example the average of n elements. There is also a continuous retraction from the convex set onto the metric space, since it is an absolute retract. Now we define for each n a social choice rule on the metric space by taking first the average of n elements and then the image of that average by the retraction.

Corollary 5. *Assume that P is a finite dimensional pathconnected compact metric space. Then if P has a majority function there is for each $n \in \mathbb{N}$ and any unanimous upper semicontinuous decomposition D of P^n a social rule on P^n which is D-compatible.*

Proof. By the theorem of Van Mill and Van de Vell P is an absolute retract. □

Now we look at possible interpretations of homotopy.

Contractibility could be interpreted as some kind of uniqueness of the social rule and homotopy as some kind of dynamic transformation of one social rule in another, but this might be open to questions. Indeed, an arbitrary homotopy can be a very wild function and not only are two social choice rules homotopic, they are homotopic to any arbitrary continuous function from P^n into P, since if P is contractible then no matter what the topological space X is any two continuous functions from X into P are homotopic. If two social rules are homotopic but if the functions obtained during the deformation of one social rule in the other are arbitrary continuous functions it is hard to see what kind of interpretation could be given to such a transformation. What one at least needs is a homotopy for which at each stage of the deformation we have a social rule. We show that under the conditions of the previous theorem this is always possible.

Theorem 6. *Assume that the preference space P is an* **AR***(Stratifiable). Let D be a unanimous upper semicontinuous decomposition of P and let $\Phi_0, \Phi_1 : P^n \to P$ be two social choice rules that are D-compatible. Then there exists a homotopy $\Theta : [0,1] \times P^n \to P$ such that:*

(i) $\Theta_0 = \Phi_0$ and $\Theta_1 = \Phi_1$
(ii) For each $t \in [0,1]$ the function $\Theta_t : P^n \to P$ is a social rule which is D-compatible.

Proof. As before let P^n/D be the quotient space of P^n with respect to D, identify P with its image in the quotient space and let $p : P^n \to P^n/D$ be the projection. There are continuous functions $\hat{\Theta}_i : P^n/D \to P, i = 0, 1$ such that $\Phi_i = \hat{\Theta}_i \circ p$.

$(\{0,1\} \times P^n/D) \cup ([0,1] \times P)$ is a closed subspace of $[0,1] \times P^n/D$, call it Z. Now define a continuous function $\Lambda : Z \to P$ in the following way:

(a) $\Lambda(0, y) = \hat{\Theta}_0(y)$
(b) $\Lambda(1, y) = \hat{\Theta}_1(y)$
(c) $\Lambda(t, x) = x$

Since Z is a closed subspace of the stratifiable space $[0,1] \times P^n/D$ and P is by hypothesis an absolute extensor for the class of stratifiable spaces the function Λ has a continuous extension $\hat{\Theta}$ from $[0,1] \times P^n/D$ into P. The function $\Theta(t, x_1, \ldots, x_n) = \hat{\Theta}(t, p(x_1, \ldots, x_n))$ yields the homotopy Θ. $\square$

As in Corollary 3 the previous theorem holds if P is an **AR**.

Next we give two examples of social choice rules when the preference space is an **AR**, a metrizable and contractible CW-complex, or a contractible manifold for example.

The first examples shows that there is always for each $n \in \mathbb{N}$ a social choice rule Φ_n that depends only on the underlying set of individual preferences and not on the society's profile.

The second example shows that in case P is finite dimensional we always have a sequence of social choice rules $\Phi_n, n \in \mathbb{N}$, such that Φ_n and Φ_m coincide on any two profiles having the same underlying set, and furthermore there is a preference which is the social choice whenever it appears in the profile. That preference can be chosen a priori.

The constructions are not restricted to **AR** or to finite dimensional spaces, but they are simpler in this case.

(1) If the preference space is a compact absolute retract then there is for each n a social choice rule $\Phi_n : P^n \to P$ such $\Phi_n(x_1, \ldots, x_n) = \Phi_n(y_1, \ldots, y_n)$ whenever the sets $\{x_i : i = 1, \ldots, n\}$ and $\{y_i : i = 1, \ldots, n\}$ are equal.

This can be seen by noticing that the space of non empty closed subspaces of P is a metric space, with the Hausdorff distance, that it contains P as a closed subspace and that there is therefore a continuous retraction from the space of non empty closed subspaces onto P. The function which assigns to an element of P^n the set of its components is continuous. Composition with the previous retraction yields Φ_n.

(2) If the preference space is a finite dimensional compact absolute retract then there is for each n a social choice function $\Phi_n : P^n \to P$ such:

(a) $\Phi_n(x_1, \ldots, x_n) = \Phi_n(y_1, \ldots, y_n)$ whenever the sets $\{x_i : i = 1, \ldots, n\}$ and $\{y_i : i = 1, \ldots, n\}$ are equal.

(b) If $(x_1, \ldots, x_n)$ has exactly k different components $\{x_{i_1}, \ldots, x_{i_k}\}$ then $\Phi_n(x_1, \ldots, x_n) = \Phi_k(x_{i_1}, \ldots, x_{i_k})$

(c) There exists $x^* \in X$ such that if $x^* \in \{x_1, \ldots, x_n\}$ then $\Phi_n(x_1, \ldots, x_n) = x^*$

The construction of $\Phi_n(x_1, \ldots, x_n)$ is done as follows. First P can be identified with a compact subspace of some euclidean space E and since P is an absolute retract there is a retraction r from E onto P. Now let $\Psi_n(x_1, \ldots, x_n)$ be the projection of x^* on the convex hull of $\{x_1, \ldots, x_n\}$ and let $\Phi_n(x_1, \ldots, x_n)$ be the image under r of that projection. It can be shown that $\Phi_n(x_1, \ldots, x_n)$ depends continuously on $(x_1, \ldots, x_n)$.

We close this section with a result on partial social choice rules.

A *partial social choice rule* on P^n is given by a subset $S \subseteq P^n$ and a function $\Phi_n : S \to P$ such that for any permutation σ of $\{1, \ldots, n\}$ and any $(x_1, \ldots, x_n) \in S$ we also have $(x_{\sigma_1}, \ldots, x_{\sigma_n}) \in S$, ($S$ is invariant under the group of permutations), $(x, \ldots, x)$ always belongs to S, and the function Φ_n is anonymous and unanimous.

Theorem 7. *Assume that P is an* **AR***(Stratifiable) and that we are given a finite family $(F_i, \Phi_{n,i}), i = 1, \dots, m$ of partial social choice rules on P^n such that each F_i is closed in P^n and such that for each pair of indices (i, j) the functions $\Phi_{n,i}$ and $\Phi_{n,j}$ coincide on $F_i \cap F_j$. Then there exists a social choice rule on P^n whose restriction to F_i is $\Phi_{n,i}$ for each $i \in \{1, \dots, m\}$.*

Proof. Let P^n/S_n be the quotient of P^n under the action of the permutation group and as before identify P with a closed subspace of P^n/S_n.

The projection of F_i on P^n/S_n is a closed subspace, call it G_i.

From $\Phi_{n,i}$ we get a continuous function $\Psi_{n,i}$ on G_i, furthermore $\Psi_{n,i}$ and $\Psi_{n,j}$ coincide on $G_i \cap G_j$. To see this take $y \in G_i \cap G_j$. Then $y = p(x_1, \dots, x_n)$ and $y = p(x'_1, \dots, x'_n)$ where $(x_1, \dots, x_n) \in F_i$ and $(x'_1, \dots, x'_n) \in F_j$ and p is the projection.

Since $y = p(x_1, \dots, x_n) = p(x'_1, \dots, x'_n)$ there is a permutation σ such that $(x_1, \dots, x_n) = (x'_{\sigma_1}, \dots, x'_{\sigma_n})$ and therefore $(x_1, \dots, x_n) \in F_i \cap F_j$.

We now have $\Psi_{n,i}(y) = \Phi_{n,i}(x_1, \dots, x_n) = \Phi_{n,j}(x_1, \dots, x_n) = \Psi_{n,j}(y)$.

Let $G = \bigcup_{i=1}^{i=m} G_i$, it is a closed subspace P^n/S_n, and by taking $(\Psi_{n,0})_{|G_{n,i}} = \Psi_{n,i}$ we get a well defined and continuous function $\Psi_{n,0} : G \to P$. Since P is an **AR**(*Stratifiable*) this function has a continuous extension $\Psi_n : P^n/S_n \to P$ to all of P^n/S_n.

The composition of Φ_n with the projection of P^n onto P^n/S_n yields Φ_n. □

Corollary 6. *Assume that P is an* **AR***(Stratifiable). Then for any open neighbourhood U of P in P^n and for any $x^* \in P$ there is a social choice rule $\Phi_n : P^n \to P$ such that for any $(x_1, \dots, x_n) \in P^n \setminus U$ we have $\Phi_n(x_1, \dots, x_n) = x^*$.*

Proof. Let $H_1 = P \cup (P^n \setminus U)$ and let F_1 be the set of all permutations of elements of H_1, it is a closed subspace of P^n.

Now define $\Phi_{1,n}$ on F_1 by taking it to be identically x^* on $P^n \setminus U$ and apply the theorem with $m = 1$. □

Of course the theorem and its corollary hold for the generalized social choice problem and also if P is assumed to be an **AR**, see the proofs of 3.

The corollary could be given the following interpretation:

If the society's profile is too far from unanimity, (if it is not in the open subset U of P in P^n), then the choice is imposed: it is x^*.

For example if P is a metric space then U could be an ε-neighbourhood of P with ε arbitrarily small.

In conclusion if the preference space has a nice topological structure, a contractible CW-complex, a contractible manifold or more generally an **AR**(*Stratifiable*) or an **AR** then any social choice rule can be continuously deformed into a social rule having very peculiar properties. The interpretation of this is not so clear, at any rate translating contractibility of P as uniqueness of the social choice rule does not in view of this seem very natural. Let us also say that contractible spaces are very far from being topologically trivial and that they can indeed have very strange structures, like being the common boundary of three open disjoint sets in euclidean space.

The last example of this section provides a space P which is metric and contractible, but is neither a CW-complex nor an **AR**, so the results of this section do not apply in this case, and has for each n a social choice rule.

Let $I = [0,1]$, $J_\infty = \{0\} \times I$, and for each integer $n > 0$ let $J_n = \left\{\frac{1}{n}\right\} \times I$.

Take $C = (\bigcup_{i=1}^{i=\infty} J_i) \cup I$, it is the *comb space*, it is a contractible compact subspace of the plane. Compactness is obvious since it is closed and bounded. It is contractible since one can first retract C along the vertical segments onto the unit interval, and then retract the unit interval to the origin. It is not an **AR** since it is not even locally contractible, consider a point on J_∞, and therefore it is not a CW-complex, a CW-complex is locally contractible.

Now let us construct a social choice rule on C^n.

Let $x_i = (a_i, b_i), i = 1, \ldots n$ be points of C, and define

$\Phi_n(x_1, \ldots, x_n) = (\max(a_1, \ldots, a_n), \min(b_1, \ldots, b_n))$.

What makes things work in that example is clear. We have a topological space with an order structure for which the max and the min functions are continuous. Any topological space on which we would have a partial order with a continuous min function would support for each n a social choice rule. We will see partial orders reappear naturally in the next section where we will look for particular social choice rules.

Lastly let us say that there is an obvious drawback with the previous results, they are purely existential. Take for example the theorem of Van Mill and Van de Vel. From the theorem we know that there is a majority function on any Banach space, proving this is actually quite easy, but as they notice such a function has not been yet explicitly found.

So, on the one hand we know that there are social choice functions under very general hypotheses on the structure of the preference space, we also know that we can impose very general conditions on these social choice functions, anonymity being only one such set of conditions. But on the other hand, in the simple case of a majority function on a convex subspace of a Banach space we might be unable to produce an example. A social rule on a convex subspace of a topological vector space is easy to produce, take the average of the points. In the general case we might have to compose with a retraction, and even though for good enough spaces such retractions are plenty we might have little to say passed their existence.

3 Social choice rules on noncontractible spaces

It is rather obvious that there exist noncontractible spaces on which there is a social choice rule, just take any set with the discrete topology. A less trivial example is given by topological semilattices. A topological semillatice is a topological space with a partial order for which each pair of elements has a least upper bound and the function which assigns to a pair its least upper bound is continuous.

We will be concerned in this section exclusively with social choice rules, not with the generalized social choice problem.

Recall that in the introduction we gave two examples of non trivial and noncontractible topological spaces on which there is a social choice rule, the space of rational numbers and the space of irrational numbers. In one case the construction used the order structure of the space. Beside being ordered topological spaces the space of rationals and the space of irrationals are both zero dimensional spaces. Also there is a social choice rule of a very particular kind: if $(x_1, \dots, x_n) \in Z^n$, where Z is either the space of rational or irrational numbers, let $\Phi(x_1, \dots, x_n) = \max\{x_1, \dots, x_n\}$. So the rule Φ picks out one of the individual choices as the social choice. Let us ask on which topological spaces such a social choice function does exist.

For a topological space X let $\mathscr{F}_n(X)$ be the space of non empty subsets of X of cardinality at most n. It is a topological subspace of the space of non empty closed subspaces of X with the Vietoris topology.

The function from X^n to $\mathscr{F}_n(X)$ which assigns to the n-uple $(x_1, \dots, x_n)$ the set $\{x_1, \dots, x_n\}$ is continuous, therefore any continuous function from $\mathscr{F}_n(X)$ to X which picks out a member of the set will give through composition with the previous function a social choice rule. Such a continuous function will be called a *selection* of $\mathscr{F}_n(X)$. More generally a *selection of a subfamily* $\mathscr{C}$ of the space of non empty closed subspaces is a continuous function $f : \mathscr{C} \to X$ such that for any $C \in \mathscr{C}$ one has $f(C) \in C$.

We will also denote by $\mathscr{K}(X)$ the space of non empty compact subspaces of the topological space X with the Vietoris topology.

At this point it is natural to ask the following question: Given the integer n, for which topological spaces X is there a selection of $\mathscr{F}_n(X)$? When X is connected the answer depends essentially on what happens for $n = 2$.

Theorem 8 (Michael). *If X is a connected topological space there is either no selection of $\mathscr{F}_2(X)$ or there are exactly two selections.*

Each selection of $\mathscr{F}_2(X)$ has for each n a unique extension to $\mathscr{F}_n(X)$ and also an extension to $\mathscr{K}(X)$.

The implications of that result for the social choice problem are now given as a corollary.

Corollary 7. *If P is connected and if there is a social rule $\Phi_2 : P^2 \to P$ such that $\Phi_2(x_1, x_2) \in \{x_1, x_2\}$ for any pair (x_1, x_2), then there is for each integer n a unique social rule Φ_n such that*

(a) $\Phi_n(x_1, \dots, x_n) = \Phi_m(y_1, \dots, y_m)$ *whenever the sets* $\{x_i : i = 1, \dots, n\}$ *and* $\{y_i : i = 1, \dots, m\}$ *are equal.*

(b) $\Phi_n(x_1, \dots, x_n) \in \{x_i : i = 1, \dots, n\}$.

Proof. On P^n we define an equivalence relation by identifying two points having the same set of coordinates. For Hausdorff spaces it is known that the quotient space is homeomorphic to $\mathscr{F}_n(P)$, see for example [17] or [7]. Since Φ_2 is compatible with the equivalence relation on P^2 and is continuous we get a selection of $\mathscr{F}_2(P)$, call it Ψ_2. By the theorem that selection has for each n a unique extension to a selection Ψ_n of $\mathscr{F}_n(P)$. The composition of that selection with the function which assigns to a point of P^n the set of its coordinates is Φ_n. The uniqueness of the rules Φ_n is a consequence of the uniqueness of the selections Ψ_n, since (a) and (b) imply that each Φ_n induces a continuous function on the quotient space of P^n which is a selection of $\mathscr{F}_n(P)$ extending Ψ_2. □

Compact connected metric spaces for which $\mathscr{F}_2(X)$ has a selection are known.

Theorem 9 (Kuratowski-Nadler-Young). *If X is metric compact and connected and if there exists a selection of* $\mathscr{F}_2(X)$ *then X is homeomorphic to the unit interval* $[0, 1]$.

If X is metric locally compact and separable and if there is a selection of $\mathscr{F}_2(X)$ *then X is homeomorphic to a subspace of the real line.*

On the other hand, if the space is zero dimensional, for example the space of rational numbers or the space of irrational numbers then there is a selection of $\mathscr{F}_n(X)$ for each integer n. More generally we have the following result.

Theorem 10 (Kuratowski-Ryll-Nardzeswki). *If X is a separable zero dimensional metric space then there exists a selection of the space of nonempty closed subspaces of X.*

The rationals and the irrationals are linearly ordered topological spaces. It turns out that the existence of a selection of $\mathscr{F}_n(X)$ is closely linked to the existence of a linear order which is compatible with the topology of X.

Theorem 11 (Michael). *Let X be a Hausdorff topological space. Then for a selection of* $\mathscr{K}(X)$ *to exist it is sufficient, and if the connected components of X are open in X it is also necessary, that there exists a linear ordering of X for which the order topology is coarser than the topology of X.*

We have seen how to obtain social choice rules by looking for selections of $\mathscr{F}_n(P)$, such social choice rules elect one of the individual preferences as the social preference and they are associated to linear orderings of the preference space.

Now consider a social choice rule $\Phi_2 : P^2 \to P$.

This is the same as a continuous retraction from $\mathscr{F}_2(P)$ onto P. For example if P is the circle then it is known that $\mathscr{F}_2(P)$ is homeomorphic to the Moebius strip and that a retraction of $\mathscr{F}_2(P)$ onto P would give a retraction of the Moebius strip on its boundary and this is known to be impossible. This is the argument that is used by Candeal and Indurain [8].

More generally we could look for a continuous retraction from $\mathscr{F}_n(P)$ onto P. Such a retraction composed with the function from P^n to $\mathscr{F}_n(P)$ sending an n-uple to the set of its elements would yield a social choice rule.

It would be interesting to find necessary and sufficient conditions for such a retraction to exist. It seems to be a difficult problem. A sufficient condition that has been used in the preceding section is that the space be an **AR**.

The next result owes some of its inspiration to the theory of games.

If a strategy space is not convex then by going from pure to mixed strategies we introduce convexity and then the existence of Nash equilibria can be established.

A similar idea will be used here. The preference space P will be replaced by the space of probability measures on P.

We recall some known facts.

Let X be a topological space and $\mathscr{B}$ the σ-algebra of its Borel sets.

$Pr(X)$ will be the space of probability measures on $(X, \mathscr{B})$. Recall that a sequence of probability measures $\mu_n \in Pr(X)$ converges to $\mu \in Pr(X)$ if for any bounded continuous function $f : X \to \mathbb{R}$ the sequence $\int f \, d\mu_n$ converges to $\int f \, d\mu$.

To an element x of X we can associate a probability measure, δ_x the Dirac measure concentrated at x. One can see that $\delta : X \to Pr(X)$ is continuous. The topological space will be identified with the set of Dirac measures on X.

$Pr_f(X)$ will be the space of probability measures with finite support on X.

Now consider a probability space $(\Omega, \mathscr{A}, \mu)$, and let $L^0(\Omega, X)$ be the space of equivalence classes of measurable functions. If X is a separable metric space, with metric d, recall that a sequence of measurable function $f_n : \Omega \to X$ converges in probability to a measurable function f_0 if for every $\varepsilon > 0$, $\mu\{d(f_n, f_0) > \varepsilon\}$ converges to 0 as n goes to ∞. It is known that on $L^0(\Omega, X)$ convergence in probability is metrizable, for example by the Ky Fan metric $d_{KF}(f, g) = \inf\{\varepsilon > 0 : \mu\{d(f, g) > \varepsilon\} \leq \varepsilon\}$, here $d(f, g)$ stands for the measurable function $\omega \mapsto d(f(\omega), g(\omega))$.

To each element f of $L^0(\Omega, X)$ one can associate its law $\mu \circ f^{-1} \in Pr(X)$, moreover with respect to the Ky Fan metric on $L^0(\Omega, X)$ and the Prohorov metric on $Pr(X)$ the function $f \mapsto \mu \circ f^{-1}$ is continuous.

Theorem 12. *Assume that P is a separable* **ANR**. *Then for each $n \in \mathbb{N}$ there exist a continuous function $\Phi_n : P^n \to Pr(P)$, neighbourhoods W_1 and W_2 of P in P^n, and $x^* \in P$ such that:*

(a) *For any permutation $(m_1, \ldots, m_n)$ of $(1, \ldots, n)$ one has $\Phi(x_1, \ldots, x_n) = \Phi(x_{m_1}, \ldots, x_{m_n})$.*
(b) *For any $x \in P$ one has $\Phi(x, \ldots, x) = x$.*
(c) $\Phi_n(W_1) \subseteq P$
(d) *For any $(x_1, \ldots, x_n) \in (P^n \backslash W_2)$ we have $\Phi(x_1, \ldots, x_n) = x^*$.*

Proof. Since P is an **ANR** there is an open neighbourhood W_2 of P in P^n and a continuous function $\Phi : W_2 \to P$ which is a partial social choice rule.

To see this we first consider P^n/S_n, the quotient space of P^n under the group of permutations, it contains P as a closed subspace. Since P is an **ANR** there is neighbourhood U of P in P^n/S_n and a continuous function $f : U \to P$ sending each point of P to itself. Let W_2 be the inverse image of U under the projection on the quotient space, and let Φ be the composition of the projection with f.

Let W_1 be a closed neighbourhood of P, which we can assume to be the inverse image of a closed neighbourhood of P in P^n/S_n, contained in W_1.

Now we have two partial social choice rules, $\Phi_{n,1} : W_1 \to P$ which is the restriction of Φ to W_1 and $\Phi_{n,2} : P^n \backslash W_2 \to P$ which is the constant function x^*.

The space of measurable functions $L^0([0,1], P)$ is an **AR**, Bessaga and Pelczinski [4] Theorem 7.1, and if we identify each point of P with the corresponding constant function from $[0,1]$ into P we can identify the previous two partial social choice functions with two functions $\Phi_{n,1} : W_1 \to L^0([0,1], P)$ and $\Phi_{n,2} : (P^n \backslash W_2) \to L^0([0,1], P)$.

Now, as before we let P^n/S_n be the quotient space of P^n under the group of permutations, and as we deed in the previous section we match the two functions to obtain a function $\Phi_n^* : P^n \to L^0([0,1], P)$ for which (c) and (d) are satisfied.

We let $\Phi_n(x_1, \cdot, x_n) = \lambda \circ (\Phi(x_1, \dots, x_n))^{-1}$ where λ is the Lebesgue measure on $[0,1]$. □

Here there is no contractibility. P could be any topological manifold, any metrizable CW-complex, any open subset of a normed space, any finite product of such spaces, or any retract of such spaces. For example any locally contractible subset of $\mathbb{R}^n$

That theorem says that we can always find a rule which is a social choice rule as long as society's profile is not too far from unanimity. If the profile is far from unanimity then society's choice is imposed. There is an in between zone where to each society's profile is associated a probability distribution on the preference space P. It could be shown that the probability distribution can always be chosen to have finite support.

4 Extension of the social choice problem to infinitely many individuals and other generalizations

In this section we will consider exclusively social choice rules.

Once we have for each $n \in \mathbb{N}$ a social choice rule $\Phi_n : P^n \to P$ it is natural to inquire into the evolution of the society's choice as the number of individuals evolves along with their individual choices. We consider that time is divided into periods, during each period the number of individuals is considered constant as well as the corresponding social choice. The dynamic is therefore represented by the parameters n_k, the number of individuals in period k and $(x_1, \dots, x_{n_k})$, the individual choices in period k. Society's choice in period k is therefore the element $\Phi_{n_k}(x_1, \dots, x_{n_k})$ of P. We have a discrete dynamical system on P given by $f(k) = \Phi_{n_k}(x_1, \dots, x_{n_k})$.

Within the purely abstract and topological framework offered by the previous theorems very little can be done, since there is no relationship between the different Φ_n.

On the other hand we will show that for separable metric spaces a much more satisfactory approach is possible. The construction will also show that all the Φ_n can be derived from a single function.

As previously if X is a topological space and $\mathscr{B}$ the σ-algebra of its Borel sets then $Pr(X)$ will be the space of probability measures on $(X, \mathscr{B})$.

To an element $(x_1, \ldots, x_n)$ of X^n we can associate a probability measure $\alpha_n((x_1, \ldots, x_n)) = \frac{1}{n}\sum_{i=1}^{i=n} \delta_{x_i}$, where δ_x is the Dirac measure concentrated at x. One can see that $\alpha_n : X^n \to Pr(X)$ is continuous.

Let X^ω be the disjoint union of the topological spaces X^n, $n \in \mathbb{N}$.

The function $\alpha : X^\omega \to Pr(X)$ whose restriction on X^n is α_n is continuous.

The evolution of the society's profile is represented by a function $\gamma : \mathbb{N} \to P^\omega$. For each $k \in \mathbb{N}$ we have an integer n_k, the size of the population in period k and an element $\gamma(k) \in P^{n_k}$, the society's profile in period k.

Theorem 13. *Let P be a separable* **AR**. *Then there exist social choice rules $\Phi_n : P^n \to P$ such that the following holds:*

If $\gamma : \mathbb{N} \to P^\omega$ is such that the sequence of probability measures $\alpha(\gamma(k))$ converges to a probability measure $\mu \in Pr(P)$ then the sequence $\Phi_{n_k}(\gamma(k))$ converges in P.

Proof. Since P is metrizable and separable, convergence of laws in $Pr(P)$ is metrizable, by the Prohorov metric. The function $\delta : X \to Pr(P)$ which sends the point $x \in P$ to the Dirac measure δ_x is a continuous embedding of P into $Pr(P)$. Since P is an absolute retract there is a continuous function $\Psi : Pr(P) \to P$ such that $\Psi(\delta_x) = x$ for any $x \in P$.

Let $\Phi : P^\omega \to Pr(P)$ be the composition of Ψ with α and finally let Φ_n be the composition of Φ with the embedding $j_n : P^n \to P^\omega$.

Now, let $\gamma : \mathbb{N} \to P^\omega$ be such that the sequence of laws $\alpha(\gamma(k))$ converges to $\mu \in Pr(P)$. Then the sequence $\Psi(\alpha(\gamma(k)))$ converges in P to $\Psi(\mu)$.

But $\Psi(\alpha(\gamma(k))) = \Phi_{n_k}(\gamma(k))$. To complete the proof one only has to see that the functions Φ_n are continuous, anonymous and unanimous. □

It is interesting to notice that within the previous context uncertainty can be accomodated. Indeed, assume that there is some uncertainty on the number of individuals and their individual preferences. That uncertainty can be represented by a probability measure on P^ω with finite support. We will see that one can associate to such a random profile a preference in such a way that by restriction to P^n, which can be identified with a space of Dirac measures on P^ω, we get a social choice rule on P^n.

As previously let $Pr_f(P^\omega)$ be the space of probability measures with finite support on P^ω.

Corollary 8. *If P is a separable* **AR** *then there is a continuous function $\Lambda : Pr_f(P^\omega) \to P$ such that $\Lambda \circ d_n$ is a social choice rule for each $n \in \mathbb{N}$, where d_n is the embedding of P^n into $Pr(P^\omega)$ which sends a point to the corresponding Dirac measure.*

Proof. If $\mu \in Pr_f(X^\omega)$ then $\mu = \sum_{i=1}^{i=m} t_i\delta_{u_i}$ where $u_i \in X^{n_i}$, $0 \le t_i$ and $\sum_{i=1}^{i=m} t_i = 1$.

Now let $\Lambda(\mu) = \Phi(\sum_{i=1}^{i=m} t_i\alpha(u_i))$, one can see that Λ is a continuous function and that $\Lambda \circ d_n = \Phi_n$, where Φ_n is the social choice rule constructed in the previous theorem. □

In case the number of individuals is fixed we can consider arbitrary probability measures. There are then two forms of uncertainties. There could be an uncertainty on the society's profile as a whole and that would be represented by a probability measure on P^n or there could be uncertainties on each individual's preference and that would be represented by an n-uple of probability measures, an element of $(Pr(P))^n$. We show below that in each case there is an appropriate social choice function. If μ is a measure on P^n and σ is a permutation of the first n integers, μ_σ is the image of the measure μ under the obvious function from P^n into itself associated to σ.

Theorem 14. *Assume that P is a separable* **AR**. *Then*

(A) *There is an embedding $j_n : P^n \to Pr(P^n)$ and a continuous function $\Phi : Pr(P^n) \to P$ such that for each permutation σ of $\{1, \ldots, n\}$ and each $\mu \in Pr(P^n)$ one has $\Phi(\mu_\sigma) = \Phi(\mu)$ and $\Phi \circ j_n : P^n \to P$ is a social choice rule.*

(B) *There is an embedding $k_n : P^n \to (Pr(P))^n$ and a continuous function $\Phi : (Pr(P))^n \to P$ such that for each permutation σ of $\{1, \ldots, n\}$ and each $(\mu_1, \ldots, \mu_n) \in (Pr(P))^n$ one has $\Phi(\mu_1, \ldots, \mu_n) = \Phi(\mu_{\sigma_1}, \ldots, \mu_{\sigma_n})$ and $\Phi \circ k_n : P^n \to P$ is a social choice rule.*

Proof.

(A) j_n assigns to $(x_1, \ldots, x_n)$ the Dirac measure $\delta_{(x_1,\ldots,x_n)}$.

Let S_n be the group of permutations of the first n integers and denote by $s_n : Pr(P^n) \to Pr(P^n)$ the operation of symetrization, $s_n(\mu) = \frac{1}{n!}\sum_{\sigma \in S_n}(\mu_\sigma)$.

For each $i \in \{1, \ldots, n\}$ let $p_i : P^n \to P$ be the corresponding projection and for $\mu \in Pr(P^n)$ let $\mu \circ p_i^{-1} \in Pr(P)$ be the image of the measure μ.

With $\Psi : Pr(X) \to X$ being the function from Theorem 13 let $\Phi(\mu) = \Psi\left(\frac{1}{n}\sum_{i=1}^{i=n} s_n(\mu) \circ p_i^{-1}\right)$.

(B) k_n assigns to $(x_1, \ldots, x_n)$ the n-uple of Dirac measures $(\delta_{x_1}, \ldots, \delta_{x_n})$.

The function $\Phi : (Pr(P))^n \to P$ is then given by $\Phi(\mu_1, \ldots, \mu_n) = \Psi\left(\frac{1}{n}\sum_{i=1}^{i=n} \mu_i\right)$. □

Now we presente a framework for the social choice problem which could accomodate an infinite number of individual choices. There is some arbitrariness in that model inasmuch as we have to pick a probability space for which there is no natural choice. Nevertheless, that models captures the idea that what matters in a society's profile is the probability measure that it induces on the preference space, this is anonymity, the measure of a subset of the preference space could then be interpreted as the proportion of individuals whose preferences are in that subset.

The space of individuals is a probability space $(\Omega, \mathscr{A}, \mu)$, if the number of individuals is finite, say n, Ω is the set of the first n integers with the uniformly distributed probability measure.

A society's profile is then a random variable $f : \Omega \to P$, P is of course a topological space and we consider measurability with respect to Borel sets, or more exactly an equivalence class of measurable functions, two functions being equivalent if they are equal almost everywhere with respect to μ, $L^0(\Omega, P)$ is the space of equivalence classes of measurable functions.

Now we have to introduce in this context the concepts of anonymity and unanimity in such a way that they are compatible with the finite case and uniform probability measure.

A social choice rule assigns in a continuous way a preference, an element of P, to a society's profile, an element of $L^0(\Omega, P)$.

Consider a continuous function $\Phi : L^0(\Omega, P) \to P$.

(a) Φ is unanimous if for any constant function $c_x : \Omega \to P$ one has $\Phi(c_x) = x$.
(b) Φ is anonymous if $\Phi(f) = \Phi(g)$ whenever $\mu \circ f^{-1} = \mu \circ g^{-1}$.

Theorem 15. *If (P, d) is a separable* **AR** *and $(\Omega, \mathscr{A}, \mu)$ an arbitrary probability space then there exists a function $\Phi : L^0(\Omega, P) \to P$ which is anonymous and unanimous.*

Proof. There is a continuous function $\mathscr{L} : L^0(\Omega, P) \to Pr(P)$ which assigns to a random variable f its law $\mu \circ f^{-1}$. Taking the composition with the function $\Psi : Pr(P) \to X$ from Theorem 13 we obtain the function Φ. □

If the metric space (P, d) is not only separable but also complete then, on the one hand the spaces $L^0(\Omega, P)$ and $Pr(P)$ are also complete repectively for the Ky Fan metric and for the Prohorov metric and on the other hand any probability measure $v \in Pr(P)$ is the law of a random variable $f : [0, 1] \to P$ (this result is due to Halmos and Von Neumann). So if the preference space is separable and completely metrizable we have at our disposal a universal space of individuals, the unit interval with the Lebesgue measure. Notice also that we can derive anonymous and unanimous social choice functions $\Phi_n : P^n \to P$ from the single function $\Phi : L^0([0, 1], P) \to P$.

Indeed, to $(x_1, \ldots, x_n)$ one associates the function $f_{(x_1,\ldots,x_n)} : [0, 1] \to P$ whose law is given by $v_{(x_1,\ldots,x_n)}(B) = \sum_{i=1}^{i=n} \frac{1}{n}\delta_{x_i}(B)$ for a Borel subset B of P. It is clear that $v_{(x_1,\ldots,x_n)} = v_{(x_{\sigma_1},\ldots,x_{\sigma_n})}$ for any permutation σ and that $v_{(x,\ldots,x)} = \delta_x$.

Now, let $\Phi_n(x_1, \ldots, x_n) = \Phi(f_{(x_1,\ldots,x_n)}) = \Psi(v_{(x_1,\ldots,x_n)})$.

5 Conclusion

The topological social choice problem, at least as far as the existence of a social choice rule is concerned, is a problem about continuous extensions of a continuous function given on a closed subspace of a topological space. Spaces for which this kind of problem has a solution with respect to a given class of domains, the absolute extensors for the domains in question, have under mild conditions to be contractible, see Hanner [19] or Hu [20]. This explains the pervasiveness of the contractibility condition.

On the other hand, if we restrict our attention to spaces which have a particular structure, zero dimensional spaces, topological semilattices or products or retracts of spaces thereof, contractibility is not needed.

Even when the space is quite general and not contractible something of interest can be said, if P is an **ANE**(*Stratifiable*) then there is a partial social choice rule.

Another possibility is to restrict the set of individual preferences. There is no general framework for doing that, it is best understood on a particular case.

Let us say that P is the m-dimensional sphere, choose an arbitrary point of P and let W be the space obtained by removing that single point from P. Then there is a social choice rule on W since it is homeomorphic to euclidean space.

It might be interesting to study the preservation of the existence, or non existence, of social choice rules with respect to some kind of topological modification of the preference space, the robustness problem. But the kind of conclusion that one could expect is not quite clear. Indeed as we have just seen by removing a single point one can go from nonexistence to existence. The other direction is also possible, remove a single point from a euclidean cube or a euclidean sphere, if that point is on the boundary then existence of a social choice function is preserved, on the other hand if that point is an interior point then there is no social choice function on the new space. On the other hand remove any point from the Hilbert cube, there is still a social choice function. The difference here comes from the specific topological properties of the spaces under consideration, in one case we have a space that is not topologically homogeneous and in the other case we have a space that is topologically homogeneous. One could also make "big holes", for example by removing the interior of the unit ball in a Banach space which is not finite dimensional then one gets the unit sphere for which there is a social choice function since it is an absolute retract.

Let us conclude by saying that we do not feel that there is anything paradoxical in the results that have been obtained, quite to the contrary. If we expect the social choice problem to have a solution within the class of CW-complexes, or within the much larger class of absolute neighborhood extensors for stratifiable spaces, then contractibility is natural, as a matter of fact it is necessary if we want to go from neighborhood extensors to extensors, see Hanner [19]. It also seems that anonymity and unanimity are not very stringent restrictions. The generalized problem shows that one could had infinitly many conditions of the same type and still have a compatible social choice rule.

6 Appendix

If $\mathscr{C}$ is a class of topological spaces, for example the class of metrizable or stratifiable spaces, a topological space X is called an *absolute neighbourhood extensor for the class* $\mathscr{C}$ if for any space Y of $\mathscr{C}$ and any closed subspace

$A \subseteq Y$, any continuous function $f : A \to X$ has a continuous extension to a neighbourhood U of A in Y, if one can always take $U = Y$ then X is called an *absolute extensor for the class* $\mathscr{C}$.

A space X of the class $\mathscr{C}$ is an *absolute neighbourhood retract for the class* $\mathscr{C}$ provided it is a neighbourhood retract of every space Y in $\mathscr{C}$ containing it as a closed subspace. If it is a retract of any space in $\mathscr{C}$ containing it as a closed subspace then it is an *absolute retract for the class* $\mathscr{C}$.

A subspace X of Y is a *neighbourhood retract* of Y if there is a neighbourhood U of X in Y and a continuous function $r : U \to X$ such that $r(x) = x$ for any $x \in X$, if one can take $U = Y$ then X is a *retract* of Y.

If $\mathscr{C}$ is either the class of metrizable spaces or the class of stratifiable spaces then a contractible absolute neighbourhood retract is an absolute retract and the absolute neighbourhood retracts are exactly the absolute neighbourhood extensors which belong to $\mathscr{C}$. This statement combines results of J. Dugundji, for metrizable spaces, and C. Borges, for stratifiable spaces.

A metrizable space X is $\mathbf{C^n}$ provided that for every $0 \leq m \leq n$ every continuous function $f : S^m \to X$ from the m-dimenional sphere into X extends to a continuous function $g : B^{m+1} \to X$ on the unit ball. The space X is $\mathbf{C}^\infty$ provided it is $\mathbf{C^n}$ for every n.

The space X is $\mathbf{LC^n}$ if for every $x \in X$ and for every neighbourhood U of x and for every $0 \leq m \leq n$ there exists a neighbourhood V of x such that every continuous function $f : S^m \to V$ has a continuous extension $g : B^{m+1} \to U$.

If the metric space X is of dimension at most n, with $n < \infty$, then it is an absolute neighborhood retract if and only if it is $\mathbf{LC^n}$ and it is an absolute retract if and only if it is $\mathbf{C^n}$.

For the definitions of CW-complexes and relative CW-complexes one can look at [27] page 65 and page 71.

Now we give an example of a space X and two subspaces A and B such that A is a deformation retract of X, B is homeomorphic to A and B is not a deformation retract of X. As a matter of fact we give two such examples, the first one is rather simple but the spaces under consideration are not contractible.

(1) X is the real line with the point zero removed, $A = \{-1, 1\}$, $B = \{1, 2\}$.
(2) $X = [0,1] \times [0,1]^{\mathbb{R}}$, $A = \{\frac{1}{2}\} \times [0,1]^{\mathbb{R}}$ $B = ([0,1] \times \{0\}) \cup (\{0\} \times [0,1]^{\mathbb{R}})$.

The function $h : [0,1] \times X \to X$ defined by $h(t,s,x) = \left((1-t)s + \frac{t}{2}, x\right)$ is a deformation retraction of X onto A.

$[0,1]^{\mathbb{R}}$ is homeomorphic to $[0,1]^{\mathbb{R}} \times [0,1]$ and B is homeomorphic to $(\{\frac{1}{2}\} \times [0,1]^{\mathbb{R}}) \cup (\{\frac{1}{2}\} \times [0,1])$, B is therefore homeomorphic to A. But B is not a retract of X, see [20] proposition 12.6.

References

[1] Allen B (1996) A remark on the social choice problem. Soc Choice Welfare 13: 11–16

[2] Aumann G (1943) Über Räume mit Mittelbildungen. Math Ann 119: 210–215
[3] Baryshnikov YM (1993) Unifying impossibility theorems: A topological approach. Adv App Math 14: 404–415
[4] Bessaga C, Pelczinski A (1975) Selected topics in infinite dimensional topology. Polish Scientific Publishers
[5] Borges CJR (1966) On stratifiable spaces. Pacific J Math 17: 1–16
[6] Borsuk K (1966) Theory of retracts. Polish Scientific Publishers
[7] Borsuk K, Ulam S (1931) On symmetric products of topological spaces. Bull AMS 37: 875–882
[8] Candeal JC, Indurain E (1994) The Moebius strip and a social choice paradox. Econ Lett 45: 407–412
[9] Cauty R (1972) Sur le prolongement des fonctions continues à valeurs dans les CW-complexes. CR Acad Sci Paris 274: 271–274
[10] Cauty R (1973) Convexité topologique et prolongement des fonctions continues. Composito Math V27: 233–271
[11] Cauty R (1976) Sur les espaces d'applications dans les CW-complexes. Arch Math 27: 306–311
[12] Ceder JG (1961) Some generalizations of metric spaces. Pacific J Math 11: 105–126
[13] Chichilnisky G, Heal G (1983) Necessary and sufficient condition for a resolution of the social choice paradox. J Econ Theory 31: 68–87
[14] Dugundji J, Granas G (1982) Fixed point theory. Polish Scientific Publishers
[15] Eckman B (1954) Räume mit Mittelbildungen. Comment Math Helv 28: 329–340
[16] Engelking R (1989) General topology. Heldermann Verlag
[17] Ganea T (1954) Symmetrische Potenzen Topologischer Räume. Math Nach 11: 305–316
[18] Fuchs L (1970) Infinite abelian groups. V.I. Academic Press
[19] Hanner O (1952) Retraction and extension of mappings of metric and nonmetric spaces. Ark Mat 16: 315–360
[20] Hu SC (1965) Theory of retracts. Wayne State University Press
[21] Kuratowski K, Nadler SB, Young GS (1970) Continuous selections on locally compact separable metric spaces. Bull Acad Pol Sci 17: 5–11
[22] Kuratowski K, Ryll-Nardzewski C (1965) A general theorem on selectors. Bull Acad Pol Sci 13: 397–403
[23] Le Breton M, Uriarte JR (1990) On the robustness of the impossibility result in the topological approach to social choice. Soc Choice Welfare 7: 131–140
[24] Michael E (1951) Topologies on spaces of subsets. Trans Amer Math Soc 71: 152–182
[25] Michael E (1953) Some extension theorems for continuous functions. Pac J Math 3: 789–806
[26] Nagata J (1985) Modern general topology. North-Holland
[27] Switzer RM (1975) Algebraic topology-homotopy and homology. Springer, Berlin Heidelberg New York
[28] van Mill J (1989) Infinite-dimensional topology. North-Holland
[29] van Mill J, van de Vel M (1982) On an internal property of absolute retracts II. Topology Appl 13: 59–68

Soc Choice Welfare (2001) 18: 251–267

Social Choice and Welfare

A linear algebra approach to non-transitive expected utility

Matjaž Omladič[1], **Vesna Omladič**[2]

[1] Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, Ljubljana, Slovenia (e-mail: Matjaz.Omladic@uni-lj.si)
[2] Faculty of Social Sciences, University of Ljubljana, Kardeljeva ploščad 5, Ljubljana, Slovenia (e-mail: Vesna.Omladic@uni-lj.si)

Received: 14 February 1996/Accepted: 22 January 1997

Abstract. The purpose of this paper is twofold. One of the aims is to give a view of P. C. Fishburn's non-transitive expected utility through geometric linear algebra. The other is to give an application of these ideas to a social choice problem. Actually, the problem under consideration arises from the theory of dynamical systems of structures in a society as studied in [All] and [Om1]. A variation of this model as presented in [Om2] leads to the problem of how macro structures of a society form coalitions. Their possible irrational behaviour may be treated using non-transitive expected utility.

1 Introduction

Decision theory is one of the central sources of applications of mathematics in the social sciences. One of the classic approaches to decision making uses von Neumann-Morgenstern linear utility theory. This approach is based on some assumptions including transitivity of the considered preference relation that have been criticized during the last few decades (see [Fi2], [Fi3], [FuI], [Rap], say). Besides statistically analyzed behavioral evidence showing that some of the underlying suppositions are violated by preference relations of many individuals, there may be other reasons for this dissatisfaction. Namely, applications in social choice theory lead to various impossibility results such as the famous Arrow's impossibility theorem [Arr], thus raising doubts about their foundations.

We study here an approach that was proposed by Fishburn [Fi2]. Instead of defining the preference relation $\succ$ by letting $p \succ q$ for alternatives p, q iff $u(p) > u(q)$, where u is a real function defined on the set of alternatives and called a "utility function", he proposes to study a relation $\succ$ given by a function Φ defined on pairs of alternatives. Such a function represents the relation by $p \succ q$ iff $\Phi(p,q) > 0$. The von Neumann-Morgenstern assumption of "convexity" or linearity which says that $u(\lambda p + (1-\lambda)q) = \lambda u(p) + (1-\lambda)u(q)$ for $\lambda \in (0,1)$, where $\lambda p + (1-\lambda)q$ denotes a lottery yielding alternative p with probability λ and alternative q with probability $1-\lambda$, enables us to compute "expected utility". By this we mean the mathematical expectation of the utility of any lottery p yielding alternative p_i with probability λ_i, where $\lambda_i > 0$ for $i = 1,2,\ldots,n$ and $\sum_i \lambda_i = 1$. Convexity yields $E(p) = \sum_i \lambda_i u(p_i) = u(p)$ by induction. Extending this situtation, Fishburn assumes that the functional Φ is convex or linear in both factors thus giving an appealing formula for the expected "paired" utility $E(p,q) = \sum_{ij} \lambda_i \Phi(p_i, q_j)\mu_j = \Phi(p,q)$. Here, p is a lottery yielding alternative p_i with probability λ_i, where $\lambda_i > 0$ for $i = 1,2,\ldots,n$ and $\sum_i \lambda_i = 1$, while q is a lottery yielding alternative q_j with probability μ_j, where $\mu_j > 0$ for $j = 1,2,\ldots,m$ and $\sum_j \mu_j = 1$. He also gives necessary and sufficient conditions, called axioms, on the relation $\succ$ for the existence of a functional Φ that represents $\succ$ as above.

The aim of our paper is to present a point of view on his approach which is closer to linear algebra. Our considerations use slightly different axioms and give the analogous result in a way that may be somewhat shorter. We believe that this is a natural framework for that theory. Basic definitions and assumptions as well as the comparison between the original axioms and our axioms are given in Sect. 2, the consequences of the first two axioms in Sect. 3, the main result in Sect. 4, and some applications in Sect. 5. Section 6 deals with an application to a social choice problem arising from a model of dynamics for power and control in society. The problem deals with the way that macro structures of a society form their coalitions. We study their irrational behaviour using the techniques presented in the paper.

2 Comparison of the two approaches

We call the *convex combination* of two points p and q of a real vector space any vector of the form $\lambda p + (1-\lambda)q$ for $\lambda \in [0,1]$. The set of all convex combinations of these two points will be denoted by $[p,q]$ and called a *segment*. The set of all convex combinations of this kind for $\lambda \in (0,1)$ will be denoted by (p,q). A subset of a real vector space will be called *convex* whenever it contains the segment between any two of its points.

Let P be a convex subset of a real vector space. Assume that a binary relation $\succ$ is given on the set P. Using this relation we define two new relations on P. First, for $p,q \in P$ let $p \sim q$ whenever neither $p \succ q$ nor $q \succ p$ and second, let $p \succeq q$ whenever either $p \succ q$ or $p \sim q$. For a given $p \in P$, the set of all $q \in P$ such that $q \sim p$ will be denoted by $C(p)$. This set is sometimes referred to as *the contour set* of p. Also, $U(p)$, *the upper contour set* of p, is defined as the set of all $q \in P$ such that $q \succ p$ and $L(p)$, *the lower contour set* of p, is the set of all $q \in P$ such that $p \succ q$. Notice that for any $p \in P$ the set P is a disjoint union of these three sets when the relation $\succ$ is asymmetric. We are asking for necessary and sufficient conditions for existence of a form Φ on

$P \times P$ such that

(a) Φ is skew-symmetric, i.e., $\Phi(p,q) = -\Phi(q,p)$ for all $p,q \in P$;
(b) Φ is convex in the first argument, i.e., $\Phi(\lambda p + (1-\lambda)q, r) = \lambda\Phi(p,r) + (1-\lambda)\Phi(q,r)$ for all $\lambda \in (0,1)$ and $p,q,r \in P$;
(c) Φ determines the relation $\succ$ via $p \succ q$ if and only if $\Phi(p,q) > 0$.

Observe that requirements (a) and (b) automatically imply:

(b′) Φ is convex in the second argument, i.e., $\Phi(r, \lambda p + (1-\lambda)q) = \lambda\Phi(r,p) + (1-\lambda)\Phi(r,q)$ for all $\lambda \in (0,1)$ and $p,q,r \in P$.

This approach was introduced in [Fi2]. There, it was shown that such a function exists if and only if the following three axioms, called "continuity", "dominance", and "symmetry", are fulfilled:

Axiom C. *For any $p,q,r \in P$ such that $p \succ q$ and $q \succ r$ there exists a $\lambda \in (0,1)$ such that $q \sim \lambda p + (1-\lambda)r$.*

Axiom D. *For any given $p \in P$ the set $C(p)$ is convex. Also, if $q \in U(p)$ and $r \in C(p) \cup U(p)$, then $(q,r) \subset U(p)$, and if $q \in L(p)$ and $r \in C(p) \cup L(p)$, then $(q,r) \subset L(p)$.*

Axiom S. *For any given $p,q,r \in P$ such that $p \succ q$, $q \succ r$, $p \succ r$, and that $q \sim \frac{1}{2}p + \frac{1}{2}r$, it is true that $\lambda p + (1-\lambda)r \sim \frac{1}{2}p + \frac{1}{2}q$ if and only if $\lambda r + (1-\lambda)p \sim \frac{1}{2}q + \frac{1}{2}r$.*

We give a proof of a similar result using geometric aspects of linear algebra more substantially. To this end we will find a bilinear extension of the form Φ defined on the linear span of P. In order to be able to do so we will have to require that the set P *avoids the origin of the space*. By this we mean that kP, i.e. the set of all points kp when p runs through P, has empty intersection with P, whenever $k \neq 1$. Further, we will suppose that the binary relation $\succ$ on P is *non-trivial* in the sense that the set of all points $p \in P$ having non-empty upper and lower contour set spans the whole space. Our axioms will be somewhat different from the original ones and the proofs may be somewhat shorter. We will assume that the relation $\succ$ (together with the corresponding relations $\sim$ and $\succeq$) satisfies the following conditions:

Axiom 1. *For any $p,q,r \in P$ such that $p \succ q$ and $q \succ r$ there exists a unique $\lambda \in (0,1)$ such that $q \sim \lambda p + (1-\lambda)r$.*

The unique λ of this axiom will be denoted by $\phi_q(p,r)$. We will also denote $\phi_q^-(p,r) = \dfrac{1-\phi_q(p,r)}{\phi_q(p,r)}$ and $\phi_q^+(p,r) = \dfrac{\phi_q(p,r)}{1-\phi_q(p,r)}$. We can extend partially these definitions to the case $p \sim q$ and $q \succ r$ by putting $\phi_q(p,r) = 1$ and $\phi_q^-(p,r) = 0$, and to the case $p \succ q$ and $q \sim r$ by putting $\phi_q(p,r) = 0$ and $\phi_q^+(p,r) = 0$.

Axiom 2. *For any given $p \in P$ the set $C(p)$ is convex.*

Axiom 3. *For any given* $p, q, r, s \in P$ *such that* $p \succ q$, $q \succ r$, $r \succ s$, $p \succ r$, $q \succ s$, *and such that the segment* $[q, r]$ *is parallel to the segment* $[p, s]$, *it is true that* $\phi_q(p, r) + \phi_r(q, s) = 1$ *if and only if* $\phi_q(p, s) + \phi_r(p, s) = 1$.

The term "parallel" is used in Axiom 3 in the usual geometric sense. Observe that these axioms imply that the relation $\succ$ is asymmetric. Indeed, if $p \succ q$ and simultaneously $q \succ p$, Axiom 1 with $r = p$ implies the existence of a unique λ such that $q \sim \lambda p + (1 - \lambda)p = p$, contradicting the definition of relation $\sim$. Note also that Axiom 1 is a simple consequence of Axioms C and D as noticed in [Fi2], so that a relation $\succ$ satisfying those two axioms is asymmetric as well.

It seems to us that "most" of Axiom D is redundant. Indeed, we will show in Sect. 3 that our Axioms 1 and 2 imply Axiom D for all $p \in P$ having non-empty both lower and upper contour set (see Lemma 4). Although Axiom 2 contains only "one third" of Axiom D, the two triples of axioms are almost, but not completely equivalent. Namely, at the end of Sect. 4 we will give an (artificial) example of a relation satisfying Axioms 1, 2, and 3, which does not satisfy Axioms C, D, and S. The idea of the example will be to take "many" points with empty upper and lower contour set. However, if the relation is non-trivial in the above sense, then the two triples of axioms are equivalent. Axiom S seems to us the one that can be modified in the largest variety of possible ways. Nevertheless, let us point out that the structure presumed for our main theorem (see Sect. 4) is stronger than Fishburn's, so in a sense the original result is more inclusive.

3 Implications of continuity and dominance

We assume in this section that P is a convex subset of a real vector space that avoids the origin of the space. Further, we suppose that there is a binary relation $\succ$ given on P. Using this relation we define two new relations $\sim$ and $\succeq$ on P in the usual way. We also assume that the relation $\succ$ (together with the corresponding relations $\sim$ and $\succeq$) satisfies Axioms 1 and 2 of Sect. 2.

It is not hard to see that $U(p) \cup C(p)$ is always convex. Namely, let q, r be any two points in this set. If both belong to $C(p)$, so does the segment $[q, r]$ by Axiom 2. Thus, assume that at least one of them, say q, belongs to $U(p)$, and suppose that a point from (q, r), say s, is not in $U(p) \cup C(p)$, so it is in $L(p)$. It follows by Axiom 1 that a point from (q, s), say t, belongs to $C(p)$. Now, $r \in C(p)$ yields $s \in [r, q] \subset C(p)$ by Axiom 2 in contradiction to $s \in L(p)$. So, $r \in U(p)$, and by Axiom 1 a point from $[s, r]$, say u, belongs to $C(p)$. This implies by Axiom 2 that $[t, u]$ belongs to $C(p)$, contradicting the fact that it contains s which is in $L(p)$. Similar arguments show that $L(p) \cup C(p)$ is convex. Under certain conditions we can actually show more.

Lemma 1. *If for some* $p \in P$ *the set* $L(p)$ *is non-empty, then* $U(p)$ *is convex. If* $U(p)$ *is non-empty, then* $L(p)$ *is convex.*

Proof. Choose a point s from the non-empty set $L(p)$. If $U(p)$ contains less than two points, it must be convex. So, assume there are two different points $q, r \in U(p)$. To show that this set is convex, it suffices to see that the segment $[q, r]$ is its subset. Let us confine ourselves to the case in which P is the triangle generated by the points q, r and s. Notice that the points from the segment $[q, s]$ of the form $q_\lambda = \lambda q + (1 - \lambda)s$ with $\lambda > \phi_p(q, s)$ belong to $U(p)$. Indeed, if this were not so, a point of the kind would belong to $C(p)$ (since $C(p) \cup U(p)$ is convex), contradicting uniqueness of λ in Axiom 1. Similarly, $q_\lambda \in [q, s]$ with $\lambda < \phi_p(q, s)$ belong to $L(p)$. Now, $C(p)$ contains a point, say t, from (q, s), and also a point, say u, from (r, s). Thus, it contains $[u, t]$ by Axiom 2. However, the segment $[r, q_\lambda]$ has exactly one intersection with $C(p)$ for every λ going from 0 to $\phi_p(q, s)$. So, these intersections all lie on $[u, t]$ and consequently, $[u, t] = C(p)$. Therefore, the segment $[q, r]$ belongs to $U(p)$. □

Recall that a functional $\theta(p)$, defined on a convex set, is called *convex* whenever $\theta(\lambda p + (1 - \lambda)q) = \lambda\theta(p) + (1 - \lambda)\theta(q)$ for all $\lambda \in (0, 1)$ and all p, q from the domain of θ.

Lemma 2. *Let $p \in P$ be such that $U(p)$ and $L(p)$ are both non-empty.*

(a) *For any $s \in L(p)$, the functional $\phi_p^-(q, s)$ is convex for $q \in U(p) \cup C(p)$.*
(b) *For any $s \in U(p)$, the functional $\phi_p^+(s, q)$ is convex for $q \in L(p) \cup C(p)$.*

Proof. Let us rewrite the two assertions of the lemma using the above definition of convexity of a functional.

(a′) If $q, r \in U(p) \cup C(p)$ and $s \in L(p)$ then for every $\lambda \in (0, 1)$ it is true that $\phi_p^-(\lambda q + (1 - \lambda)r, s) = \lambda\phi_p^-(q, s) + (1 - \lambda)\phi_p^-(r, s)$.
(b′) If $q, r \in L(p) \cup C(p)$ and $s \in U(p)$ then for every $\lambda \in (0, 1)$ it is true that $\phi_p^+(s, \lambda q + (1 - \lambda)r) = \lambda\phi_p^+(s, q) + (1 - \lambda)\phi_p^+(s, r)$.

The assumption in (a′) together with the above notation shows that

$$\phi_p(q, s)[q + \phi_p^-(q, s)s] \sim p$$

and

$$\phi_p(r, s)[r + \phi_p^-(r, s)s] \sim p.$$

Multiply the left-hand side of the first of these relations by $v\phi_p(r, s)\lambda$ and the second one by $v\phi_p(q, s)(1 - \lambda)$, where positive v is chosen in such a way that the sum of these two coefficients equals 1. By Axiom 2 it follows that

$$v\phi_p(q, s)\phi_p(r, s)[(\lambda q + (1 - \lambda)r) + (\lambda\phi_p^-(q, s) + (1 - \lambda)\phi_p^-(r, s))s] \sim p,$$

which gives the desired conclusion. Assertion (b′) is proved similarly. □

Lemma 3. *For any $q, r \in U(p)$ and $s, t \in L(p)$,*

$$\phi_p^+(q, s)\phi_p^-(q, t) = \phi_p^+(r, s)\phi_p^-(r, t).$$

Proof. For any $\lambda, \mu \in (0, 1)$ we have

$$1 = \phi_p^+(\lambda q + (1 - \lambda)r, \mu s + (1 - \mu)t)\phi_p^-(\lambda q + (1 - \lambda)r, \mu s + (1 - \mu)t),$$

and by Lemma 2 we have

$$1 = [\mu\phi_p^+(\lambda q + (1-\lambda)r, s) + (1-\mu)\phi_p^+(\lambda q + (1-\lambda)r, t)]$$
$$\times [\lambda\phi_p^-(q, \mu s + (1-\mu)t) + (1-\lambda)\phi_p^-(r, \mu s + (1-\mu)t)].$$

Use Lemma 2 again to see that

$$1 = \left[\frac{\mu}{\lambda\phi_p^-(q,s) + (1-\lambda)\phi_p^-(r,s)} + \frac{(1-\mu)}{\lambda\phi_p^-(q,t) + (1-\lambda)\phi_p^-(r,t)}\right]$$
$$\times \left[\frac{\lambda}{\mu\phi_p^+(q,s) + (1-\mu)\phi_p^+(q,t)} + \frac{(1-\lambda)}{\mu\phi_p^+(r,s) + (1-\mu)\phi_p^-(r,t)}\right].$$

Insert $\lambda = \mu = \frac{1}{2}$ into this equation and denote $\alpha = \phi_p^-(q,s)$, $\beta = \phi_p^-(r,s)$, $\gamma = \phi_p^-(q,t)$ and $\delta = \phi_p^-(r,t)$ to get

$$1 = \left[\frac{1}{\alpha+\beta} + \frac{1}{\gamma+\delta}\right]\left[\frac{\alpha\gamma}{\alpha+\gamma} + \frac{\beta\delta}{\beta+\delta}\right].$$

Let us rearrange this equation into

$$(\alpha+\beta)(\gamma+\delta)(\alpha+\gamma)(\beta+\delta) - (\alpha+\beta+\gamma+\delta)(\alpha\beta\gamma + \alpha\beta\delta + \alpha\gamma\delta + \beta\gamma\delta) = 0.$$

The left-hand side of this equation may be viewed as a quadratic polynomial in α. Its coefficient at α^2 is $(\gamma+\delta)(\beta+\delta) - (\beta\gamma + \beta\delta + \gamma\delta) = \delta^2$. Its coefficient at α equals $(\beta+\gamma)(\gamma+\delta)(\beta+\delta) - \beta\gamma\delta - (\beta+\gamma+\delta)(\beta\gamma+\beta\delta+\gamma\delta) = -2\beta\gamma\delta$. And, its constant coefficient is $\beta\gamma(\gamma+\delta)(\beta+\delta) - (\beta+\gamma+\delta)\beta\gamma\delta = (\beta\gamma)^2$. The equation therefore simplifies to $(\alpha\delta - \beta\gamma)^2 = 0$. The definitions of α through δ yield the desired conclusion. □

Lemma 4. *For any $p \in P$ such that $L(p)$ and $U(p)$ are both non-empty, there exists a linear functional v_p, defined on the linear span of P, such that $v_p(q) > 0$, respectively $v_p(q) = 0$, respectively $v_p(q) < 0$, if and only if $q \in U(p)$, respectively $q \in C(p)$, respectively $q \in L(p)$. If w is any linear functional with this property, defined on the span of P, then there exists a positive constant c such that $w = cv_p$ on the span of P.*

Proof. Choose a point $q \in U(p)$ and a point $s \in L(p)$. Define $v_p(t) = \phi_p^-(t,s)$ for $t \in U(p)$, $v_p(t) = 0$ for $t \in C(p)$, and $v_p(t) = -\phi_p^+(q,t)\phi_p^-(q,s)$ for $t \in L(p)$. By Lemma 2 the partial definitions of v_p are giving convex functionals for $t \in U(p) \cup C(p)$ as well as for $t \in L(p) \cup C(p)$. Moreover, the third partial definition is independent of the choice of $q \in U(p)$ by Lemma 3. We show that the functional v_p is convex on P. Choose any $r, t \in P$. We would like to show that $v_p(\lambda r + (1-\lambda)t) = \lambda v_p(r) + (1-\lambda)v_p(t)$ for all $\lambda \in (0,1)$. If the points r and t simultaneously belong to either $U(p) \cup C(p)$, or $L(p) \cup C(p)$, then this equation is satisfied automatically by the above. Let us now consider the case in which one of them, say r, is an element of $U(p)$, while the other, say t, belongs to $L(p)$. It follows that

$$\phi_p(r,s)r + (1-\phi_p(r,s))s \sim p \quad \text{and} \quad \phi_p(r,t)r + (1-\phi_p(r,t))t \sim p. \qquad (*)$$

Fix $\lambda \in (0,1)$ and let $r_\lambda = \lambda r + (1-\lambda)t$. If $r_\lambda \in C(p)$, we have $\lambda = \phi_p(r,t)$, so that $\lambda v_p(r) + (1-\lambda)v_p(t) = \phi_p(r,t)\phi_p^-(r,s) - (1-\phi_p(r,t))\phi_p^+(r,t)\phi_p^-(r,s) = 0 = v_p(r_\lambda)$. Next, consider the case of $r_\lambda \in U(p)$. This implies that $\lambda > \phi_p(r,t)$. Multiply the left-hand side of the first relation of $(*)$ by $v(\lambda - (1-\lambda)\phi_p^+(r,t))/\phi_p(r,s)$ (which is necessarily positive as soon as v is positive) and the left-hand side of the second relation in $(*)$ by $v(1-\lambda)/(1-\phi_p(r,t))$, where v is chosen in such a way that the sum of these two coefficients is 1, and sum them up. By Axiom 2 we have that

$$v(\lambda r + (1-\lambda)t) + v(\lambda\phi_p^-(r,s) - (1-\lambda)\phi_p^+(r,t)\phi_p^-(r,s))s \sim p$$

and the desired conclusion follows. It remains to consider the case of $r_\lambda \in L(p)$. In this case rewrite the second of the two relations in $(*)$ as

$$\frac{\phi_p(r,t) - \lambda}{1-\lambda} r + \frac{1-\phi_p(r,t)}{1-\lambda}(\lambda r + (1-\lambda)t) \sim p.$$

By definition of v_p we get $\lambda v_p(r) + (1-\lambda)v_p(t) = \phi_p^-(r,s)(\lambda - (1-\lambda)\phi_p^+(r,t)) = \phi_p^-(r,s)(\lambda - \phi_p(r,t))/(1-\phi_p(r,t)) = v_p(r_\lambda)$ and the convexity of v_p on P is proved. The linear span of P clearly equals the set of all vectors of the form $\alpha q - \beta r$ when q and r run through P, and α and β run through positive reals. Define $v_p(\alpha q - \beta r) = \alpha v_p(q) - \beta v_p(r)$. If $\alpha q - \beta r = \gamma s - \delta t$, we must have $(\alpha+\delta)(\lambda q + (1-\lambda)t) = (\beta+\gamma)(\mu r + (1-\mu)s)$, where $\lambda = \alpha/(\alpha+\delta)$ and $\mu = \beta/(\beta+\gamma)$. Since P avoids the origin of the space, it follows that $\alpha+\delta = \beta+\gamma$ and that therefore $\alpha v_p(q) - \beta v_p(r) = \gamma v_p(s) - \delta v_p(t)$. This shows that v_p is well-defined. Since it is necessarily additive, positively homogeneous and homogeneous in -1, it is linear. Uniqueness up to a multiplicative constant now follows easily. □

Lemma 5. *Let $p,q,r \in P$ be such that $p \succ q$, $q \succ r$, and $r \succ p$. Then $\phi_p^+(r,q)\phi_q^+(p,r)\phi_r^+(q,p) = 1$.*

Proof. Assume with no loss of generality that P is the triangle generated by the points p,q,r. By Lemma 4, $C(p)$, $C(q)$ and $C(r)$ are segments of straight lines connecting corresponding vertices with the opposite sides. We will show that the three segments have a common point of intersection and the lemma will then follow by an exercise in elementary geometry. Denote by t the point of intersection of $C(r)$ and $C(q)$ and by s the point of intersection of $C(p)$ with the side $[q,r]$. By definition, $t \sim q$ and $t \sim r$, and, by Axiom 2, $C(t)$ contains the triangle generated by q,r,t, so that $s \sim t$, since s belongs to this triangle. On the other hand $s \in L(q)$ and $s \in U(r)$, so that $U(s)$ and $L(s)$ are both non-empty and Lemma 4 tells us that $C(s)$ is a segment. Hence, $C(p) = C(s)$ and it contains t, as was to be proved. □

4 Implications of symmetry

In this section we assume that Axiom 3 holds in addition to Axioms 1 and 2.

Lemma 6. *For any given $p, q, r, s \in P$ such that $p \succ q$, $q \succ r$, $r \succ s$, $p \succ r$, $q \succ s$, and such that the segment $[q, r]$ is parallel to the segment $[p, s]$, it is true that $\phi_q^+(p, r)\phi_r^+(q, s) = \phi_r(p, s)/(1 - \phi_q(p, s))$.*

Proof. If $\alpha := \phi_q(p, s) + \phi_r(p, s) = 1$, the lemma follows easily from Axiom 3. Thus, we need only consider the cases of $\alpha < 1$ and $\alpha > 1$. Assume $\alpha < 1$ first and define $u = \alpha p + (1 - \alpha)s$. Then, $q \sim t := \phi_q(p, s)p + (1 - \phi_q(p, s))s = \dfrac{\phi_q(p, s)}{\alpha}u + \dfrac{\phi_r(p, s)}{\alpha}s$, and similarly $r \sim \dfrac{\phi_r(p, s)}{\alpha}u + \dfrac{\phi_q(p, s)}{\alpha}s$, so that $\phi_r(u, s) + \phi_q(u, s) = 1$. Now, $C(q)$ meets the segment $[u, s]$ in the point t that is further away from p than u. So, we have $u \succ q$ and similarly $u \succ r$. Axiom 3 with p interchanged by u then tells us that

$$\phi_r^+(q, s) = \phi_q^-(u, r). \qquad (*)$$

Now, $p, t \in C(q) \cup U(q)$ and u is clearly a convex combination of p and t, namely, $u = \dfrac{\phi_r(p, s)}{1 - \phi_q(p, s)}p + \dfrac{1 - \alpha}{1 - \phi_q(p, s)}t$. Use this convex combination in $(*)$ and apply Lemma 2 to see that $\phi_r^+(q, s) = \phi_q^-(u, r) = \dfrac{\phi_r(p, s)}{1 - \phi_q(p, s)}\phi_q^-(p, r)$ and the lemma follows.

The case of $\alpha > 1$ goes similarly. This time, let $u = (\alpha - 1)p + (2 - \alpha)s$ and observe that $q \sim \dfrac{1 - \phi_r(p, s)}{2 - \alpha}p + \dfrac{1 - \phi_q(p, s)}{2 - \alpha}u$, and $r \sim \dfrac{1 - \phi_q(p, s)}{2 - \alpha}p + \dfrac{1 - \phi_r(p, s)}{2 - \alpha}u$, so that $\phi_r(p, u) + \phi_q(p, u) = 1$. Apply Axiom 3 with s replaced by u to see that $\phi_q^-(p, r) = \phi_q^+(q, u)$. Insert $u = \dfrac{1 - \alpha}{\phi_r(p, s)}t + \dfrac{1 - \phi_q(p, s)}{\phi_r(p, s)}s$, where $t = \phi_r(p, s)p + (1 - \phi_r(p, s))s \sim r$, into this equation and apply Lemma 2 to get $\phi_q^-(p, r) = \dfrac{1 - \phi_q(p, s)}{\phi_r(p, s)}\phi_q^+(q, s)$, so that Lemma 6 follows in this final case as well. □

Lemma 7. *For any fixed $p, r \in P$ such that $p \succ r$ the function $\phi_q^+(p, r)$ is a continuous function of q such that $p \succ q$ and $q \succ r$.*

Proof. Fix $p, q, r \in P$ as in the lemma and let (q_n) be a sequence of points from P converging to the point q and such that $p \succ q_n$ and $q_n \succ r$. Let u denote the point in which $C(q)$ meets the segment $[p, r]$. Moreover, let (u_n) denote the sequence of points in which $C(q_n)$ respectively meet the segment $[p, r]$. To get the Lemma it suffices to see that u_n converges to u. Now, choose any $s \in (p, u)$ and any $t \in (u, r)$. Since s lies on the same side of $C(q)$ as p, we have $s \succ q$ and similarly $q \succ t$. Since $p \succ q$, we have $s \succ u$. Hence $C(s)$, being a (subset of a) hyper-plane by Lemma 4, does not meet the segment $[q, s]$. Thus, the terms of (q_n) that are close enough to q are on the same side of $C(s)$ as q. Therefore, the corresponding terms of (u_n) lie in $[s, u]$. Similarly, the terms of (u_n) with large enough indices lie in $[u, t]$. The lemma now follows by letting s and t go to u. □

Lemma 8. *For any given $p, q, r, s, t \in P$ such that $p \succ q$, $q \succ t$, $t \succ r$, $r \succ s$, $p \succ t$, $q \succ r$, $t \succ s$, $p \succ r$ and $q \succ s$, it is true that* $\phi_q^+(p,t)\phi_t^+(q,r)\phi_r^+(t,s) = \phi_q^+(p,r)\phi_r^+(q,s)$.

Proof. Assume at first that the segment $[q,r]$ is parallel to the segment $[p,s]$, the segment $[q,t]$ is parallel to the segment $[p,r]$, and the segment $[t,r]$ is parallel to the segment $[q,s]$. Use Lemma 6 once with r replaced by t and once with q replaced by t to get

$$\phi_q^+(p,t)\phi_t^+(q,r)\phi_r^+(t,s) = \phi_t^-(q,s)\phi_t^-(p,r)\phi_t^+(q,r)\phi_t^+(p,s)\frac{\phi_r(p,s)}{1-\phi_q(p,s)}.$$

Lemma 3 now tells us that $\phi_q^+(p,t)\phi_t^+(q,r)\phi_r^+(t,s) = \dfrac{\phi_r(p,s)}{1-\phi_q(p,s)}$, and another application of Lemma 6 finishes the proof of this case.

Consider now the general situation. Rewrite the desired equation as

$$\phi_t^+(q,r) = \frac{\phi_q^+(p,r)\phi_q^-(p,t)}{\phi_r^-(q,s)\phi_r^+(t,s)}. \qquad (*)$$

We first show that we may assume with no loss of generality that the points q and r belong to the interior of the set P. Namely, by Lemma 2, the denominator of the right-hand side of $(*)$ is a convex and therefore continuous functional of $q \in U(r)$. Similarly, the left-hand side is continuous in $q \in U(t)$. By Lemma 7 the two functions in the numerator of the right-hand side are continuous in q. Thus, if Eq. $(*)$ holds for all the appropriate points q in the interior of P, it must be true for a boundary point q as well. Similar arguments apply to the point r. So, assume from now on that points q and r belong to the interior of P. Similar continuity arguments show that t may be assumed not to belong to the segment $[q,r]$, so that the points q, t and r generate a plane.

Observe that by Lemma 3 the numerator of the right-hand side of Eq. $(*)$ is independent of the choice of $p \in U(q)$. Since q is an interior point of P, it follows that $U(q)$ has non-empty intersection with the plane determined by q, r and t. So we may and will assume that the point p belongs to this plane. Similarly, apply Lemma 3 to the denominator of the right-hand side of Eq. $(*)$ to see that it is independent of the choice of $s \in L(r)$, and assume with no loss of generality that s belongs to the same plane by the fact that $L(r)$ has non-empty intersection with that plane. So, we may restrict ourselves to the case in which P itself is a subset of that plane. Now, denote by u the point of intersection of $C(t)$ with the segment $[q,r]$; actually, we can write it down: $u = \phi_t(q,r)q + (1-\phi_t(q,r))r$. Thus, if we replace t in the right-hand side of Eq. $(*)$ by u and apply Lemma 2, we get $\phi_t^+(q,r)$. Observe by Lemma 2 that the right-hand side of $(*)$ is a fraction of two convex functions in t. Hence, if $(*)$ is true, the right-hand side of $(*)$ has equal values at the endpoints of the segment $[t,u]$ and is therefore constant on this segment. The same conclusion holds if t in the right-hand side of $(*)$ is replaced by any point from (t,u). So, after we fix the segment $[t,u]$, it suffices to prove $(*)$ with t replaced by any point from (t,u). Finally, let the point t go towards u (in the direction of this fixed segment). Then the line parallel to segment $[q, t]$ through the point r rotates towards the line containing segment $[q, r]$. Since $C(q)$ is a line meeting this segment only in the point q, the rotating line must eventually meet $U(q)$ in a point, say p. Similarly, the line parallel to segment $[t, r]$ through the point q rotates towards the line containing segment $[q, r]$. Since $C(r)$ is a line meeting this segment only in the point r, the rotating line must eventually meet $L(r)$ in a point, say s. Going with t even further towards u, if necessary, we may rotate the two lines even more, so that we may assume that the segment $[p, s]$ is parallel to the segment $[q, r]$. Thus, we have reduced the general case to the particular case from earlier in the proof. □

In the following lemma let $p, q, r \in P$ be such that each of them has non-empty upper and lower contour set. Then, by Lemma 4, there exist linear functionals v_p, v_q, v_r defined on the linear span of P that are positive exactly on the respective upper contour sets and negative exactly on the corresponding lower contour sets. Moreover, assume that none of these points belongs to the contour set of any other and that P spans the whole vector space.

Lemma 9. *If $p, q, r \in P$ are as above, then*

$$\frac{v_p(q)}{v_p(r)} \frac{v_q(r)}{v_q(p)} \frac{v_r(p)}{v_r(q)} = -1.$$

Proof. Notice that each of the functionals v_p, v_q, and v_r appears both in a numerator and in a denominator on the left-hand side of the equation to be proved. So, that expression is independent of the choice of these functionals. Actually, we may and will choose to define them as in the proof of Lemma 4 with the obvious choice of defining points. Assume at first that $p \succ q$, $q \succ r$, and $r \succ p$ and observe that $v_p(q)/v_p(r) = -\phi_p^+(r, q)$. By cyclically permuting p, q, and r, we get $v_q(r)/v_q(p) = -\phi_q^+(p, r)$ and $v_r(p)/v_r(q) = -\phi_r^+(q, p)$. This case of the lemma now follows easily from Lemma 5. Next, assume that $p \succ q$, $q \succ r$ and $p \succ r$. Observe that it now suffices to prove this case since all the other possible cases follow from this one and the above by cyclically permuting p, q, and r. To this end we first show the existence of a $t \in P$ such that $p, q, r \in L(t)$. The assumption $t \sim q$ for all $t \in U(p)$ leads to a contradiction since the non-trivial functional v_q cannot annihilate $U(p)$, which spans the whole vector space. So, for $t \in U(p)$ with $q \succ t$, there is exactly one point of the segment $[t, p]$ in $C(q)$ by Axiom 1. Hence, this segment contains also points from $U(q)$, so that we may suppose with no loss of generality that $p, q \in L(t)$. Since $U(p) \cap U(q)$ is non-empty, it spans the whole space and it cannot be annihilated by the non-trivial functional v_r. Thus, for any $t \in U(p) \cap U(q)$, the segment $[t, p]$ contains again points from $U(r)$. The existence of a $t \in P$ such that $p, q, r \in L(t)$ now follows and similar arguments show the existence of a $u \in P$ such that $p, q, r \in U(u)$. Using the appropriate definition of functionals v_p, v_q, and v_r, we obtain the following equations:

$$\frac{v_p(q)}{v_p(r)} = \frac{\phi_p^+(t, q)}{\phi_p^+(t, r)}, \quad \frac{v_q(r)}{v_q(p)} = -\phi_q^+(p, r), \quad \text{and} \quad \frac{v_r(p)}{v_r(q)} = \frac{\phi_r^+(q, u)}{\phi_r^+(p, u)}.$$

Multiply the three equations and apply Lemma 8 with p, q, t, r, s replaced respectively by t, p, q, r, u to get the desired conclusion. □

Denote by P^* the set of all $p \in P$ such that their upper and lower contour sets are both nonempty. Recall that the relation $\succ$ on P is called non-trivial whenever P^* spans the whole vector space.

Theorem. *Let P be a convex subset of a real vector space that avoids its origin. Then a non-trivial relation $\succ$ satisfying Axioms 1, 2, and 3 exists on P if and only if there is a skew-symmetric bilinear form Φ on the space such that $\Phi(p,q) > 0$ iff $p \succ q$, for $p, q \in P$.*

Proof. Assume at first that a form Φ of the kind claimed exists. To show that Axiom 1 is satisfied, choose any $p, q, r \in P$ such that $\Phi(p,q) > 0$ and $\Phi(q,r) > 0$. For any $\lambda \in [0,1]$, we have $\Phi(\lambda p + (1-\lambda)r, q) = \lambda\Phi(p,q) - (1-\lambda)\Phi(q,r)$ so that $\lambda = \dfrac{\Phi(q,r)}{\Phi(p,q)+\Phi(q,r)}$ is the unique element of $(0,1)$ for which $\lambda p + (1-\lambda)r \sim q$. Axiom 2 is clear, since for any given $p \in P$ the set of all $q \in P$ such that $p \sim q$ is exactly the intersection of the convex set P with the linear subspace $\{q \mid \Phi(p,q) = 0\}$ and is therefore convex. For Axiom 3, choose $p, q, r, s \in P$ as stated therein. By the above computation, we have $\phi_q(p,r) = \dfrac{\Phi(q,r)}{\Phi(p,q)+\Phi(q,r)}$ and similarly $\phi_r(q,s) = \dfrac{\Phi(r,s)}{\Phi(q,r)+\Phi(r,s)}$. Permute the points to get $\phi_q(p,s) = \dfrac{\Phi(q,s)}{\Phi(p,q)+\Phi(q,s)}$ and $\phi_r(p,s) = \dfrac{\Phi(r,s)}{\Phi(p,r)+\Phi(r,s)}$. Hence, $\phi_q(p,r) + \phi_r(q,s) = 1$ if and only if $\dfrac{\Phi(q,r)}{\Phi(p,q)+\Phi(q,r)} = \dfrac{\Phi(q,r)}{\Phi(q,r)+\Phi(r,s)}$, which is true if and only if $\Phi(p,q) = \Phi(r,s)$. A similar computation gives $\phi_q(p,s) + \phi_r(p,s) = 1$ if and only if

$$\frac{\Phi(q,s)}{\Phi(p,q)+\Phi(q,s)} = \frac{\Phi(p,r)}{\Phi(p,r)+\Phi(r,s)}. \tag{$*$}$$

The supposition that the segment $[q,r]$ is parallel to the segment $[p,s]$ tells us that any linear functional which is constant on one of the segments is also constant on the other. Since the functional $x \mapsto \Phi(x,r) - \Phi(x,q)$ is constant on the segment $[q,r]$ by the fact that Φ is skew-symmetyric, it is constant on $[p,s]$, showing that

$$\Phi(p,r) + \Phi(r,s) = \Phi(p,q) + \Phi(q,s). \tag{$**$}$$

Thus, Eq. $(*)$ reduces to $\Phi(q,s) = \Phi(p,r)$, so that $\phi_q(p,s) + \phi_r(p,s) = 1$ if and only if $\Phi(q,s) = \Phi(p,r)$. Using $(**)$ we see that this is true if and only if $\Phi(r,s) = \Phi(p,q)$ and together with the above we derive Axiom 3.

Let us now prove the Theorem in the opposite direction. For any $p \in P^*$ choose a linear functional v_p as in Lemma 4. Fix an $r \in P^*$ and for any $q \in P^*$ and any vector p define $\Phi(p,q) = \dfrac{v_q(p)}{v_q(r)} v_r(q)$. It follows that $\dfrac{\Phi(p,q)}{\Phi(q,p)} =$

$\frac{v_q(p)}{v_q(r)}\frac{v_r(q)}{v_r(p)}\frac{v_p(r)}{v_p(q)} = -1$ by Lemma 8 for all $p, q, r \in P^*$ such that none of these points belongs to the contour set of any other. Thus, $\Phi(p, q) = -\Phi(q, p)$ for any p, q of the kind. Since $p \mapsto \Phi(p, q)$ is linear by definition for every such q, so is $q \mapsto \Phi(p, q)$ for any such p by this equation. It follows that this equation extends to all vectors p, q. The fact that $p \succ q$ if and only if $\Phi(p, q) > 0$ now follows by the definition of Φ and by Lemma 4. □

We now give a simple but somewhat artificial example showing that Axioms 1, 2, and 3, are not equivalent to Axioms C, D, and S in general. Take for P the segment connecting the points $(0, 1)$ and $(1, 0)$ of the real plane. Let the relation on this segment be defined so that the point $(0, 1)$ is preferred to every other point of P, while indifference holds between any other two points. It is clear that P avoids the origin of the plane, that it spans the plane, and that the relation satisfies Axioms 1, 2, and 3. However, observe that Axiom D fails for this example. So, it is not surprising that there is no non-transitive utility function Φ that would yield this relation.

5 Applications

Assume that we have a finite set of outcomes or prizes $X = \{x_1, x_2, \ldots, x_n\}$. Let u_{ij} measure the relative utility of the preference of prize x_i over the prize x_j. Thus, in particular, $u_{ij} > 0$ means that prize x_i is preferred to prize x_j, while $u_{ij} < 0$ means the opposite. Actually, it makes sense to take $u_{ij} = -u_{ji}$. Now, assume two lotteries $p = \langle p_1, x_1; p_2, x_2; \ldots; p_n, x_n\rangle$ and $q = \langle q_1, x_1; q_2, x_2; \ldots; q_n, x_n\rangle$. The notation means that lottery p yields prize x_i with probability p_i for $i = 1, 2, \ldots, n$. What is the expected relative utility for the preference of lottery p over lottery q? One attractive possibility is

$$\sum_{i=1}^{n}\sum_{j=1}^{n} p_i u_{ij} q_j, \qquad (*)$$

and we consider it henceforth.

We write the relative utilities u_{ij} in matrix form as

$$A = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n1} & u_{n2} & \cdots & u_{nn} \end{bmatrix}$$

and think of lottery $p = \langle p_1, x_1; p_2, x_2; \ldots; p_n, x_n\rangle$ as the n-tuple $(p_1, p_2, \ldots, p_n)$. Denote by P the set of all n-tuples such that $p_i \geq 0$ for all $i = 1, 2, \ldots, n$, and $\sum_{i=1}^{n} p_i = 1$. Then the relative expected utility of Eq. $(*)$ can be written as $\Phi(p, q) = pAq^{\mathrm{tr}}$, where q^{tr} denotes the transpose of q.

Observe that a permutation of prizes yields a matrix permutationally congruent to the matrix A, i.e., a matrix $Q^{\mathrm{tr}}AQ$ for some permutation matrix Q.

If the sequence of prizes can be rearranged as $x_1 \succ x_2 \succ \cdots \succ x_n$, then the corresponding matrix A has positive entries above the main diagonal (and necessarily negative entries below it). This will clearly give rise to a transitive preference relation on P. On the other hand, if there is a cycle among prizes, say, $x_1 \succ x_2 \succ x_3 \succ x_1$, this will be seen from matrix A by the fact that entries u_{12} and u_{23} above the diagonal are positive and u_{31} below the diagonal is positive. It is clear that no matrix permutationally congruent to this matrix can have all the entries above the diagonal nonnegative and all the entries below the diagonal nonpositive.

This application extends to any set of prizes X equipped with a structure of a measurable space, i.e. an arbitrary set together with a σ-algebra of its subsets. The non-transitive utility function is represented in this case by a bounded, measurable, skew-symmetric function $\Phi : X \times X \to \mathbb{R}$, while the role of lotteries is played by random variables on X via their probability measures. The underlying vector space in this case is the Banach space of all real-valued finite measures. Function Φ extends to a skew-symmetric bilinear functional on this space via $\Phi(p, q) = \int\int \Phi(x, y)\, dp(x)\, dq(y)$ for any finite measures p and q. If the two measures represent "lotteries", i.e., they are probability measures, this formula reduces to the formula of mathematical expectation of the random variable $\Phi(U, V)$, where U and V are independent (generalized) random variables with values in the set X of prizes with respective distribution laws equal to p and q.

6 Applications to a social choice problem

In [All] a model is proposed for the dynamics of political power and control in a social system. The paper introduces various socio-political phenomena such as coalitions, alliances, anarchy, and revolutions. In [Om1] the model is studied from the point of view of hierarchical relations between the macro structures of the society. It is a surprising feature of the extended model that the equilibria always exist although the asymptotic behaviour of the system may get slower as the number of hierarchical levels grows. We give here briefly Allen's macro viewpoint. There is a finite number of *structures* in the society, i.e. its control holding, power wielding categories, which possess relative independence of the others. These categories exercize power through substructures. Allen's axioms on socio-political power and control read:

Axiom 1. Each of the structures attempts to exert power to control the others.
Axiom 2. Each of the structures accedes to the others a certain fraction of its complete control.
Axiom 3. Each structure acts to align the fraction of control that another structure has over a third according to an intrinsic control matrix, which may vary from structure to structure.

Let n be the number of identifiable structures, denoted by $S_i, i = 1, 2, \ldots, n$. For any indices i and j let p_{ij} denote the fraction of power (exercized over a particular fixed structure) that structure S_j accedes to structure S_i. The matrix $P = (p_{ij})$ of power fractions is called the *power profile matrix*. It follows from this definition that $\sum_{i=1}^{n} p_{ij} = 1$, for $j = 1, 2, \ldots, n$, so that this matrix is column-stochastic. For any structure S_i we further define x_i as its fraction of control. We have

$$\sum_{i=1}^{n} x_i = 1, \tag{$*$}$$

in order to preserve the total amount of control. Each structure has certain resources or powers at its disposal to achieve its goals. Denote by r_i the primary or direct power of structure S_i. Allen's model may be represented by

$$\dot{x}_i = \sum_{j=1}^{n} r_j (p_{ij} - x_i) x_j. \tag{$**$}$$

Introduce the resource matrix $R = \mathrm{diag}(r_1, r_2, \ldots, r_n)$, and the power fraction vector $x = (x_j)$. In addition, let $\mathbf{e}$ denote the n-tuple of 1's, and define the resource vector $r = R\mathbf{e}$. Denote by $\langle x, y \rangle$ the usual scalar product of the real n-tuples $x, y \in \mathbb{R}^n$, and let $x \otimes y$ denote the linear mapping of rank one defined for $z \in \mathbb{R}^n$ by $(x \otimes y)z = \langle y, z \rangle x$. Using this notation, model $(**)$ can be rewritten as

$$\dot{x} = (PR - x \otimes r)x, \tag{$***$}$$

and condition $(*)$ as $\langle x, \mathbf{e} \rangle = 1$. Thus, we want to find a solution $x(t)$ of the system of differential Eqs. $(***)$ satisfying a given initial condition $x(0)$. It turns out that condition $\langle x, \mathbf{e} \rangle = 1$ is satisfied by any solution of this system as soon as the initial condition $x(0)$ satisfies it.

In [Om2] this model was modified into a model with additive resources. It was supposed there that in addition to Allen's axioms the following axioms are satisfied.

Axiom 4. Every structure has its individual resources.

Axiom 5. The resources of any structure in a coalition equal the sum of the individual resources of its members, but it accedes all the gained power to the other members of the coalition.

Axiom 6. The goal of the structures is to increase maximal possible control in the long run.

Each of the two models may be viewed as a decision model. From this point of view the columns of the power profile matrix become strategies of the macro structures. Through these strategies they decide how much of their power they want to accede to other structures and to which ones. The outcome of their decisions is the eigenvector u which represents the fractions of control that the structures gain in the long run. It turns out that in these models the control is given to some coalitions of structures called *coalitions in power*. In both models, the aim of the structures can be considered to maximize their individual fractions of control in the long run. However, it turns out that there is a substantial difference between the two models in how to achieve this goal. Namely, the original model as presented in [All] and [Om1] behaves antagonistically since a structure can only gain some power at the expense of some other structure, while in the model with additive resources, as presented in [Om2], there is an advantage in forming a coalition since the individual powers of structures add up to the power of a coalition.

The question of what kind of coalitions the structures will choose to form reduces to a social choice problem. In [Om2] some suggestions were given to treat the question, including the one using utility functions. Denote by $w_i(\mathscr{C})$ the utility of the structure S_i to enter the coalition $\mathscr{C}$. Then, a posssible way to get a social welfare function, or at least a social choice correspondence in the sense of [Mou], is to introduce a "social choice utility" such as $w(\mathscr{C}) = \sum_{i=1}^n w_i(\mathscr{C})$. This way we are taking care of the behaviour of the macro structures only as long as it is rationally explainable via their resources and their greed for power and control. We now propose to add to this a term that would take care of their irrationality as well. Let the relative "irrational utility" $\omega_i(\mathscr{C}, \mathscr{D})$ measure the amount of irrational preference that the structure S_i has for coalition $\mathscr{C}$ over coalition $\mathscr{D}$ and assume that this function is skew-symmetric on the set of all possible coalitions in power. The irrational part of the social choice utility is then naturally defined as $\omega(\mathscr{C}, \mathscr{D}) = \sum_{i=1}^n \omega_i(\mathscr{C}, \mathscr{D})$, while the total relative utility is $\Phi(\mathscr{C}, \mathscr{D}) = w(\mathscr{C}) - w(\mathscr{D}) + \omega(\mathscr{C}, \mathscr{D})$.

In [Om2] the social choice utility was computed for a possible utility function. Let r_i denote the resources of the structure S_i, for $i = 1, 2, \dots, n$, and assume that $\sum_{i=1}^n r_i = 1$. It turns out that the fraction of control $u_i(\mathscr{C})$ that the structure S_i gains in the long run whenever it decides for a coalition $\mathscr{C}$ equals

$$u_i(\mathscr{C}) = \begin{cases} \dfrac{r_i}{\rho(C)} & \text{if } S_i \in \mathscr{C} \text{ and } \rho(C) > \rho(D) \\ 0 & \text{otherwise} \end{cases}$$

where

$$\rho(C) = \sum_{j \in I} r_j \quad \text{and similarly} \quad \rho(D) = \sum_{j \notin I} r_j.$$

Here and below, I is the set of indices of structures in the coalition $\mathscr{C}$. The condition $\rho(\mathscr{C}) > \rho(\mathscr{D})$ means that the coalition $\mathscr{C}$ is in power. If we decide for the utility function $w_i(\mathscr{C}) = u_i(\mathscr{C})^2$, we get

$$w(\mathscr{C}) = \begin{cases} \dfrac{\rho_2(C)}{\rho(C)^2} & \text{if } \rho(C) > \rho(D) \\ 0 & \text{otherwise} \end{cases}$$

where

$$\rho(C) = \sum_{j \in I} r_j, \quad \rho_2(C) = \sum_{j \in I} r_j^2, \quad \text{and} \quad \rho(D) = \sum_{j \notin I} r_j.$$

As an illustration consider three structures with respective resources r, s, and t, such that $r \geq s \geq t$ and $s + t > r$. The possible coalitions in power are $\mathscr{A} = \{S_1, S_2\}$, $\mathscr{B} = \{S_1, S_3\}$, $\mathscr{C} = \{S_2, S_3\}$, and the "big" coalition $\mathscr{D} = \{S_1, S_2, S_3\}$. The social choice utilities of these coalitions are clearly

$$w(\mathscr{A}) = \frac{r^2 + s^2}{(r+s)^2}, \quad w(\mathscr{B}) = \frac{r^2 + t^2}{(r+t)^2}, \quad w(\mathscr{C}) = \frac{r^2 + s^2}{(r+s)^2},$$

$$\text{and} \quad w(\mathscr{D}) = \frac{r^2 + s^2 + t^2}{(r+s+t)^2}.$$

We now introduce irrationality into this example. Assume that structure S_1 hates cooperating with S_2 and strongly prefers to join its forces with S_3. This implies existence of a relative utility $\omega_1(\mathscr{B}, \mathscr{A}) = \alpha$, where α is a fixed positive value measuring the level of this preference. To get this utility function skew-symmetric, we have to define $\omega_1(\mathscr{A}, \mathscr{B}) = -\alpha$. Let all other values of ω_1 be zero. Similarly, assume that structure S_2 prefers to cooperate with S_1 more than with S_3, and that structure S_3 prefers to unite its resources with S_2 rather than with S_1. This implies $\omega_2(\mathscr{A}, \mathscr{C}) = \beta$, $\omega_2(\mathscr{C}, \mathscr{A}) = -\beta$, $\omega_3(\mathscr{C}, \mathscr{B}) = \gamma$, and $\omega_3(\mathscr{B}, \mathscr{C}) = -\gamma$, with all other values of ω_2 and ω_3 equal to zero. Here, β and γ are fixed positive values measuring values of the level of respective preferences. Using the above we get the matrix of relative social choice utilities

$$\begin{bmatrix} 0 & w(\mathscr{B})-w(\mathscr{A})+\alpha & w(\mathscr{C})-w(\mathscr{A})-\beta & w(\mathscr{D})-w(\mathscr{A}) \\ w(\mathscr{A})-w(\mathscr{B})-\alpha & 0 & w(\mathscr{C})-w(\mathscr{B})+\gamma & w(\mathscr{D})-w(\mathscr{B}) \\ w(\mathscr{A})-w(\mathscr{C})+\beta & w(\mathscr{B})-w(\mathscr{C})-\gamma & 0 & w(\mathscr{D})-w(\mathscr{C}) \\ w(\mathscr{A})-w(\mathscr{D}) & w(\mathscr{B})-w(\mathscr{D}) & w(\mathscr{C})-w(\mathscr{D}) & 0 \end{bmatrix}.$$

Now, if $r = s = t = \frac{1}{3}$, then the rational part of the utilities become $w(\mathscr{A}) = w(\mathscr{B}) = w(\mathscr{C}) = \frac{1}{2}$ and $w(\mathscr{D}) = \frac{1}{3}$. Thus the above matrix reduces to

$$\begin{bmatrix} 0 & \alpha & -\gamma & -\frac{1}{6} \\ -\alpha & 0 & \beta & -\frac{1}{6} \\ \gamma & -\beta & 0 & -\frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \end{bmatrix}.$$

In this case, the structures have equal power and rational reasons give no clear outcome since all the coalitions of two structures are to occur equally likely and the "big" coalition will happen less often. So, the outcome depends solely on the irrational causes. Since structure S_1 likes to cooperate with S_3, they may form the coalition $\mathscr{B}$. However, this creates dissatisfaction on the side of the structure S_3 as they would prefer to be in a coalition with S_2 rather than with S_1. They may therefore break the coalition $\mathscr{B}$ and form the coalition $\mathscr{C}$. However, now structure S_2 is not satisfied, so it breaks $\mathscr{C}$ and forms $\mathscr{A}$. Clearly, this behaviour may create a cycle of coalitions exchanging their positions in power.

We have thus seen that irrationality can be introduced in the model of dynamical interactions of socio-political structures. The above example shows that when irrational causes are not adequately compensated by rational reasons, they can prevail and create a high level of instability. This aspect of the model may be useful in some applications. It is an interesting question whether this kind of irrationality can be introduced into the dynamical model directly to obtain the cyclic behaviour of the model in terms of time *t*.

References

[All] Allen GD (1992) Toward a dynamics for power and control in society. J Math Sociol 17: 1–38

[Arr] Arrow KJ (1963) Social choice and individual values, 2nd Ed. Wiley, London

[Bri] Bridges DS (1994) The constructive theory of preference relations on a locally compact space II. Math Social Sci 27: 1–9

[Dom] Dombi J (1994) Universal characterization of non-transitive preferences. Math Social Sci 27: 91–104

[Fi1] Fishburn PC (1970) Utility theory for decision making. Wiley, New York

[Fi2] Fishburn PC (1982) Nontransitive measurable utility. J Math Psychol 26: 31–67

[Fi3] Fishburn PC (1984) SSB utility theory and decision-making under uncertainty. Math Social Sci 8: 253–285

[Fre] French S (1993) Decision theory: An introduction to the mathematics of rationality, 2nd Ed. Ellis Horwood, Chichester

[FuI] Fukuba Y, Ito K (1984) The so-called expected utility theory is inadequate. Math Social Sci 7: 1–12

[Mou] Moulin H (1983) The strategy of social choice. North-Holland, Amsterdam

[Om1] Omladič M, Omladič V (1994) Hierarchical dynamics for power and control in society. J Math Sociol 18: 315–332

[Om2] Omladič V (1995) Asymptotics of dynamics for control in society. In: Ferligoj A, Kramberger A (eds) Contributions to methodology and statistics. Faculty of Social Sciences, Ljubljana, pp 165–180

[Rap] Rapaport A (1994) Problems of normative and descriptive decision theories. Math Social Sci 27: 31–47

Soc Choice Welfare (2001) 18: 269–288

Social Choice and Welfare

A cognitive model of individual well-being

Itzhak Gilboa[1], **David Schmeidler**[2]

[1] Department of Economics and Recanati School of Business, Tel Aviv University, Tel Aviv 69978, Israel (e-mail: igilboa@post.tav.ac.il)
[2] Department of Statistics and Recanati School of Business, Tel Aviv University, Tel Aviv, Israel, and
Department of Economics, Ohio State University, Ohio, OH, USA
(e-mail: schmeid@tav.ac.il)

Received: 3 April 1998/Accepted: 23 March 1999

Abstract. This paper offers a model of individual well-being that takes into account cognitive factors. It postulates that individuals compare payoffs to aspiration levels. The latter are determined by past experiences (adaptation), by other people's performance (interpersonal comparison), as well as by reasoning (justifications and excuses). We axiomatize a measure of well-being defined on real-valued vectors of various lengths. It is a linear combination of differences between payoffs and aspiration levels, where the aspiration level at each instance is a linear function of past payoffs.

1 Introduction

Modern microeconomic theory is behavioral in nature. For the most part, it relies on the neo-classical utility function, which is in principle derived from observed preferences, for both descriptive and normative applications. Further, this utility function is often defined on product bundles, disregarding psychological factors that individuals seem to relate to in describing their well-being.

For descriptive purposes, this approach is consistent and it may be tested for empirical validity. By contrast, it is doubtful that a behaviorally-defined utility function is sufficient when normative considerations are involved. Welfare economics ultimately deals with cognitive concepts such as "well-

We wish to thank Elchanan Ben-Porath, Eva Gilboa-Schechtman, Daniel Kahneman, Edi Karni, Dan Levin, Donald O. Parsons, James Peck, Andrew Postlewaite, Drazen Prelec, Peter Wakker, an anonymous referee, and an associate editor for discussions that influenced this work, as well as for comments and references.

being," "happiness," and "satisfaction." These relate to notions such as aspirations and needs, contentment and disappointment. The literature does not seem to offer a convincing justification for substituting revealed preference for these concepts. Moreover, the notion of "well-being" involves too many cognitive variables to allow us to infer their values from choice data. For instance, if we accept the view that one's satisfaction is determined by one's consumption relative to one's needs, we are faced with two subjective, cognitive variables, namely, "needs" and "satisfaction," that cannot be disentangled based on objective consumption data alone.

The goal of this paper is to suggest a model of individual well-being that explicitly introduces some of the relevant cognitive variables. Specifically, the key assumption is that an individual's well-being depends on her aspirations, where these are endogenously determined. One may identify three major factors that participate in the determination of an individual's aspiration level for a given experience at a given time. First, people adapt to circumstances. Hence the individual's own history of payoffs affects her aspirations. For instance, when an individual is accustomed to a certain standard of living, her well-being depends mostly on deviations from it. Second, people compare themselves to others in what they perceive to be their peer group. Thus, other people's payoffs are also determinants of an individual's aspirations. An increase in one's income may make one worse off if it is accompanied by a decrease in one's income ranking in society. Finally, various facts that do not directly relate to any single agent's performance may give an individual reasons to expect higher or lower payoffs. For instance, if the economy is predicted to boom, individuals may be led to expect higher income. The fact that one grows old may decrease one's aspirations regarding one's health. Aging is a fact that directly (and negatively) affects one's well-being. But it also helps one to accept, say, a decline in one's physical fitness. Similarly, being discriminated against has a direct negative impact on an individual's well-being. But it can also account for perceived failures. By this process of "psychological accounting" an excuse can adjust the individual's aspiration level, and thereby mitigate the impact of low payoffs.

These considerations call into question some of the accepted tenets of welfare economics. For instance, it is not obvious that more choices or more opportunities necessarily make an individual better off. Choice comes with responsibility, and it may result in regret. Opportunity generates expectations, and may lead to low self-esteem. More generally, welfare analysis should take into account subjective aspirations as well as objective performance. While having more choices and opportunities may improve the latter, it also tends to raise the former. Its net effect therefore need not always be positive.

For a concrete example, consider an integrative educational system, in which children of different neighborhoods and of different socio-economic status (SES) are put in the same school class. Such a system presumably allows children of lower-income families to have the same level of education as those of higher-income families, thereby giving them an equal opportunity to succeed in their future careers. Undoubtedly, some children may benefit from the system in terms of their objective performance. However, many may experience a reduction in their subjective well-being for two reasons. First, the mere exposure to lifestyles of higher-income families may make lower SES children view their life differently. Their neighborhood, house, and consumption opportunities are likely to be seen as less satisfactory than prior to their exposure to those of high SES children. Second, the semblance of "equal opportunity" deprives them of potential justifications for low performance on objective scales. Should a child end up with a low-paying job, believing that she indeed had "equal opportunity" leaves her with no one but herself to blame for her "failure." Overall, it is not clear that such a system does more good than harm. It does seem clear, however, that economic theory lacks the language to address this question.

A clarification is due before it becomes too late. We do not intend to suggest that people are better off in slavery than in liberty, that competition should be abolished, or that all students should always get "A" grades. Our goal is merely to offer a theoretical framework within which some of the above phenomena can be discussed.

To this end, we suggest the following model. An individual's history consists of a sequence of facts. Some are events that have a temporal aspect. Some are atemporal, such as religious or national identification. Each fact t has an "objective" payoff attached to it, denoted x_t. The individual's instantaneous payoff is defined as $x_t - a_t$, where a_t is the individual's aspiration level pertaining to that fact. Overall well-being is defined as a weighted sum of these instantaneous payoffs. However, the aspiration levels a_t are themselves linear functions of preceding objective payoffs. Explicitly, given a history $x = (x_1, \dots, x_T)$, the measure of well-being is

$$U(x) = \sum_t w_t(x_t - a_t)$$

for given weights w_t, where

$$a_t = \sum_{i=1}^{t-1} s_{it} x_i$$

for given coefficients s_{it}.

Suppose that x is the individual's income stream. An increase in income at period i has a direct effect on the individual's well-being. The larger is w_i (assumed positive), the happier will be the individual. However, with positive coefficients s_{it} for $t > i$, the aspiration levels in future periods will be higher. If the individual will not experience high payoffs in these periods, she may be disappointed. Moreover, if the aspiration level is an average of recent periods' payoffs, a permanent increase in income will have an effect mostly in the first periods, whereas in later ones the aspiration level will be correspondingly adjusted and the instantaneous payoff will diminish.

Assume now that x describes experiences of consumption of various goods, including meals, entertainment, social life, and so forth. Within each category, similar logic applies. Many of the coefficients s_{it} may be null, reflecting the fact that the individual's aspiration level for a certain experience depends only on experiences of the same kind. Yet, the measure of overall well-being is an aggregation of all such experiences.

Note that, in the present formulation, current payoffs only affect future aspiration levels. The individual does not re-assess past experiences in view of current payoffs or aspirations. Thus, an individual who has a constant high income will still be better off than one who has a constant low income, even though both have become adjusted to their respective circumstances. One may wish, however, to consider a more general model, according to which aspirations are adjusted retrospectively as well.

To capture the impact of interpersonal comparison, we incorporate into the model facts involving protagonists other than the individual in question. For instance, assume that we are measuring the well-being of an individual A, who knows that another individual, B, has a certain annual income. Suppose that fact i is that B has income x_i. The weight w_i assigned to this payoff in the measurement of A's well-being reflects A's attitude toward B. Should it be positive, A would rejoice at B's successes and lament her failures. Naturally, a negative w_i would reflect a less benevolent attitude. Finally, if w_i is zero, B's income has no direct impact on A's well-being.

However, even if w_i is indeed zero, or if one maintains that, for normative purposes, w_i should be assumed zero (see Harsanyi (1992)), B's income may affect A's aspiration level, and thus his well-being, via the coefficients s_{it}. A positive value of s_{it} would make A aspire to have a higher income, the higher is B's income. Further, this social aspect may interact with the process of adaptation discussed above. Assume that every two weeks both A and B get their pay checks, and that A knows B's previous payoffs. We obtain a sequence of facts, in which the odd-numbered ones are, say, A's salary, whereas the even-numbered ones reflect B's salary. Assume that A's salary is lower than B's, and that the aspiration level is a weighted average of past period's payoffs. If left alone, A would have adjusted to his income, and his U measure would tend to zero. However, in the presence of the constant reminder of B's higher income, A's U measure remains negative.

Observe that our interpretation of "aspiration level" is emotional rather than rational. That is, the aspiration level does not attempt to capture the individual's reply to "What do you think you will get?" but rather "What would make you content?" In the example above, A cannot fail to notice that, pay period after pay period, B gets a higher salary. If asked, A would certainly be able to correctly predict his next paycheck. However, to the extent that A has not come to terms with this salary difference, to the extent that he is bothered by it, feels that he is discriminated against, and so forth – his aspiration level is affected by B's income.

Next consider examples of psychological accounting. An old woman finds that her mental capacities are becoming limited. She faces problems remembering names, numbers, tasks, and so forth. Of course, this is not good news. However, she might not be as unhappy about it as she would have been had these phenomena occurred when she was young. In our model, "growing old" would be one of the facts in the woman's memory. This fact is associated with a low payoff and thus reduces well-being. But, to the extent that the woman comes to terms with aging, this low payoff also reduces her aspiration levels regarding her memory's performance. This would make future failures less painful, and mediocre successes – a source of joy.

Psychological accounting also applies to one's self-esteem. Consider a prospective student who belongs to a minority group, say, defined by religion. He is not admitted to a top school, and he suspects that he might have been discriminated against. Being denied admission is not good news. As in the previous example, discrimination also means that the student suffers from unfavorable interpersonal comparison. Every equally qualified candidate who was admitted makes the rejection more painful. Moreover, the injustice of discrimination is infuriating in and of itself. However, *given* that one is not admitted, there may be some consolation in the fact that this failure is not a reflection of one's true merit. As far as responsibility and self-esteem are concerned, the student may "deduct" his religion from his objectively poor performance. He is only responsible for the unaccounted part of this low payoff.

The phenomena we attempt to capture in our model have been discussed in the past. Indeed, some of these discussions date back to religious thinkers of previous millennia, and some have pervaded popular culture to the point of banality. To the extent that this paper makes a contribution, it is in offering a simple, formal model that allows a discussion of well-being and that captures the various determinants of aspirations in a unified way.

The rest of this paper is organized as follows. Section 2 presents the model and an axiomatization of the evaluation rule. The main goal of the axiomatic treatment is to show that theoretical concepts such as "aspiration level" can be derived from a more primitive and presumably observable preference relation. Section 2 also discusses monotonicity. Section 3 discusses the relationship of our model to existing theories, and the extent to which it can describe some empirical psychological findings. We devote Sect. 4 to a discussion of welfare economics in light of our model. Section 5 concludes.

2 Model and result

A "fact" is a pair, of circumstances and payoff. To simplify notation, we assume that the circumstances are ordered, and identify them with the natural numbers. It should be borne in mind that they are not necessarily repetitions of identical situations. For instance, "circumstance" 3 may be belonging to a minority group, "circumstance" 5 may be having a meal, and so forth.

The objects of comparison are histories, represented by real-valued vectors, with generic notation $x^T = (x_1^T, \ldots, x_T^T)$. Since histories of different lengths are involved, we will use superscripts to denote their lengths. Thus, a symbol such as x^T denotes a real-valued vector of length $T \geq 1$, with generic component x_i^T. We will use this notation also for real-valued vectors that are not interpreted as histories (for instance, vectors of coefficients). Finally, $(x^T, y^{T'})$ denotes vector concatenation.

When considering different values of x_i^T, we refer to different possible facts, all of which share the circumstances, and vary only in the objective payoff. For instance, x_3^T may take a low value if the individual in the example above was discriminated against because he belonged to a minority group. It would take a high value if the individual still belonged to the same group, but benefited from this identification. In the sequel we assume that any real-valued payoff can be attached to any circumstance, generating a meaningful fact.

Implicit in this notation is the assumption that all possible histories agree on the order of circumstances. That is, for any two histories, the vector of circumstances of the shorter history is a prefix of the corresponding vector for the longer one. The model can be extended to deal with functions on arbitrary sets of circumstances. Yet, the present set-up contains sufficient information to derive the functional form we are interested in, and its notation is simpler. Further, no loss of generality is involved in assuming that facts are linearly ordered, as opposed to partially ordered.

Thus the set of objects of comparison is $X \equiv \bigcup_{T=1}^{\infty} \Re^T$, and $\geq$ will denote a binary order on it. $x^T \geq y^{T'}$ is interpreted to mean that a history x^T makes an individual at least as happy as a history $y^{T'}$. Our primary interpretation of $\geq$ is cognitive: it should reflect an individual's responses, based on introspection, to questions of the type "what life history would make you happier?" It is also possible to interpret the preference relation behaviorally, provided that choice data are given not only between outcomes but also between entire histories. For instance, major life decisions can be viewed as choices between life paths, for oneself or for one's children. But it is hard to imagine that the preference between any two life histories may be revealed by actual choices. Be that as it may, we emphasize that the preference between life histories can not be subsumed by revealed preferences between single outcomes.

One may wonder, how can histories of different lengths be compared? For instance, if $T < T'$, x^T may appear "better" than $y^{T'}$ simply because we are not told what happens in history x after T. Shouldn't we first find out what x entails? The answer is no. If the interpretation of $\geq$ is cognitive, there is no difficulty in comparing well-being or happiness at different times. Indeed, many people feel that they were happier in their youth than in old age. But even if $\geq$ is behaviorally interpreted, "facts" in our model need not be temporally defined, nor are they required to be complete descriptions of one's life. Rather, they are the subjectively relevant facts.

We are interested in the following axioms on $\geq$ ($\leq$, $>$, $<$, and $\approx$ are defined as usual):

A1: $\geq$ is a weak order.
A2 (Continuity): For every $T, T' \geq 1$, and every x^T, the sets $\{y^{T'} | x^T > y^{T'}\}$ and $\{y^{T'} | x^T < y^{T'}\}$ are open (in the standard topology) in $\Re^{T'}$.
A3 (Additivity): For every $T, T' \geq 1$, and every x^T, $y^{T'}$, z^T, $w^{T'}$, if $x^T \geq y^{T'}$

$(x^T > y^{T'})$ and $z^T \geq w^{T'}$, then $x^T + z^T \geq y^{T'} + w^{T'}$ $(x^T + z^T > y^{T'} + w^{T'})$.
A4 (Neutral Continuation): For every $T \geq 1$ and every x^T there exists $a \in \Re$ such that $x^T \approx (x^T, a)$.

A1 and A2 are standard. A3 is a straightforward additivity assumption. It makes sense mostly when the payoffs are interpreted as if they were measured in "utiles." Finally, A4 guarantees that for every history there is a continuation that does not affect its desirability. Intuitively, the value a reflects the individual aspiration level following the experience x^T. Should the individual now experience the payoff a, she would be just as content as before. In the presence of A2, one can derive A4 from the assumption that any history can be continued in a way that would improve it, as well as in a way that would impair it.

Our main result is

Theorem. *The following are equivalent:*
(i) $\geq$ *satisfies A1–A4;*
(ii) *There are real numbers* $(w_t)_{t\geq 1}$ *and* $(s_{it})_{t>i\geq 1}$ *such that for every* $T, T' \geq 1$, *and every* x^T, $y^{T'}$,

$$x^T \geq y^{T'} \quad \text{iff } U(x^T) \geq U(y^{T'})$$

where

$$U(x^T) = \sum_{t=1}^{T} w_t(x_t^T - a_t(x^T))$$

and

$$a_t(x^T) = \sum_{i=1}^{t-1} s_{it} x_i^T.$$

Furthermore, in this case the weights $(w_t)_{t\geq 1}$ *are unique up to a multiplication by a positive number, whereas, for every* $t > 1$, $(s_{it})_{t>i\geq 1}$ *are unique whenever* $w_t \neq 0$ *(and arbitrary otherwise).*

The proof is relegated to an Appendix.

Under A1–A4, U need not be monotone with respect to the instantaneous payoffs. Indeed, since some of the payoffs may reflect other agents' experiences, there is no need to assume that U is always monotone. However, if we restrict attention to facts that relate to the individual under consideration, lack of monotonicity may be theoretically troublesome. For instance, it allows that an individual be better off not to get a one-time increase in income, due to future disappointments. Similarly, an individual might be better off when discriminated against, or at least when believing he is. While such rankings may be observed, most people would not consider it ethical to rely on them in decision making regarding other people's welfare. For example, it is not very convincing to argue that one should not give charity to homeless people, in order not to expose them to future disappointments, or that all cases of discrimination are made up by weaklings who cannot confront their own inadequacies. We are therefore interested in the following:

A5 (Monotonicity): For every $T \geq 1$ and every x^T, y^T, if $x_t^T \geq y_t^T$ for all $t \leq T$, then $x^T \geq y^T$.
Observation: Under the assumptions of the Theorem, $\geq$ satisfies A5 iff

$$w_i \geq \sum_{t=i+1}^{T} s_{it} w_t \quad \text{for all } i \leq T \tag{$*$}$$

(in particular, $w_T \geq 0$ for all T).

Condition $(*)$ states that the direct (positive) impact a payoff x_i has outweighs the potential indirect (negative) impact it has on future aspiration levels. In view of the uniqueness result in the Theorem, this condition does not depend on the specific numbers chosen to represent $\geq$.

Monotonicity does not imply that the coefficients s_{it} are nonnegative. On the contrary, the lower are the s_{it}, the larger is the cone of $(w_t)_{t \geq 1}$ that define a monotone relation. A negative value of s_{it} may result from several reasons. First, payoff in circumstance i may be perceived to be negatively correlated with performance in circumstance t. For instance, one may be happy that one is honest, but consequently one may not expect to be considered very polite. Second, an individual may experience satiation. In this case, high payoffs in some early circumstances may reduce her aspiration level in later ones. (This may also result in changing the coefficients w_T. Such dependencies are beyond the scope of the linear model.)

3 Psychological evidence and related theories

The term "aspiration level" is borrowed from Simon (1957), who argued that people "satisfice," rather than optimize in decision making. A payoff that exceeds the aspiration level makes the decision maker "satisficed," and thereby prone to retain the status quo, whereas a payoff that falls below it prods her to experiment. Gilboa and Schmeidler (1995, 2001) suggest a model of decision under uncertainty that provides a behavioral definition of "aspiration levels." However, we do not focus on the decision-making aspect in this paper, and the present interpretation of "aspirations" differs from the behavioral one. A rational decision maker may behave as if she were satisficed because she knows that she has no better option than her current choice. Yet, if she is unhappy, she is below the "emotional" aspiration level we use in this paper.

Adaptation level theory, developed by Helson (1964), suggests that people adjust to various stimuli. The theory deals with phenomena ranging from perception to happiness and to learning and creative thinking. It postulates that the adaptation level is a geometric average of a "background" level and a series of stimuli. (These terms are borrowed from the case of visual perception.) Since we focus on aspiration level adjustments, the formal concept of "aspiration" in our model bears kinship to adaptation level. Specifically, on a logarithmic scale, Helson's formula for the adaptation level becomes very similar to the linear one we use for the aspiration level.

Adaptation level theory focuses on the measurement of an instantaneous sensation, rather than on the concept of overall well-being. It thus corresponds to expressions of the type $x_t - a_t$ in our model, and does not deal with aggregation thereof. In contrast, our overall measure of well-being, U, is cumulative in nature. It attempts to measure how happy an individual is throughout a sequence of experiences, or how desirable such a sequence is. Correspondingly, two individuals who are fully adjusted to their respective income levels might be equally happy according to adaptation level theory, while the richer individual would still have a higher U value in our model. Still, the assessment of instantaneous payoffs relative to an adapting aspiration level finds support in the psychological literature on adaptation level theory. (See Appley 1971)

Brickman et al. (1978) studied lottery winners and their self-reported happiness. They found that recent winners were not happier than were control subjects. They explain this phenomenon using adaptation level theory, by the concepts of contrast and of habituation: shortly after the lottery, in contrast to the moment in which the winners were informed of their gains, they were no longer as happy. Similarly, it is postulated that they became habituated to their new wealth level. In our model, habituation is captured by the adjustment of the aspiration level. In order to satisfactorily model contrast, however, one may wish to extend the model by allowing the s_{it} to depend on the instantaneous payoffs: a single period in which the payoff was exceptional may, because of this payoff, play a more major role in the determination of future aspiration levels. (Other studies relating to aspiration level adjustments include Payne et al. 1980, 1981; March 1988, and Mezias 1988.)

There is ample evidence that subjective well-being can hardly be approximated by real income. Campbell (1981) argues that "people react to the world as they perceive it, not as it objectively is" (p. 23), and studies self-reported happiness across various income and education levels in the US over time. For instance, from 1957 to the early 1970's, the proportion of American population who were prepared to describe themselves as "very happy" declined from 35% to 24%; by 1978, it had risen to 30%. These shifts "are in direct opposition to the national economic trends" (p. 28). Perhaps more significantly, these proportions are at a very low correlation with income and with education levels. (See also Schoemaker 1982, Diener 1984, Duncan 1975, and Easterlin 1974) This evidence is consistent with our model: a high-income individual is likely to be better off than a low-income one, but, since the aspiration levels of both adjust over time, measures of their long-run well-being need not be drastically different.

Kahneman and Varey (1991) discuss adaptation level theories, and suggest that utility may be derived from transitions no less than, and perhaps more than from states (see also Kahneman 1994). Relatedly, Fredrickson and Kahneman (1993) find that the way people retrospectively evaluate affective episodes (viewing film clips in their studies) is barely related to their duration.

They dub this phenomenon "duration neglect." Moreover, Kahneman et al. (1993) have found that, due to duration neglect, people may choose more pain over less. They exposed subjects to a painful experience (60 seconds of holding one's hand in water at 14 °c), and then to the same experience followed by a better, but still painful, end (additional 30 seconds of holding the hand in water that gradually warmed up to 15 °c). A significant majority of the subjects preferred to repeat the longer experience.

Duration neglect is captured by our model if the aspiration level is adjusted fast enough. Fredrickson and Kahneman (1993) also report that individuals' retrospective evaluations of affective episodes significantly depend on the peak and the end levels of affect. End effects are compatible with our model.[1] However, modeling peak effects (as well as modeling the concept of "contrast" mentioned above) calls for a generalization according to which the weights w_t and s_{it} may depend on the instantaneous payoffs. It should be noted that Kahneman and his colleagues discuss the effects of memory, as well as one's ability to predict one's taste. It is not clear that the way experiences are remembered is the yardstick by which well-being should be measured. Yet, the evidence provided by these studies supports the principle on which our model is based.

Kahneman et al. (1997) suggest a normative model for the evaluation of experiences. Their normative suggestion is to avoid duration neglect. By contrast, our model is more descriptive in nature. As discussed below, we tend to be more agnostic on normative questions. When histories consist solely of past consumption, our model bears resemblance to other models such as Ryder and Heal (1973), Abel (1990), and Becker's (1996) model of addiction, in that preferences are history-dependent.

Interpersonal comparisons of income, status, and utility have probably received economists' attention more than other determinants of an individual's aspiration level. Veblen (1899) has already discussed conspicuous consumption (see also Leibenstein 1948). Duesenberry (1949) has formulated the relative income hypothesis, stating that saving rates depend on a family's percentile position in the income distribution. Hirsch (1976) further emphasized the role of relative social status. Hayakawa and Venieris (1977) discuss the consumer's "reference group" in this context. Interpersonal comparisons are also at the basis of the concept of "envy-free" allocations introduced by Foley (1967) (see also Schmeidler and Vind 1972, Pazner and Schmeidler 1974, and Varian 1974). Frank (1985a, 1985b, 1989) argues that an individual's consumption is compared to that of others, as well as to past consumption, thus combining adaptation with interpersonal comparisons. Kapteyn et al. (1980) and Kapteyn and Wansbeek (1982, 1985) argue that utility is completely relative. Ng and Wang (1993) present a model that captures aspiration level and interpersonal comparisons effects, as well as envi-

[1] This may be more clearly seen from an equivalent mathematical representation of the evaluation functional. See Appendix.

ronmental factors. (Other studies include Rainwater 1974, Layard 1980, Tomes 1986, Seidman 1987, Congleton 1989, and Persky and Tam 1990.)

In the context of existing literature, the present paper suggests a formal, axiomatically-based model that captures the effects of past consumption and of interpersonal comparisons, as well as other factors affecting aspirations, in a unified way.

Another phenomenon our model attempts to capture is that one's well-being depends on reasons, justification, and excuses one may have for objectively poor performance. When pushed to an extreme, this cognitive phenomenon may be reflected in behaviorally observed preference for situations in which one is constrained, or "objectively" disadvantaged. Such preferences are called "self-handicapping" in the psychological literature. For instance, Berglas and Jones (1978) have conducted an experiment in which subjects were administered a drug prior to engaging in a problem-solving task. The subjects were asked to choose between two drugs. One was described as enhancing performance, whereas the second was allegedly interfering with performance. Some subjects (mostly male) chose the drug that was alleged to induce poorer performance. According to Berglas and Jones, the only plausible explanation for this choice is to provide oneself with an excuse for failure to solve the problem. (See also Tice and Baumeister 1990). In our model, such a choice is explained by a higher U value in the presence of a fact that may serve as an excuse. (Note, however, that this type of preference violates the monotonicity axiom.)

Berglas and Jones also allowed subjects to choose the level of difficulty of the task. They found that "subjects high in fear of failure show a tendency to prefer either very simple or very difficult tasks, tasks typically low in diagnosticity" (pp. 405–406). McClelland (1961) has argued that not everyone wants to have "concrete knowledge of results of choices and actions" (p. 231). (See also Sedikides 1993 and Trope 1979 regarding preference for diagnosticity.)

The tendency to choose a task that would not provide much information regarding one's abilities may be explained by our model: failing on a task that is known to be difficult is less painful than failing on one that is supposed to be manageable. The fact that the task is difficult reduces one's aspiration level regarding one's performance on it. Thus, in case of failure, the resulting U value would be higher in case of a difficult task than it would be in case of a simpler one.

4 Welfare economics

Measuring well-being by a "utility" index introduces the temptation to adopt various axioms that are well accepted in the context of the neo-classical utility function, or of income. For instance, Pigou (1920) has introduced an axiom of social welfare, according to which a transfer of income from a high-income individual to a low-income one, that does not reverse their income ranking, enhances equality. Applying this axiom to our model verbatim may result in the suggestion that a satisficed individual be taxed in order to subsidize an unsatisficed one.

Such a suggestion is patently absurd as a general principle. The satisficed individual may be much poorer than the unsatisficed one. In this case, the transfer, which is Pigouvian on a subjective scale, is clearly non-Pigouvian on an objective one. It hardly seems just to tax someone simply because they happen to have a low aspiration level. Further, this policy is easily manipulable, since it provides individuals with an incentive to overstate their aspirations.

However, in some cultures it is not uncommon to observe transfers from elderly parents to adult children that follow this pattern, and appear to be viewed as morally acceptable by both sides. Especially if the parents have consumption habits and aspiration levels that were shaped in less affluent periods, they may find that they "have no use" for part of their income, whereas their children, despite their higher income, can easily spend more. (For a related phenomenon regarding consumption patterns, see a front-page story in the *Wall Street Journal* of July 8, 1996.) Our model is not designed to explain the phenomenon of intergenerational transfers. It only attempts to show that such transfers may not be viewed as unfair even when they are non-Pigouvian.

Moreover, any reference to a "status quo" as a basis for a moral right or claim may be viewed as an implicit recognition of (adjusted) aspiration levels as an ethical reference point. Specifically, the legal system appears to accept a divorced spouse's claim for a certain standard of living, a tenant's right to a certain bound on rent increase, and so forth.

Aspiration levels in their original psychological meaning may be too subjective and too easily manipulable to serve as a basis for normative arguments in interpersonal interactions. But they do seem to capture some intuitive notions of "what is fair" that economic theory tends to ignore.

Another dangerously simplistic conclusion that may be associated with our model is that, since aspirations adapt, there is no point in attempting to improve objective conditions of human life. Indeed, based on adaptation level theory, Brickman and Campbell (1971) make a similar point, and argue that "there is no true solution to the problem of happiness" but "getting off the Hedonic Treadmill." (See also Thurow 1971, 1973, 1975, 1980 and Ittleson et al. 1974.)

While many philosophers and religious preachers have offered similar arguments over the years, this is a rather dangerous position for a social scientist. First, as noted by Kahneman and Varey (1991), not all experiences are subject to adaptation. On the contrary, some aversive experiences escalate with time. Examples such as hunger and other forms of physical deprivation illustrate. Second, adaptation and habituation are constrained by the social context. An agent may never be satisficed with a given level of income should all around her enjoy higher levels of income. Finally, a distinction should be drawn between a normative recommendation for an individual and for a society. An individual may choose to step off the "Hedonic Treadmill"; but one can hardly be excused for pushing others off it, leaving more space for oneself.

Yet, aspiration levels can be modified by education, exposure to information, and adaptation. Further, it is sometimes easier to reduce aspirations than it is to improve objective payoffs. The psychological literature provides evidence that people typically are not fully aware of potential effects of adaptation and of aspiration level adjustments. Studies such as Kahneman and Snell (1990, 1992), and Loewenstein and Adler (1995) conclude that, partly due to these processes, many individuals are poor predictors of their own future tastes. It follows that normative economics should take these considerations into account. (See also Scitovsky 1976 who argued that consumers need not know "what's best for them.")

In our model, the lower are an individual's aspiration levels, the happier she is. One may conclude that, given the choice, as, say, in the case of educating a child, we should select the lowest aspirations possible. But this would be premature. First, due to adaptation and interpersonal comparisons, aspiration levels are never fully controlled. Hence we cannot claim to have found a "shortcut to happiness" that is independent of objective payoffs. Second, the latter are not independent of the aspiration levels. Specifically, higher aspiration levels prod experimentation, which may lead to objectively higher payoffs. Indeed, Gilboa and Schmeidler (1996) shows that ambitious but realistic aspiration level adjustments lead to objectively optimal choice. (See also Gilboa and Schmeidler 2001) Thus, reducing a person's aspiration level makes her happier given the same objective payoffs, but may also negatively affect the payoffs she is likely to get. Striking the balance between the positive and the negative effects of ambition is therefore a delicate matter.

Our model highlights welfare implications that information might have. For example, information about other agents' income may have direct effects on one's well-being. Moreover, if aspiration levels are more readily adjusted upward than they are downward, one may argue that, on average, people are better off knowing less; or that a more segregated society would allow more people to feel that they are "Number One." This is a dangerous idea. It may serve various political causes. As explained above, it may also hinder objective progress. But it would be wrong to pretend that information has no subjective cost.

Similarly, one may argue, with various degrees of honesty, that segregation and even discrimination are beneficial to disadvantaged, or discriminated-against populations. Again, an outrageous notion. Yet, it is a mistake to avoid such arguments by choosing a theoretical model that does not even allow their formalization.

To conclude, there appear to be many more claims that we are not willing to make than that we are. It is entirely possible that, from a normative point of view, little can be said at this level of generality. We hope, however, that the model presented here may be of help in discussing some specific normative issues.

5 Concluding comments and further research

5.1 According to cognitive dissonance theories (Festinger 1957), people tend to provide post-hoc explanations for various facts. For instance, when a certain activity results in a low monetary payoff x_t, the individual may attempt to believe that there were other reasons to engage in this activity. In our model, this can be viewed as an adjustment of the coefficients s_{it} for $i < t$. Assuming that preceding facts should have decreased the aspiration level for fact t, the individual's current well-being increase. (See also Shafir et al. (1993) on preferences for choices that can be more readily justified.)

5.2 When facts involve other agents as protagonists, they may be ordered in more than one way. In particular, one may wish to list a temporal fact according to its time of occurrence, rather than the time in which it became known to the agent. Alternatively, one can use an extension of the model in which, as above, the coefficients s_{it} may be adjusted retrospectively.

5.3 Another extension of this model would differentiate between upward and downward adjustments of the aspiration levels. It appears that people "get used" to higher payoffs more readily than to lower ones. In particular, this assumption may partly explain such economic phenomena as wage rigidity. Also, under the assumption that aspiration levels adjust upward faster that they do downward, applying utilitarian criteria to our measure of individual well-being would result in preference for equality.

5.4 Modeling justifications explicitly allows capturing some aspects of regret theories. For instance, having fewer options may serve as a justification for a given choice and for its outcomes. Correspondingly, it can reduce regret.

5.5 With normative applications in mind, one might wonder to what extent our theory describes "rational" evaluation of well-being. For instance, can such phenomena as minimization of cognitive dissonance be part of a rational evaluation model?

We tend to answer this particular question in the negative. Knowingly adjusting one's beliefs to match actual performance has a flavor of irrationality. Specifically, such an exercise may fail should the individual be fully aware of the analysis of her cognitive processes. (See Gilboa 1991 and Gilboa and Schmeidler 2001 for a related definition of "rationality.") However, there appears to be nothing irrational in an individual taking into account her future habituation to, say, high income.

Our approach is to start out with a cognitively plausible descriptive theory, and to study within it the boundaries of rationality, and of practical normative recommendations.

5.6 The term "normative science" is used with more than one meaning. Should a normative scientist devise algorithms to obtain goals that were dictated to her, as would an engineer? Or should she tell people what goals they should have, as a moral preacher might do? We find it useful to define the role of the normative scientist as separate from that of the engineer, as well as from that of the preacher. Rather than taking goals as given, or determining what they should be, we focus on the scientist's task of modeling and analyzing the moral and ethical preferences of decision makers, such as the preferences over societies one might belong to.[2] To the extent that our model might have normative implications, it is in this light that they should be construed.

Appendix
Proofs and related comments

Proof of the Theorem. The necessity of the axioms and the uniqueness result are straightforward. We therefore prove only sufficiency, i.e., that (i) implies (ii). It may be convenient to prove this result in two steps. Noting that U is a linear functional, we first derive a more explicit linear representation. We then show that this representation is algebraically equivalent to U. Specifically, we state two propositions:

Proposition 1. *Assume that $\geq$ satisfies A1–A4. There exist real valued vectors $(\beta^T)_{T\geq 1}$ such that:*

(a) *for every $T > 1$, $\beta_T^T \neq 0$ or $[\beta_t^T = \beta_t^{T-1}$ for all $t < T]$;*
(b) *for every $T, T' \geq 1$, and every $x^T, y^{T'}$,*

$$x^T \geq y^{T'} \text{ iff } \beta^T \cdot x^T \geq \beta^{T'} \cdot y^{T'}$$

where $\cdot$ denotes inner product.

Proposition 2. *For every $T \geq 1$, define $w_T = \beta_T^T$. For $T > 1$, if $\beta_T^T \neq 0$, define*

$$s_{iT} = \frac{\beta_i^{T-1} - \beta_i^T}{\beta_T^T},$$

and if $\beta_T^T = 0$, define $s_{iT} = 0$.

Then, for every $T \geq 1$, and every x^T,

$$U(x^T) = \beta^T \cdot x^T$$

where

$$U(x^T) = \sum_{t=1}^{T} w_t(x_t^T - a_t(x^T))$$

and

$$a_t(x^T) = \sum_{i=1}^{t-1} s_{it} x_i^T.$$

Thus Proposition 2 guarantees that the numerical representation of Proposition 1 and that in part (ii) of the Theorem are equivalent.

[2] In this sense, normative theories are descriptive.

Proof of Proposition 1. Assume that $\geq$ satisfies A1–A4. The proof proceeds in several steps.

Lemma 1. *For every* $T \geq 1$, *and every* x^T, y^T, *and* $z^T : x^T \geq y^T$ *iff* $x^T + z^T \geq y^T + z^T$.

Proof. A repeated application of A3.

Lemma 2. *For every* $T \geq 1$ *there exists a vector* β^T *such that for every* x^T, y^T,

$$x^T \geq y^T \quad \textit{iff} \quad \beta^T \cdot x^T \geq \beta^T \cdot y^T.$$

Moreover, each β^T *is unique up to a multiplication by a positive number.*

Proof. Note that the restriction of $\geq$ to $\Re^T$ is a continuous weak order (by A1 and A2). In view of Lemma 1, the proof is standard.

Select a sequence of vectors $(\hat{\beta}^T)_{T\geq 1}$ provided by Lemma 2. We wish to show that each can be re-scaled such that together they represent preferences across different $\Re^T$'s as well. To this end, we use a few auxiliary lemmata.

Lemma 3. *For every* $T, T' \geq 1$, *the set* $\{(x^T, y^{T'}) \mid x^T > y^{T'}\}$ *is open (in the standard topology) in* $\Re^{T+T'}$.

Proof. Endow the set $X = \bigcup_{T=1}^{\infty} \Re^T$ with the topology whose base is the union of the standard topologies on each $\Re^T$. (Convergence in this topology requires that a net consist of vectors of identical lengths from some element on, and that they converge in the corresponding topology.) X is a separable metric space; a metric for it can be defined by

$$d(x^T, y^{T'}) = \begin{cases} |T - T'| & \text{if } T \neq T' \\ l(|x^T - y^{T'}\|) & \text{if } T = T' \end{cases}$$

where $l(a) = \min(a, 1)$ and $|\cdot\|$ denotes the Euclidean distance in the appropriate space. Hence, by Debreu (1983) (Chapt. 6, Theorem II, p. 109), $\geq$ admits a continuous real-valued representation. This implies the desired result.

In the following, 0^T denotes the origin in $\Re^T$.

Lemma 4. *For every* $T \geq 1 : 0^T \approx 0^{T+1}$.

Proof. If not, consider a pair x^T and y^{T+1} such that $x^T \approx y^{T+1}$. (The existence of which is guaranteed by A4.) Using A3, one obtains a contradiction.

Lemma 5. *For every* $T, T' \geq 1$, *and every* $x^T, y^{T'} : x^T \geq y^{T'}$ *iff* $-x^T \leq -y^{T'}$.

Proof. Otherwise, A3 would lead to a contradiction to Lemma 4.

Lemma 6. *For every* $T, T' \geq 1$, *every* $x^T, y^{T'}$, *and every* $\lambda > 0 : x^T \geq y^{T'}$ *iff* $\lambda x^T \geq \lambda y^{T'}$.

Proof. It is sufficient to prove the "only if" part. Inductive application of A3 proves if for a natural λ. Similarly, A3 proves that $x^T > y^{T'}$ implies $nx^T > ny^{T'}$. Thus the "only if" part follows for every positive rational λ. Lemma 3 concludes.

Lemma 7. *For every $T \geq 1$, if $\hat{\beta}^T \neq 0^T$, then $\hat{\beta}^{T+1} \neq 0^{T+1}$.*

Proof. Otherwise, there is $x^T > 0^T$, but for all $a \in \Re$, $(x^T, a) \approx 0^{T+1} \approx 0^T < x^T$, contradicting A4.

Completion of the Proof of Proposition 1. We now turn to define the vectors $(\beta^T)_{T\geq 1}$ as positive multiples of the respective $(\hat{\beta}^T)_{T\geq 1}$. If all the $\hat{\beta}^T$ vanish, then, by Lemmata 2 and 4, all vectors in X are equivalent, and the proof is complete. Let T_0 be the first number for which $\hat{\beta}^T$ does not vanish. By Lemma 7, $\hat{\beta}^T \neq 0^T$ for every $T > T_0$ as well.

For $T < T_0$, $\beta^T = \hat{\beta}^T$ is uniquely defined (as the origin). For T_0, set $\beta^{T_0} = \hat{\beta}^{T_0}$. We now wish to show by induction on $T \geq T_0$ that there exists $\alpha > 0$ such that, by defining $\beta^T = \alpha\hat{\beta}^T$, the vectors $(\beta^t)_{t \leq T}$ represent $\geq$ for all vectors of length T or less (where representation is defined as in part (b) of Proposition 1). Note that this claim holds for T_0.

Assume that the claim is true for $T(\geq T_0)$, and we prove it for $T+1$. In view of A4, it suffices to find $\alpha > 0$ such that $\beta^{T+1} = \alpha\hat{\beta}^{T+1}$ would satisfy

$$x^T \geq y^{T+1} \quad \text{iff} \quad \beta^T \cdot x^T \geq \beta^{T+1} \cdot y^{T+1}.$$

Choose $\bar{x}^T > 0^T$. Let $\bar{y}^{T+1}$ satisfy $\bar{x}^T \approx \bar{y}^{T+1}$. (Again, existence of such $\bar{y}^{T+1}$ is guaranteed by A4.) By Lemmata 2 and 4, $\hat{\beta}^{T+1} \cdot y^{T+1} > 0$. We can therefore define

$$\alpha = \frac{\beta^T \cdot \bar{x}^T}{\hat{\beta}^{T+1} \cdot \bar{y}^{T+1}} > 0,$$

so that $\beta^T \cdot \bar{x}^T = \beta^{T+1} \cdot \bar{y}^{T+1}$.

Let there be given x^T, y^{T+1}. Distinguish among three cases:

(i) If $x^T \approx 0^T \approx 0^{T+1}$, i.e., $\beta^T \cdot x^T = 0$, then $x^T > (\approx, <) y^{T+1}$ iff $0^{T+1} > (\approx, <) y^{T+1}$. By Lemma 2, this is the case iff $\beta^T \cdot x^T = 0 = \beta^{T+1} \cdot 0^{T+1} > (=, <) \beta^{T+1} \cdot y^{T+1}$.

(ii) If $x^T > 0^T \approx 0^{T+1}$, i.e., $\beta^T \cdot x^T > 0$, there exists $\lambda > 0$ such that $\lambda x^T \approx \bar{x}^T$. By Lemma 6, $x^T > (\approx, <) y^{T+1}$ iff $\lambda x^T > (\approx, <) \lambda y^{T+1}$, i.e., iff $\bar{y}^{T+1} > (\approx, <) \lambda y^{T+1}$, which (by Lemma 2 again) is equivalent to $\beta^{T+1} \cdot \bar{y}^{T+1} > (=, <) \lambda\beta^{T+1} \cdot y^{T+1}$. Since $\beta^{T+1} \cdot \bar{y}^{T+1} = \beta^T \cdot \bar{x}^T = \lambda\beta^T \cdot x^T$, it follows that $x^T > (\approx, <) y^{T+1}$ iff $\beta^T \cdot x^T > (=, <) \beta^{T+1} \cdot y^{T+1}$.

(iii) Finally, if $x^T < 0^T \approx 0^{T+1}$, i.e., $\beta^T \cdot x^T < 0$, by Lemma 5, $x^T > (\approx, <) y^{T+1}$ iff $-x^T < (\approx, >) - y^{T+1}$. Using case (ii) above, this is equivalent to $-\beta^T \cdot x^T < (=, >) - \beta^{T+1} \cdot y^{T+1}$, and therefore also to $\beta^T \cdot x^T > (=, <) \beta^{T+1} \cdot y^{T+1}$.

Finally, we prove that the vectors $(\beta^T)_{T\geq 1}$ satisfy part (a) of the Proposition. For every $T \geq 1$, if $\beta^T_T \neq 0$, we are done. If, however, $\beta^T_T = 0$, use A4 to conclude that $y^T \approx (y^T, a)$ for every y^T and every $a \in \Re$. By the representation of $\geq$ as in part (b),

$$x^T \geq y^T \approx (y^T, a) \quad \text{iff} \quad \sum_{t=1}^{T} \beta_t^T x_t^T \geq \sum_{t=1}^{T} \beta_t^{T+1} y_t^T$$

for every x^T, y^T. This implies that $\beta_t^{T+1} = \beta_t^T$ for all $t \leq T$, and thus completes the proof of Proposition 1. $\diamond$

Proof of Proposition 2. For every $a \in \Re$, $U(a) = w_1 a = \beta_1^1 a$. For $T > 1$, assume that $U(x^t) = \beta^t \cdot x^t$ has been proved for every $t < T$ and every x^t. Consider x^T, and denote its $(T-1)$-long prefix by x^{T-1}. Then

$$\begin{aligned}\beta^T \cdot x^T &= \sum_{i=1}^{T-1} \beta_i^T x_i^T + \beta_T^T x_T^T = \sum_{i=1}^{T-1} \beta_i^{T-1} x_i^T + \sum_{i=1}^{T-1} (\beta_i^T - \beta_i^{T-1}) x_i^T + \beta_T^T x_T^T \\ &= \sum_{i=1}^{T-1} \beta_i^{T-1} x_i^T + \beta_T^T \left[x_T^T - \sum_{i=1}^{T-1} \left(\frac{\beta_i^{T-1} - \beta_i^T}{\beta_T^T} \right) x_i^T \right] \\ &= U(x^{T-1}) + w_T \left[x_T^T - \sum_{i=1}^{T-1} s_{iT} x_i^T \right] = U(x^T). \qquad \diamond\end{aligned}$$

Remark 1. *By definition, w_T are linear functions of $(\beta^T)_{T \geq 1}$. Conversely, it can be seen that $(\beta^T)_{T \geq 1}$ are uniquely defined by $\beta^T \cdot x^T = U(x^T)$, and are linear functions of $(w_T)_{T \geq 1}$. Explicitly,*

$$\beta_i^T = w_i - \sum_{t=i+1}^{T} s_{it} w_t.$$

Remark 2. *In the absence of A4, the Theorem does not hold. For instance, $\geq$ may rank vectors lexicographically, first by length, and then according to β^T within each $\Re^T$.*

Further, A4 cannot be replaced by assuming, say, that $0^T \approx 0^{T+1}$ for every T. To see this, define $\geq$ by the following function:

$$U(x) = x \quad \textit{for all } x \in \Re$$

$$U(x^T) = 2x_1^T \quad \textit{for all } x^T \textit{ with } T > 1.$$

$\geq$ satisfies A1–A3, and the condition $0^T \approx 0^{T+1}$, but does not satisfy A4.

References

Abel AB (1990) Asset prices under habit formation and catching up with the joneses. Am Econ Rev 80: 38–42

Appley MH (ed) (1971) Adaptation level theory: A symposium. Academic Press, New York

Becker GS (1996) Accounting for tastes. Harvard University Press, Cambridge, MA
Berglas S, Jones EE (1978) Drug choice as a self-handicapping strategy in response to noncontingent success. J Personality Soc Psychol 36: 405–417
Brickman P, Campbell DT (1971) Hedonic relativism and planning the good society. In: Appley MH (ed.) Adaptation-level theory: A symposium. Academic Press, New York
Brickman P, Coates D, Janoff-Bulman R (1978) Lottery winners and accident victims: Is happiness relative. J Personality Soc Psychol 36: 917–927
Campbell A (1981) The sense of well-being in America: Recent patterns and trends. McGraw-Hill, New York
Congleton RD (1989) Efficient status seeking: Externalities and the evolution of status games. J Econ Behav Organiz 11: 175–190
Debreu G (1983) Mathematical economics: Twenty paper of Gerard Debreu. Cambridge University Press, Cambridge
Diener E (1984) Subjective well-being. Psychol Bull 95: 542–575
Duesenberry JS (1949) Income, saving, and the theory of consumer behavior. Harvard University Press, Cambridge, MA
Duncan O (1975) Does money buy satisfaction? Soc Indicators Res 2: 267–274
Easterlin RA (1974) Does economic growth improve the human lot? Some empirical evidence. In: David PA, Reder MW (eds.) Nations and households in economic growth. Academic Press, New York
Festinger L (1957) A theory of cognitive dissonance. Stanford University Press, Stanford
Foley D (1967) Resource allocation and the public sector. Yale Econ Essays 7: 45–98
Frank RH (1985a) Choosing the right pond: Human behavior and the quest for status. Oxford University Press, New York
Frank RH (1985b) The demand for unobservable and other non-positional goods. Am Econ Rev 75: 101–116
Frank RH (1989) Frames of reference and the quality of life. Am Econ Rev 79: 80–85
Fredrickson BL, Kahneman D (1993) Duration neglect in retrospective evaluations of affective episodes. J Personality Soc Psychol 65: 45–55
Gilboa I (1991) Rationality and ascriptive science. mimeo
Gilboa I, Schmeidler D (1996) Case-based optimization. Games Econ Behav 15: 1–26
Gilboa I, Schmeidler D (1995) Case-based decision theory. Q J Econ 110: 605–639
Gilboa I, Schmeidler D (2001) A theory of class-based decisions. Cambridge University Press (forthcoming)
Harsanyi J (1992) Game and decision-theoretic models in ethics. In: Aumann RJ, Hart S (eds.), Handbook of game theory with applications to economics, I. Elsevier Science Publishers, New York 669–707
Hayakawa H, Venieris Y (1977) Consumer interdependence via reference groups. J Polit Econ 85: 599–615
Helson H (1964) Adaptation level theory. Harper and Row, New York
Hirsch F (1976) Social limits to economic growth. Harvard University Press, Cambridge, MA
Ittleson WH, Proshansky HM, Rivlin LG, Winkel GH (1974) An introduction to environmental psychology. Holt, Rinehart, and Winston, New York
Kahneman D (1994) New challenges to the rationality assumption. J Institutional Theoret Econ 150: 18–36
Kahneman D, Fredrickson BL, Schreiber CA, Redelmeier DA (1993) When more pain is preferred to less: Adding a better end. Psychol Sci 4: 401–405
Kahneman D, Snell J (1990) Predicting utility. In: Hogarth RM (ed.) Insights in decision making. University of Chicago Press, Chicago
Kahneman D, Snell J (1992) Predicting a changing taste. J Behav Decision Making 5: 187–200

Kahneman D, Varey C (1991) Notes on the psychology of utility. In: Elster J, Roemer J (eds.), Interpersonal comparisons of utility. Cambridge University Press, New York 127–163
Kahneman D, Wakker PP, Sarin R (1997) Back to bentham? Explorations of experienced utility. Q J Econ 112: 375–405
Kapteyn A, Wansbeek TJ (1982) Empirical evidence on preference formation. J Econ Psychol 3: 137–154
Kapteyn A, Wansbeek TJ (1985) The individual welfare function. J Econ Psychol 6: 333–363
Kapteyn A, Wansbeek TJ, Buyze J (1980) The dynamics of preference formation. J Econ Behav Organiz 1: 123–157
Layard R (1980) Human satisfaction and public policy. Econ 90: 737–750
Leibenstein H (1948) Bandwagon, snob, and veblen effects in theory of consumer's demand. Q J Econ 63: 165–201
Loewenstein G, Adler D (1995) A bias in the prediction of taste. Econ J 105: 929–937
March JG (1988) Variable risk preferences and adaptive aspirations. J Econ Behav Organiz 9: 5–24
McClelland DC (1961) The achieving society. Van Nostrand, Princeton, NJ
Mezias JS (1988) Aspiration level effects: An empirical investigation. J Econ Behav Organiz 10: 389–400
Ng YK, Wang J (1993) Relative income, aspiration, environmental quality, individual and policy myopia. Math Soc Sci 26: 3–23
Payne J, Laughunn D, Crum R (1980) Translation effects of gamblers and aspiration level effects in risky choice behavior. Manag Sci 26: 1029–1060
Payne J, Laughunn D, Crum R (1981) Further tests of aspiration level effects in risky choice behavior. Manag Sci 27: 953–958
Pazner E, Schmeidler D (1974) A difficulty in the concept of fairness. Rev Econ Stud 41: 441–443
Persky J, Tam M-Y (1990) Local status and national social welfare. J Reg Sci 30: 229–238
Pigou AC (1920) Economics of welfare. Macmillan, London
Rainwater L (1974) What money buys. Basic Books, New York
Ryder HE, Heal GM (1973) Optimal growth with intertemporally dependent preferences. Rev Econ Stud 40: 1–31
Scitovsky T (1976) The Joyless economy. Oxford University Press, Oxford
Sedikides C (1993) Assessment, enhancement, and verification determinants of self-evaluation process. J Personality Soc Psychol 65: 317–338
Schmeidler D, Vind K (1972) Fair net trades. Econometrica 40: 637–642
Schoemaker P (1982) The expected utility model: Its variants, purposes, evidence and limitations. J Econ Lit 20: 529–563
Seidman LS (1987) Relativity and efficient taxation. Southern Econ J 54: 463–474
Shafir E, Simonson I, Tversky A (1993) Reason-based choice. Cognition 49: 11–36
Simon HA (1957) Models of man. Wiley, New York
Thurow L (1971) The income distribution as a pure public good. Q J Econ 85: 327–336
Thurow L (1973) Toward a definition of economic justice. Publ Interest 31: 56–80
Thurow L (1975) Generating inequality. Basic Books, New York
Thurow L (1980) Zero-sum society. Basic Books, New York
Tice DM, Baumeister RF (1990) Self-esteem, self-handicapping and self-presentation: The strategy of inadequate practice. J Personality 58: 443–464
Tomes N (1986) Income distribution, happiness, and satisfaction: A direct test of the interdependent preferences model. J Econ Psychol 7: 425–446
Trope Y (1979) Uncertainty-reducing properties of achievement tasks. J Personality Soc Psychol 37: 1505–1518
Varian HR (1974) Equity, envy, and efficiency. J Econ Theory 9: 63–91
Veblen T (1899) The theory of the leisure class. MacMillan, New York

Soc Choice Welfare (2001) 18: 289–301

Social Choice and Welfare

Characterizing uncertainty aversion through preference for mixtures

Peter Klibanoff

Department of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Rd., Evanston, IL 60208, USA (e-mail: peterk@nwu.edu)

Received: 20 February 1998/Accepted: 25 March 1999

Abstract. Uncertainty aversion is often modelled as (strict) quasi-concavity of preferences over uncertain acts. A theory of uncertainty aversion may be characterized by the pairs of acts for which strict preference for a mixture between them is permitted. This paper provides such a characterization for two leading representations of uncertainty averse preferences; those of Schmeidler [24] (Choquet expected utility or CEU) and of Gilboa and Schmeidler [16] (maxmin expected utility with a non-unique prior or MMEU). This characterization clarifies the relation between the two theories.

1 Introduction

A large body of work has recently emerged in economics and decision theory with the goal of representing behavior in the face of subjective uncertainty that may violate the independence axiom of subjective expected utility theory. One branch of this literature, and the one that will be the focus below, considers preferences that may violate independence by displaying a preference for facing risk (or "objective" probabilities) as opposed to uncertainty. This preference is known as *uncertainty aversion*. One motivation for examining these preferences are the well-known problems posed by Ellsberg [10] and the

I thank Mark Machina, Michèle Cohen, Edi Karni, Paolo Ghirardato and other participants in the 21st Seminar of the European Group of Risk and Insurance Economists for stimulating discussions that motivated me to begin this work. Eddie Dekel and Massimo Marinacci provided very helpful comments on a preliminary draft. An anonymous referee provided useful suggestions. All errors are my own.

huge experimental literature that has followed, in which many individuals behave as if they were uncertainty averse.

There are several ways that one could imagine defining uncertainty aversion. The definition that I will use here, and the one that has dominated the literature so far, is due to Schmeidler [24][1]. It states that for any two acts that an individual is indifferent between, a mixture over these two acts is at least as preferred as either act.[2] One may interpret this requirement as saying that the individual likes smoothing expected utility across states. This smoothing has the effect of making the outcome less subjective, and therefore such a mixing operation could be called "objectifying"[3]. Thus, in a natural sense, such an individual is displaying an aversion to uncertainty. An equivalent way of stating this characteristic is to say that preferences are quasi-concave ($f \succeq g$ and $\alpha \in (0,1)$ implies $\alpha f + (1-\alpha)g \succeq g$). In particular, observe that if $f \sim g$ then quasi-concavity allows $\alpha f + (1-\alpha)g \succ g$ while independence requires $\alpha f + (1-\alpha)g \sim g$.

From this viewpoint, a theory of uncertainty averse preferences may be characterized by the set of violations of independence in the direction of strict preference for mixture that it allows. The goal of this paper is to provide a characterization of this kind for two leading axiomatic theories of uncertainty aversion, the Choquet expected utility (CEU) theory of Schmeidler [24] and the maxmin expected utility (MMEU) theory of Gilboa and Schmeidler [16]. Such a characterization is useful not only from the point of view of theoretical understanding, but also as a guide to the design of experiments testing one theory of uncertainty aversion against another. For example, the results in Sect. 3 allow the easy identification of pairs of acts over which an MMEU decision maker may have a strict preference for mixture, while a CEU decision maker cannot. Furthermore, in the emerging literature applying these theories (e.g., Dow and Werlang [7], Klibanoff [17], Lo [19], Eichberger and Kelsey [9], Marinacci [21] on game theory; Wakker [25] on optimism and pessimism; Dow and Werlang [6], Chateauneuf et al. [3], Epstein and Wang [12] on financial markets; Mukerji [22] and Ghirardato [13] on contracting; and others) too often one theory or the other is adopted without much recognition of the ways in which the theories differ.

The next section introduces the CEU and MMEU theories and points out the known, yet frequently ignored, fact that under uncertainty aversion, MMEU is a strict generalization of CEU. Section 3 presents the main theorems characterizing the acts for which no convex combination is ever strictly preferred to both acts themselves under MMEU and under CEU respectively. Section 4 concludes.

---

[1] Until the recent work of Epstein [11] and Ghirardato and Marinacci [15], Schmeidler's was essentially the *only* definition used in the literature.

[2] A mixture over two acts is formally defined in Sect. 2.1 below.

[3] I thank Mark Machina for suggesting this term.

2 Two models of uncertainty aversion

2.1 Notation and set-up

Throughout the paper, preferences are a binary relation, $\succeq$, on functions (acts) $f : S \to Y$ where S is a finite set of states of the world, X a set of prizes, and Y the set of all probability measures with finite support (lotteries) on X. Thus, in each state of the world an act yields a lottery over prizes (as in Anscombe and Aumann [1]). Lotteries are evaluated according to an affine utility function $u : Y \to \Re$. Denote the probability of prize $x \in X$ in state $s \in S$ induced by act f by $f(s)(x)$. For any $\alpha \in (0,1)$, define the α-mixture over f and g, $\alpha f + (1-\alpha)g$, by

$$(\alpha f + (1-\alpha)g)(s)(x) = \alpha f(s)(x) + (1-\alpha)g(s)(x), \quad \text{for all } x \in X,\ s \in S.$$

One interpretation of the mixture $\alpha f + (1-\alpha)g$ is a randomization over the acts f and g with probabilities α and $1-\alpha$ respectively. Under this interpretation a preference for mixtures implies a preference for randomization. Such a preference for randomization is controversial in the literature. An issue is whether randomization is a way of making mixtures feasible in particular settings.[4] The correctness and (in large part) interpretation of the analysis below is independent of one's position in this debate. The objects of study are acts, and mixtures are simply particular acts. If one does not accept the randomization interpretation, preference for mixtures may be read as simply a statement about preferences over pure acts whose utility payoffs happen to be related through convex combinations.

2.2 Two models

A leading representation of uncertainty averse preferences is the CEU representation axiomatized by Schmeidler [24]. Here preferences are represented by the Choquet integral of a utility function with respect to a capacity or non-additive measure. One of the properties which characterizes such preferences is *comonotonic independence*. Two acts, f and g, are said to be comonotonic if, for no pair of states of the world s and s', $f(s) \succ f(s')$ and $g(s') \succ g(s)$.[5] Preferences satisfy comonotonic independence if, for any acts f and g, $f \succeq g$ if and only if $\alpha f + (1-\alpha)h \succeq \alpha g + (1-\alpha)h$ for all $\alpha \in (0,1)$ and all h such that f, g, h are pairwise comonotonic. This is simply a restriction of the standard independence axiom (e.g., Anscombe and Aumann [1]) to pairwise comonotonic acts. From this axiom, the following is immediate:

---

[4] For two contrasting views of the impact, in the context of uncertainty aversion and randomization, of using a model with an Anscombe-Aumann-style mixture space of acts (as here) rather than Savage-style acts, see Eichberger and Kelsey [8] and Klibanoff [18].

[5] $f(s)$ should be understood as an act which gives the lottery that act f gives in state s no matter which state occurs.

Result 1. *Suppose that preferences satisfy comonotonic independence. Then for any comonotonic acts f and g, for each* $\alpha \in (0,1)$, *either* $f \succeq \alpha f + (1-\alpha)g$ *or* $g \succeq \alpha f + (1-\alpha)g$ *or both.*

Thus, strict preference for mixtures cannot occur with comonotonic acts. Notice that this observation derives from comonotonic independence *alone* and is in no way implied by uncertainty aversion per se.

Now consider a second common representation of preferences incorporating uncertainty aversion, namely the MMEU representation axiomatized in Gilboa and Schmeidler [16]. In this work, the axiom of comonotonic independence is replaced by an alternative axiom, denoted C-independence. C-independence requires the independence axiom to hold only when the act h used to form the mixtures gives the same expected utility in every state of the world.[6] Intuitively, acts which yield the same expected utility in every state leave no room for uncertainty about which state will occur to matter. C-independence is the assumption that mixing with such an act will not change either the way in which the decision maker perceives her uncertainty or the way in which she allows her attitude towards uncertainty to affect her preferences.

Gilboa and Schmeidler [16] showed that C-independence and the standard assumptions of weak order, continuity and monotonicity together with uncertainty aversion imply that preferences can be represented by the minimum expected utility of an act, where the minimum is taken over a closed, convex set of probability measures. Notice that an act which yields constant expected utility across states is comonotonic with any other act. In fact, comonotonic independence, weak order, continuity, and monotonicity imply C-independence. This means that, under the assumption of uncertainty aversion, any preferences that can be represented by CEU can also be represented in the MMEU framework (Schmeidler [23], [24]). The converse is not true, however, as the following example makes clear.

	s_1	s_2	s_3
f	1.5	2	3.5
g	0	2.1	4
h	0.75	2.05	3.75

Set of measures: $B = \left\{(p_1, p_2, p_3) \mid p_1 = p_3, \sum_{i=1}^{3} p_i = 1, 0 \leq p_i \leq 1\right\}$

Example 1

[6] Technically the axiom is more restrictive, requiring h to give the same lottery over outcomes in each state of the world, but together with the assumptions of weak order, continuity and monotonicity the axiom as described is implied. Note that the assumptions of weak order, continuity and monotonicity were also assumed in the Choquet expected utility theory of Schmeidler [24].

In Example 1, an individual must choose over two possible pure acts, *f* and *g*, which give the expected utilities indicated above in the three possible states of the world. Observe that *f* and *g* are comonotonic. Suppose that the individual's preferences can be represented by minimum expected utility over the set of measures *B* (i.e., the set of all probability measures which assign equal weight to s_1 and s_3). Straightforward calculation shows that *f* and *g* each give a minimum expected utility of 2, while, for example, *h*, a half-half mixture between *f* and *g*, gives a minimum expected utility of 2.05. Therefore, this uncertainty averse individual will strictly prefer the mixture *h*, compared to either *f* or *g*. Since this violates comonotonic independence it shows that these preferences cannot be represented in the CEU framework, and also demonstrates that comonotonicity is not enough, in general, to guarantee that an uncertainty averse individual will not strictly prefer to objectify by mixing over acts. What is the right condition to guarantee no strict preference for mixtures in the MMEU representation? Is the comonotonicity condition a necessary as well as sufficient condition for no preference for mixing under CEU? The next section provides results to answer these questions.

3 Characterizing preference for mixtures

In examining when strict preference for mixtures is possible (or impossible) under the two theories, it is helpful to consider a previous result characterizing preference for mixtures under MMEU for a *specified* set of probability measures. While such results are of more interest in a setting where certain beliefs are focal (e.g., equilibrium beliefs in game theory), they will be used in proving the theorems to follow that apply to the whole domain of the respective theories.

Theorem 1. *(Klibanoff [17]) For any acts f and g such that* $f \succeq g$*, no mixture over these acts will be strictly preferred to either alone if and only if there exists some measure q in the set of measures such that q minimizes the expected utility of f over the set and such that the expected utility of f with respect to q is at least the expected utility of g with respect to q.*

For acts which the decision maker is indifferent between this simplifies to:

Corollary 1. *(Klibanoff [17]) For any acts f and g such that* $f \sim g$*, no mixture over these acts will be strictly preferred to either alone if and only if there exists some measure q in the set of measures such that q minimizes the expected utility of both f and g over the set of measures.*

Now we characterize the set of acts for which *no* MMEU decision maker would have a strict preference for a mixture. This result and the corresponding result for the CEU case are provided in the next two theorems.

Theorem 2. *Fix acts f and g. No convex combination of f and g will ever be strictly preferred to either alone (given MMEU preferences) if and only if (i) f* *weakly dominates g or vice-versa (i.e.,* $u(f(s)) \geq (\leq) u(g(s))$, *for all* $s \in S$.*)* **or** *(ii) there exists an* $a \geq 0, b \in \Re$ *such that either* $u(g(s)) = au(f(s)) + b$ *for all* $s \in S$ *or* $u(f(s)) = au(g(s)) + b$ *for all* $s \in S$.

Proof. The difficult direction is to show that no strict preference implies (i) or (ii). (The opposite direction is left for the reader to verify.) The key step in the proof is to show that the conditions for a convex combination to reduce uncertainty are equivalent to the existence of a pair of probability vectors satisfying a set of linear inequalities. This is done in the lemma below. The only task remaining is then to characterize existence. To do this I apply a well known result from the theory of linear inequalities, Motzkin's Theorem of the alternative (see e.g., Mangasarian [20]). The existence of a solution to the resulting alternative system is then (after a bit of rearrangement) shown to be equivalent to the conditions of the theorem.

Let the vector of utility payoffs to the act f be denoted $\mathscr{F} (\equiv \{u(f(s))\})$ and similarly for $\mathscr{G}$. The following lemma reduces the conditions for a convex combination of f and g to possibly reduce uncertainty to a question of existence of probabilities satisfying certain linear inequalities.

Lemma 1. *Fix* $\mathscr{F}$ *and* $\mathscr{G}$. *There exists a non-empty, closed, convex set of measures B for which some mixture of* $\mathscr{F}$ *and* $\mathscr{G}$ *is strictly preferred to either alone if and only if there exist probability vectors* $p1$ *and* $p2$ *satisfying:*

(i) $\mathscr{F} \cdot p2 - \mathscr{F} \cdot p1 > 0$

(ii) $\mathscr{F} \cdot p1 - \mathscr{G} \cdot p1 < 0$

(iii) $\mathscr{G} \cdot p1 - \mathscr{G} \cdot p2 > 0$

and

(iv) $\mathscr{G} \cdot p2 - \mathscr{F} \cdot p2 < 0$

Proof.

($\Leftarrow$) Suppose such $p1$ and $p2$ exist. Let B be the set of all convex combinations of $p1$ and $p2$. Either $f \succeq g$ or $g \succeq f$ or both. If $f \succeq g$ then by (i) and (ii), $p1$ is the only minimizer in B of the expected utility of f and the expected utility of f under $p1$, $\mathscr{F} \cdot p1$, is less than the expected utility of g under $p1$, $\mathscr{G} \cdot p1$. Therefore, by Theorem 1, there exists a mixture which is strictly preferred. If $g \succeq f$ then by (iii) and (iv) and Theorem 1 the same conclusion holds.

($\Rightarrow$) Suppose such a B exists. Consider the set $A_f = \{p \mid p \in \text{argmin}_{p \in B} \mathscr{F} \cdot p\}$ and $A_g = \{p \mid p \in \text{argmin}_{p \in B} \mathscr{G} \cdot p\}$. Consider $p1 \in A_f$ and $p2 \in A_g$. By definition of these sets we must have

(a) $\mathscr{F} \cdot p2 - \mathscr{F} \cdot p1 \geq 0$

and

(b) $\mathscr{G} \cdot p1 - \mathscr{G} \cdot p2 \geq 0$.

Suppose that (a) holds with equality for some such $p1$ and $p2$. Then if

$g \succeq f$, $\mathscr{G} \cdot p2 \geq \mathscr{F} \cdot p2$ which implies that the condition in Theorem 1 is satisfied and no mixture is strictly preferred. If $f \succ g$, then $\mathscr{F} \cdot p1 = \mathscr{F} \cdot p2 > \mathscr{G} \cdot p2$ and again appealing to Theorem 1 no mixture is strictly preferred. Similar arguments show that if (b) holds with equality for some such $p1$ and $p2$ then no mixture is strictly preferred. Thus for there to be a mixture that is strictly preferred it must be that for all $p1 \in A_f$ and $p2 \in A_g$,

(i) $\mathscr{F} \cdot p2 - \mathscr{F} \cdot p1 > 0$

and

(iii) $\mathscr{G} \cdot p1 - \mathscr{G} \cdot p2 > 0.$

Can it be that $\mathscr{F} \cdot p1 - \mathscr{G} \cdot p1 \geq 0$? This and (iii) would imply $\mathscr{F} \cdot p1 - \mathscr{G} \cdot p2 > 0$ which implies $f \succ g$ and thus by Theorem 1 and the hypothesis no mixture would be strictly preferred. Therefore,

(ii) $\mathscr{F} \cdot p1 - \mathscr{G} \cdot p1 < 0$

must hold. By an analogous argument,

(iv) $\mathscr{G} \cdot p2 - \mathscr{F} \cdot p2 < 0$

must hold as well and we are done. *QED*

Now that the lemma has been proved, the next step in proving the theorem is to combine conditions (i)–(iv) with the restrictions implied by the fact that $p1$ and $p2$ must be probability vectors. To this end, let n be the number of states in S. Then $\mathscr{F}, \mathscr{G}, p1$ and $p2$ are n-vectors. Let $\mathscr{F}$ and $\mathscr{G}$ be row vectors and $p1$ and $p2$ be column vectors. Let e be a row n-vector of 1's. Let

$$p = \begin{bmatrix} p1 \\ p2 \end{bmatrix}$$

and

$$A = \begin{bmatrix} -\mathscr{F} & \mathscr{F} \\ \mathscr{G} - \mathscr{F} & 0 \\ \mathscr{G} & -\mathscr{G} \\ 0 & \mathscr{F} - \mathscr{G} \end{bmatrix}.$$

Observe that (i)–(iv) is equivalent to $Ap > 0$. Furthermore the requirement that $p1$ and $p2$ be probabilities is equivalent to $p \geq 0$ and

$$\begin{bmatrix} e & 0 \\ 0 & e \end{bmatrix} p = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{1}$$

Equivalently, we can replace the normalization (1) with the condition

$$[e \quad -e]p = 0$$

and the condition $p \geq 0$ with the equivalent

$$Ip \geq 0$$

where I is a $2n \times 2n$ identity matrix. To summarize, we would like to characterize when there exists a p such that

(a) $Ap > 0$

(b) $Ip \geq 0$

and

(c) $[e \quad -e]p = 0$.

By Motzkin's Theorem of the alternative (Mangasarian [20]), either (a), (b) and (c) has a solution p or

$$(*) \quad \left\langle \begin{matrix} A'y_1 + I'y_3 + [e \quad -e]'y_4 = 0 \\ y_1 > 0, y_3 \geq 0 \end{matrix} \right\rangle$$

has a solution y_1, y_3, y_4, but never both. (Note that $y_1 > 0$ means that each element of y_1 is greater than or equal to zero with at least one element strictly positive. $y_3 \geq 0$ means almost the same thing except that it allows all elements to be zero.)

All that remains is to rewrite system $(*)$ to get an interpretable condition (namely the one in the theorem.) First notice that since the elements of y_3 are all non-negative, $(*)$ has a solution if and only if

$$(**) \quad \left\langle \begin{matrix} A'y_1 + [e \quad -e]'y_4 \leq 0 \\ y_1 > 0 \end{matrix} \right\rangle$$

has a solution y_1, y_4. Adding up the inequalities determined by the first line of $(**)$ yields

$$(\mathscr{G} - \mathscr{F})'(y_{21}^1 - y_{41}^1) \leq 0,$$

where

$$y_1 = \begin{bmatrix} y_{11}^1 \\ y_{21}^1 \\ y_{31}^1 \\ y_{41}^1 \end{bmatrix}.$$

This implies that either $y_{21}^1 = y_{41}^1$ or one of f and g is weakly dominated by the other. So, a solution to $(*)$ exists if and only if either weak dominance between f and g holds or $(**)$ is satisfied with $y_{21}^1 = y_{41}^1$. Imposing the latter restriction and disaggregating the inequality in $(**)$ we obtain the system

$$\left\langle \begin{matrix} \mathscr{G}'(y_{21}^1 + y_{31}^1) - \mathscr{F}'(y_{11}^1 + y_{21}^1) + ey_4 \leq 0 \\ -\mathscr{G}'(y_{21}^1 + y_{31}^1) + \mathscr{F}'(y_{11}^1 + y_{21}^1) - ey_4 \leq 0 \\ y_1 > 0 \end{matrix} \right\rangle$$

which is equivalent to

$$(***) \quad \left\langle \begin{matrix} \mathscr{G}'(y_{21}^1 + y_{31}^1) - \mathscr{F}'(y_{11}^1 + y_{21}^1) + ey_4 = 0 \\ y_1 > 0 \end{matrix} \right\rangle.$$

Observe that without loss of generality y^1_{21} can be set to zero since it can be incorporated into y^1_{11} and y^1_{31}. Now, suppose that one of y^1_{11} or y^1_{31} is zero. Then a solution will exist if and only if either g or f or both are constant utility acts. Finally, consider the remaining case where both y^1_{11} and y^1_{31} are positive. Here a solution exists if and only if there exists an $\alpha > 0, \beta > 0$, and y_4 such that

$$\alpha\mathscr{G}' - \beta\mathscr{F}' + ey_4 = 0.$$

This last condition is equivalent to

$$\mathscr{G}' = a\mathscr{F}' + be', \quad \text{for some } a > 0,\ b \in \Re.$$

Now note that the case where $a = 0$ corresponds to the cases where g is a constant act. If only f is a constant act, simply reverse the roles of f and g and again set $a = 0$.

Pulling the different possibilities together, we have that a solution to $(*)$ exists if and only if either f and g are ordered by weak dominance or

$$\mathscr{G}' = a\mathscr{F}' + be', \quad \text{for some } a \geq 0,\ b \in \Re$$

or,

$$\mathscr{F}' = a\mathscr{G}' + be', \quad \text{for some } a \geq 0,\ b \in \Re.$$

Our application of Motzkin's Theorem now yields the desired conclusion.

QED

The analogue for CEU is given in the next theorem. Note that this result is related to the prior work of Bassanezi and Greco [2] who show that the Choquet integral is additive for all capacities if and only if the functions being integrated are comonotonic.

Theorem 3. *Fix acts f and g. No convex combination of f and g will ever be strictly preferred to either alone (given CEU preferences) if and only if (i) f weakly dominates g or vice-versa (i.e., $u(f(s)) \geq (\leq) u(g(s))$, for all $s \in S$.)* **or** *(ii) f and g are comonotonic.*

Proof.

($\Leftarrow$) It is straightforward that (i) implies the weakly dominant act will be at least as good as any mixture. Result 1 stated earlier says that (ii) implies no mixture strictly preferred.

($\Rightarrow$) We will show that Not ((i) or (ii)) implies there exists a mixture that may be strictly preferred to both f and g. Not ((i) or (ii)) implies f, g not comonotonic and no weak dominance between them. Since the two acts are not comonotonic, there exist states $s_f, s_g \in S$ such that $f(s_f) \succ f(s_g)$ and $g(s_g) \succ g(s_f)$. Consider the restriction of f and g to $\{s_f, s_g\}$. There are two possible cases:

Case I. Neither restricted act weakly dominates the other. In this case, without loss of generality assume that $f(s_f) \succeq g(s_g) \succ g(s_f) \succeq f(s_g)$. Consider the capacity v such that $v(\{s_f, s_g\}) = 1$ and $v(A) = 0$ for any set A such that $\{s_f, s_g\} \not\subseteq A$. Relative to this capacity, we can calculate the Choquet expected utility (CEU) of f and g: $\text{CEU}(f) = u(f(s_g))$ and $\text{CEU}(g) = u(g(s_f))$. By continuity of preferences, there exists an $\alpha^* \in (0,1)$ such that

$$\alpha^* g(s_g) + (1-\alpha^*) f(s_g) \succ g(s_f).$$

Taking the CEU of this convex combination with respect to v yields,

$$\begin{aligned}&\text{CEU}(\alpha^* g + (1-\alpha^*) f)\\ &\quad = \min[\alpha^* u(g(s_f)) + (1-\alpha^*) u(f(s_f)), \alpha^* u(g(s_g)) + (1-\alpha^*) u(f(s_g))]\\ &\quad > u(g(s_f)) \geq u(f(s_g)).\end{aligned}$$

Therefore $\alpha^* g + (1-\alpha^*) f \succ f$ and $\alpha^* g + (1-\alpha^*) f \succ g$ for this v.

This proves the claim for the case where neither restricted act weakly dominates the other. Now we examine the remaining possibility:

Case II. One restricted act weakly dominates the other. Without loss of generality assume $f(s_f) \succ f(s_g) \succeq g(s_g) \succ g(s_f)$. Since, over the whole space, S, we assumed neither act weakly dominates the other, there must exist an $s' \in S$ such that $g(s') \succ f(s')$.

There are several possibilities. First, suppose that $g(s') \succeq g(s_g)$. Then $g(s') \succ f(s')$ implies $f(s_g) \succ f(s')$ so that f and g are not comonotonic on $\{s_g, s'\}$ and Case I applies to the restriction of f and g to $\{s_g, s'\}$.

Another possible ordering of the states by g is $g(s_g) \succ g(s') \succeq g(s_f)$. Here $g(s') \succ f(s')$ implies $f(s_f) \succ f(s')$ and Case I applies to the restriction of f and g to $\{s_f, s'\}$.

Finally, assume (the only remaining possibility) that $g(s_f) \succ g(s') \succ f(s')$. Consider the capacity v such that $v(\{s_f, s_g, s'\}) = 1$, $v(\{s_f, s_g\}) = k$, and for all other sets v assigns the lowest nonnegative value consistent with monotonicity of the capacity. Choose $k \in (0,1)$ to satisfy

$$ku(f(s_g)) + (1-k)u(f(s')) = ku(g(s_f)) + (1-k)u(g(s')).$$

Such a k exists under our ordering assumptions. Using the capacity v,

$$\text{CEU}(f) = ku(f(s_g)) + (1-k)u(f(s')),$$

and

$$\text{CEU}(g) = ku(g(s_f)) + (1-k)u(g(s')).$$

Thus for this capacity v and utility u, $f \sim g$. Now, using the fact that we can represent the CEU preferences under v as the maxmin expected utility over the set of probability measures that are in the core of v (i.e., $\{p \mid p(s_f) + p(s_g) \geq k, p(s_f) + p(s_g) + p(s') = 1\}$) (Schmeidler [23], [24]), we can apply Corollary 1 to show that some convex combination will be strictly preferred to both f and g.

To summarize, in each of the possible cases where Not ((i) or (ii)) holds the above has shown that there exists a convex combination that may be strictly preferred to both f and g. *QED*

To facilitate a comparison with Theorem 2 the following corollary is provided:

Corollary 2. *Fix acts f and g. No convex combination of f and g will ever be strictly preferred to either alone (given CEU preferences) if and only if (i) f weakly dominates g or vice-versa (i.e.,* $u(f(s)) \geq (\leq) u(g(s))$*, for all* $s \in S$*.)* **or** *(ii) there exists an act h and weakly increasing functions w and x on* $\Re$ *such that, for all* $s \in S$*,* $u(f(s)) = w(u(h(s)))$ *and* $u(g(s)) = x(u(h(s)))$*.*

Proof. By Denneberg [5, Proposition 4.5], two functions $d, e : S \to \Re$ are comonotonic if and only if there exists a function $z : S \to \Re$ and weakly increasing functions w, x on $\Re$ such that $d = w(z)$ and $e = x(z)$. Let $d = u \circ f$, $e = u \circ g$, and $z = u \circ h$ and the result follows from theorem 3 and the fact that f and g are comonotonic if and only if $u \circ f$ and $u \circ g$ are. *QED*

To see how this result compares to Theorem 2, observe that if we require w and x to be affine then condition (ii) of the corollary is equivalent to condition (ii) of Theorem 2. While CEU prevents strict preference for mixture for acts that are weakly increasing tranformations of the same utility payoffs, MMEU does so only if the transformations are affine. Intuitively, this says that MMEU decision makers may care about the cardinal properties of the distribution of utilities across states when evaluating whether one act is more uncertain than another, while CEU individuals must consider distributions of utilities that (roughly) order states the same way as representing equivalent levels of uncertainty.

Remark. As the results above concern strict preference for mixture, the reader may wonder whether this addresses all the relevant possibilities for strict quasi-concavity of the preferences. Specifically, can there exist acts f and g satisfying (i) or (ii) of the appropriate theorem above such that indifference curves over mixtures are strictly quasi-concave, yet no mixture is strictly preferred? It is easily seen that the answer may be yes only if (ii) is violated. To see this note that if (ii) is satisfied then for any MMEU preferences the same probability measure will be used to evaluate all mixtures, generating linear indifference curves. Conversely, if (ii) is violated then $u(f)$ and $u(g)$ are not related by a positive affine transformation and therefore order probability measures distinctly. Given one minimizing measure for f and another for g, it follows that the measure used to evaluate $\alpha f + (1 - \alpha) g$ must generate more than the minimum expected utility level for one of the two acts, producing strict quasi-concavity of preferences. Arguments similar to the ones above could be used to show this more formally and demonstrate it for the Choquet case as well.[7] There is then no essential loss in limiting our analysis, as we have, to preference for mixtures. Furthermore, by examining the preference for mixtures case, we see that only weak dominance limits the extent of the quasi-concavity permitted by a violation of (ii).[8]

---

[7] See Ghirardato et al. [14] for elaboration.

[8] An alternative reason for interest in preference for mixtures per se is that such preferences correspond precisely to violations of the analogue for uncertain acts of the *betweeness* property for preferences under risk (e.g., Dekel [4]).

Thus we have a characterization of the uncertainty aversion (as expressed through strict preference for mixtures) that the two theories allow. It is hoped that this will further understanding of what distinguishes these two representations.

4 Conclusion

Theories of uncertainty aversion may differ in the circumstances under which they allow violations of independence, and in particular strict preference for mixtures. This paper has provided a characterization of those acts which may never admit such a strict preference for the two leading representations of uncertainty aversion: maxmin expected utility and Choquet expected utility. The fact that these characterizations are substantially different has implications for empirical testing of the theories as well as for those trying to apply one or the other model and wondering what the consequences of the modelling choice are. Fundamentally, CEU decision makers view uncertainty in terms of (roughly) how states are ordered by an act's utility payoffs. Given a set of acts which all induce the same ordering, a CEU decision maker acts exactly like an expected utility (and thus uncertainty neutral) decision maker. MMEU decision makers, in contrast, may view uncertainty not only in terms of ordering of states, but also in terms of how much better the payoff is in one state as opposed to another. MMEU allows the decision maker to be averse to such cardinal variations across states even among acts that order states in the same way.

References

[1] Anscombe FJ, Aumann RJ (1963) A definition of subjective probability. Ann Math Stat 34: 199–205
[2] Bassanezi R, Greco G (1984) Sull' additivita dell' integrale. Rend Sem Mat Univ Padova 72: 249–275
[3] Chateauneuf A, Kast R, Lapied A (1996) Choquet pricing for financial markets with frictions. Math Finance 6: 323–330
[4] Dekel E (1986) An axiomatic characterization of preferences under uncertainty: Weakening the independence axiom. J Econ Theory 40: 304–318
[5] Denneberg D (1994) Non-additive measure and integral. Kluwer Academic, Dordrecht
[6] Dow J, Werlang S (1992) Uncertainty aversion, risk aversion and the optimal choice of portfolio. Econometrica 60: 197–204
[7] Dow J, Werlang S (1994) Nash equilibrium under knightian uncertainty: Breaking down backward induction. J Econ Theory 64: 305–324
[8] Eichberger J, Kelsey D (1996) Uncertainty aversion and preference for randomization. J Econ Theory 71: 31–43
[9] Eichberger J, Kelsey D Non-additive beliefs and game theory. Games Econ Behav (forthcoming)
[10] Ellsberg D (1961) Risk, ambiguity, and the savage axioms. Q J Econ 75: 643–669
[11] Epstein L A definition of uncertainty aversion. Rev Econ Stud (forthcoming)

[12] Epstein L, Wang T (1994) Intertemporal asset pricing under knightian uncertainty. Econometrica 62: 283–322
[13] Ghirardato P (1994) Agency theory with non-additive uncertainty. mimeo, U.C. Berkeley
[14] Ghirardato P, Klibanoff P, Marinacci M (1998) Additivity with multiple priors. J Math Econ 30: 405–420
[15] Ghirardato P, Marinacci M (1998) Ambiguity made precise: A comparative foundation. mimeo, California Institute of Technology
[16] Gilboa I, Schmeidler D (1989) Maxmin expected utility with non-unique prior. J Math Econ 18: 141–153
[17] Klibanoff P (1996) Uncertainty, decision, and normal form games. mimeo, Northwestern University
[18] Klibanoff P (1998) Stochastic independence and uncertainty aversion. CMS-EMS Discussion Paper 1212, Northwestern University
[19] Lo KC (1996) Equilibrium in beliefs under uncertainty. J Econ Theory 71: 443–484
[20] Mangasarian O (1969) Nonlinear programming. McGraw-Hill
[21] Marinacci M (1996) Ambiguous games. mimeo, Northwestern University
[22] Mukerji S (1998) Ambiguity aversion and incompleteness of contractual form. American Economic Review 88: 1207–1231
[23] Schmeidler D (1986) Integral representation without additivity. Proceedings of the Am Math Soc 97: 255–261
[24] Schmeidler D (1989) Subjective probability and expected utility without additivity. Econometrica 57: 571–587
[25] Wakker P (1990) Characterizing optimism and pessimism directly through comonotonicity. J Econ Theory 52: 453–463

Soc Choice Welfare (2001) 18: 303–326

Social Choice and Welfare

Bayesian learning in repeated games of incomplete information

John H. Nachbar

Department of Economics, Washington University, One Brookings Drive, St. Louis, MO 63130, USA (e-mail: nachbar@wuecon.wustl.edu)

Received: 15 October 1997/Accepted: 17 March 1999

Abstract. In Nachbar [20] and, more definitively, Nachbar [22], I argued that, for a large class of discounted infinitely repeated games of *complete* information (i.e. stage game payoff functions are common knowledge), it is impossible to construct a Bayesian learning theory in which player beliefs are simultaneously weakly cautious, symmetric, and consistent. The present paper establishes a similar impossibility theorem for repeated games of *incomplete* information, that is, for repeated games in which stage game payoff functions are private information.

1 Introduction

Consider a discounted infinitely repeated game in which, prior to the start of repeated play, each player is privately informed of his payoff function for the underlying stage game. I will refer to this as a repeated game of incomplete information. In such a setting one can ask whether, in some sense, Bayesian players eventually learn from their repeated interaction to play an equilibrium of the realized game. The literature on this topic originated in Jordan [10] and has continued with Kalai and Lehrer [14, Sect. 6], Koutsougeras and Yanellis [15], Nyarko [23], Jordan [12], Lehrer [17], and Nyarko [25, 24]. Jackson and Kalai [8, 9] and Conlon [4] conduct similar analyses for recurring, rather than repeated, games.

Much of the research reported here was conducted while I was attending the Warwick Summer Research Workshop on Repeated Games and Their Economic Applications. I wish to thank two anonymous referees. And I wish to thank the Center in Political Economy at Washington University for its support.

In most of the literature on Bayesian learning in games of incomplete information, prior beliefs are degenerate in the sense that each player knows his opponent's incomplete information strategy, which gives the opponent's repeated game strategy as a function of the opponent's payoff parameter. (For simplicity, I will restrict attention to two-player games.) A degenerate prior rules out strategic uncertainty about how the opponent will play conditional on the realized stage game, even for repeated games, such as repeated coordination games, in which the set of pure equilibria is uncountable. In the present paper, I will consider instead learning theories in which players have nondegenerate priors.[1]

In Nachbar [20] and, more definitively, Nachbar [22], I argued that, for a large class of repeated games of *complete* information (i.e. stage game payoff functions are common knowledge), it is impossible to construct a Bayesian learning theory in which prior beliefs over repeated game strategies are simultaneously weakly cautious, symmetric, and consistent. Loosely, weak caution means that if a repeated game strategy is in the support of a player's belief then so are computationally trivial variants of that strategy, symmetry means that the supports of player beliefs contain repeated game strategies of comparable strategic complexity, and consistency means that the support of each player's belief contains one of his opponent's ε best responses. This paper establishes a similar impossibility theorem for beliefs over incomplete information strategies. The theorem does not preclude the construction of Bayesian learning theories *per se*, and it says nothing directly about convergence to equilibrium play. Rather, the theorem says that a Bayesian learning theory can be consistent only if it rules out certain types of strategic uncertainty.

This paper is largely self-contained but the proof of the impossibility theorem employs two intermediate results from Nachbar [22]. The reader is also directed to Nachbar [22] for additional discussion of motivation and interpretation.

2 Basic definitions

2.1 The repeated game form

Let A_i denote the set of *actions* available to player i in the stage game. Let a_i denote an element of A_i. I assume that $|A_i| < \infty$, where $|A_i|$ is the cardinality of A_i. Let $A = A_1 \times A_2$. An element $a = (a_1, a_2) \in A$ is an *action profile*.

Let $\Delta(A_i)$ denote the set of probability mixtures over A_i. $\Delta(A_i)$ can be

[1] Following Böge and Eisele [2], Mertens and Zamir [18], Brandenburger and Dekel [3], one could construct a type space theory in which each player's "super strategy," mapping from types to incomplete information strategies, could be taken to be common knowledge. This representation is essentially without loss of generality provided that one does not also impose a common prior. For a discussion of the difficulties of interpretation that arise in type space theories, see Dekel and Gul [5].

identified with the unit simplex in $\mathbb{R}^{|A_i|}$. An element $\alpha_i \in \Delta(A_i)$ is a *mixed action*. An element $\alpha = (\alpha_1, \alpha_2) \in \Delta(A_1) \times \Delta(A_2)$ is a *mixed action profile*.

An n-period *history*, denoted h, lists the action profiles realized in the first n periods of play. Thus, h is an element of A^n, the n-fold Cartesian product of A. Let h^0 denote the null history, the history that obtains before play begins. Let $\mathscr{H}$ denote the set of all finite histories, including h^0.

A *path of play*, denoted z, is an infinite history, an element of A^∞. Let $\mathscr{Z} = A^\infty$ denote the set of paths of play. z_n will denote the action profile played at date n under the path of play z. $\pi(z,n) \in A^n$ will denote the projection of z onto its first n coordinates, giving the initial n-period history determined by z. $\pi(z,0) = h^0$.

Make $\mathscr{Z}$ measurable by giving it the σ-algebra generated by the cylinders $C(h) \subset \mathscr{Z}$, where $C(h)$ is the set of paths of play with initial segment h. Let $\Delta(\mathscr{Z})$ denote the set of probability measures over $\mathscr{Z}$.

Players at date $n+1$ know the realized n-period history. A repeated game strategy is a complete contingent plan of action: it specifies what i plans to do in period $n+1$, for each possible n period history of the repeated game, for each n. I will adopt the convention that the term *action* refers to the stage game while the term *repeated game strategy* refers to the repeated game. Formally, a *repeated game behavior strategy* for i is a function of the form

$$\sigma_i : \mathscr{H} \to \Delta(A_i).$$

Thus, a repeated game behavior strategy allows i to randomize at each of his information sets. Given a repeated game behavior strategy σ_i and a history h, the probability that player i chooses action a_i in the period following h is $\sigma_i(h)(a_i)$. Let Σ_i denote the set of i's repeated game behavior strategies. Let $\Sigma = \Sigma_1 \times \Sigma_2$. $\sigma = (\sigma_1, \sigma_2) \in \Sigma$ is a repeated game behavior strategy profile.

I will use s_i to denote a repeated game *pure* strategy, which is a repeated game behavior strategy that, for each h, assigns probability 1 to an element of A_i. Following history h, if the repeated game pure strategy s_i assigns probability 1 to a_i then I will write $s_i(h) = a_i$. $S_i \subset \Sigma_i$ will denote the set of player i's repeated game pure strategies. S_i can be identified with $[A_i]^\infty$. Make S_i measurable by giving it the product σ-algebra.

A repeated game behavior strategy profile σ induces a probability measure $\mu_\sigma \in \Delta(\mathscr{Z})$. History h is *reachable* under the repeated game behavior strategy profile σ iff $\mu_\sigma(C(h)) > 0$. History h is reachable under σ_1 iff there exists *some* repeated game pure strategy s_2 for which $\mu_{(\sigma_1, s_2)}(C(h)) > 0$.

2.2 Stage game payoffs

Given a stage game form, stage game payoffs for player i can be represented as a real $|A_1| \times |A_2|$ matrix, which in turn can be represented as an element $\theta_i \in \mathbb{R}^{|A_1||A_2|}$. Payoffs can be normalized in a number of different ways. For present purposes, the most convenient normalization turns out to be to the unit cube centered on the origin. Call the cube Θ. An element $\theta_i \in \Theta$ is a *parameter* for player i. An element $\theta = (\theta_1, \theta_2) \in \Theta \times \Theta$ is a *parameter profile*.

Make Θ measurable by giving it the standard Borel σ-algebra inherited from $= \mathbb{R}^{|A_1||A_2|}$, and make $\Theta \times \Theta$ measurable by giving it the product σ-algebra. Given $\theta_i^* \in \Theta$ and $\lambda \geq 0$, I will denote the closed relative λ neighborhood of θ_i by

$$\bar{N}_\lambda(\theta_i^*) = \{\theta_i \in \Theta : \|\theta_i - \theta_i^*\| \leq \lambda\}.$$

In particular, $\bar{N}_0(\theta_i^*) = \{\theta_i^*\}$. Given a parameter profile $\theta^* = (\theta_1^*, \theta_2^*)$, I will let

$$\bar{N}_\lambda(\theta^*) = \bar{N}_\lambda(\theta_1^*) \times \bar{N}_\lambda(\theta_2^*).$$

Define $u_i : A \times \Theta \to \mathbb{R}$ by setting $u_i(a, \theta_i)$ equal to the component of θ_i corresponding to the action profile a. I will also let u_i denote the mixed extension of u_i; for any θ_i and any mixed action profile $\alpha \in \Delta(A_i) \times \Delta(A_2)$,

$$u_i(\alpha, \theta_i) = \mathbb{E}_\alpha u_i(a, \theta_i),$$

where $\mathbb{E}_\alpha$ denotes expectation with respect to α.

2.3 *Repeated game payoffs*

Consider any parameter $\theta_i \in \Theta$ and any profile σ of repeated game behavior strategies. Player i's expected discounted payoff is

$$V_i(\sigma, \theta_i) = \mathbb{E}_{\mu_\sigma}\left(\sum_{n=1}^{\infty} \delta^{n-1} u_i(z_n, \theta_i)\right),$$

where $\delta \in [0, 1)$ and where $\mathbb{E}_{\mu_\sigma}$ denotes expectation with respect to μ_σ.

2.4 *Continuation games*

An n-period history h defines a *continuation game*, the repeated game starting in period $n + 1$. In the continuation game following h, a repeated game behavior strategy σ_i induces a repeated game continuation behavior strategy σ_{ih} defined by

$$\sigma_{ih}(h') = \sigma_i(h \cdot h')$$

for any history h', where $h \cdot h'$ denotes the concatenation of h and h'. Given a profile $\sigma = (\sigma_1, \sigma_2)$, let $\sigma_h = (\sigma_{1h}, \sigma_{2h})$. The profile σ_h induces the probability measure μ_{σ_h} over the set of continuation paths, which is simply $\mathscr{Z}$.

Given a profile σ of repeated game behavior strategies and a parameter θ_i, the expected continuation payoff to player i following history h is

$$V_i(\sigma_h, \theta_i) = \mathbb{E}_{\mu_{\sigma_h}}\left(\sum_{n=1}^{\infty} \delta^{n-1} u_i(z_n, \theta_i)\right).$$

Note that in this definition payoffs are discounted to the start of the period following h, rather than back to the first period.

2.5 *Incomplete information strategies*

At the start of the repeated game, player i knows his own payoff parameter, θ_i, but not that of his opponent. Recall that S_i is the set of repeated game *pure* strategies. An *incomplete information strategy* for player i is a measurable function

$$\tau_i : \Theta \to S_i.$$

Given θ_i, if $\tau_i(\theta_i) = s_i$ then player i plays

$$\tau_i(\theta_i)(h) = s_i(h)$$

following history h. I will use the term "incomplete information strategy" (rather than simply "strategy") to avoid confusion with repeated game strategies. Let T_i be the set of i's incomplete information strategies and let $T = T_1 \times T_2$. An element $\tau = (\tau_1, \tau_2) \in T$ is an *incomplete information strategy profile*.

I have three reasons for restricting the range of incomplete information strategies to the repeated game pure strategies. First, for repeated games in which stage game payoffs are known, Nachbar [22] shows that an inconsistency obtains even if players can execute nonpure repeated game behavior strategies. That reasoning extends to the present setting. Second, one of the main motivations for the present paper is to consider whether the inconsistency obtains when randomization is purified via payoff uncertainty, along the lines of Harsanyi [7]. The restriction to repeated game pure strategies is in this spirit. Third, as noted in a more general context by Aumann [1], there is a measure theoretic subtlety in defining incomplete information strategies that map to nonpure repeated game behavior strategies; the problem is not serious, but I would prefer to avoid it entirely.

2.6 *The probability measure on parameter profiles*

Fix a probability measure ρ on $\Theta \times \Theta$. I will assume that ρ is independent. Let ρ_i denote the marginal corresponding to player i.

2.7 *Beliefs*

From the perspective of player 1, T_2 is the set of possible models of player 2's behavior. Informally (see below), Player 1's prior belief is a probability distribution over these alternative models. As noted in the introduction, most of the literature on learning in games of incomplete information assumes that player 1's belief is degenerate: player 2 is certain that the model is some $\tau_2 \in T_2$. In this paper, in contrast, condition CS, defined in Sect. 4, will force beliefs to be nondegenerate.

The formal definition of a belief is complicated by the fact that, because both Θ and S_i are uncountable, there are problems with defining an appropriate σ-algebra for T_i; see Aumann [1]. Following Milgrom and Roberts [19], I will define a belief about player i to be a probability measure P_i over $\Theta \times S_i$, where $\Theta \times S_i$ is given the product σ-algebra. I will assume that P_2, player 1's belief about player 2, does not vary with player 1's own parameter θ_1, and similarly for player 2. This assumption is consistent with the players believing that their parameters are drawn independently. And I will assume that the marginal of P_i on Θ equals ρ_i. I will refer to (P_1, P_2) as a *belief profile*.

Remark 1. Measure theoretic problems with defining beliefs do not arise if a player is certain that his opponent's incomplete information strategy will come from a countable set $\hat{T}_i \subset T_i$. In this case, the belief can be represented as the product of ρ_i and a probability over the discrete space $\hat{T}_i$. But I will not confine the discussion to this case. □

2.8 *Reduced forms*

Given ρ_1, a profile (τ_1, s_2) consisting of an incomplete information strategy τ_1 and a repeated game pure strategy s_2 induces a probability measure $\mu_{(\tau_1, s_2)}$ over paths of play; that is, $\mu_{(\tau_1, s_2)} \in \Delta(\mathscr{Z})$. I will say that σ_1 *is a reduced form of* τ_1 iff, for any s_2,

$$\mu_{(\tau_1, s_2)} = \mu_{(\sigma_1, s_2)}.$$

The definition of reduced form for τ_2 is analogous.

It is not hard to see that a reduced form exists for every τ_i. I have not claimed that the reduced form of τ_i is unique. But any two reduced forms of τ_i, say σ_i and σ_i', will generate the same set of reachable histories and will be identical at every reachable history.

Similarly, any pair (P_1, s_2) induces a probability measure $\mu_{(P_1, s_2)} \in \Delta(\mathscr{Z})$. I will say that σ_1 *is a reduced form of* P_1 iff, for any $s_2 \in S_2$,

$$\mu_{(P_1, s_2)} = \mu_{(\sigma_1, s_2)}.$$

The definition of reduced form for P_2 is analogous. Again, reduced forms exist for any belief but need not be unique.

From a decision theoretic standpoint, it does not matter whether a player (a) has belief P_i about his opponent's incomplete information strategy, yielding a reduced form σ_i, or (b) is certain that his opponent will play the repeated game behavior strategy σ_i. It is often more convenient to work with a reduced form σ_i than with P_i. But when formulating a theory of what beliefs ought to be, it is P_i rather than σ_i that is relevant.

Consider any parameter profile $\theta^* = (\theta_1^*, \theta_2^*)$ and any $\lambda \geq 0$ such that $\rho(\overline{N}_\lambda(\theta^*)) > 0$. I will say that σ_1 is a *reduced form of* τ_1 *on* $\overline{N}_\lambda(\theta_1^*)$ iff, for any $s_2 \in S_2$,

$$\mu_{(\tau_1, s_2 \,|\, \theta_1 \in \overline{N}_\lambda(\theta_1^*))} = \mu_{(\sigma_1, s_2)}.$$

where the notation "$|\theta_1 \in \overline{N}_\lambda(\theta_1^*)$" indicates that μ is computed conditional on $\theta_1 \in \overline{N}_\lambda(\theta_1^*)$. Similarly, σ_1 is a *reduced form of* P_1 *on* $\overline{N}_\lambda(\theta_1^*)$ iff, for any $s_2 \in S_2$,

$$\mu_{(P_1, s_2 \,|\, \theta_1 \in \overline{N}_\lambda(\theta_1^*))} = \mu_{(\sigma_1, s_2)}.$$

The definitions for player 2 are analogous.

2.9 Bayesian updating

As the game proceeds, each player learns by Bayesian updating of his prior. As I will discuss in Sect. 2.10, Bayesian updating will be inherent in optimization, rather than a separate behavioral assumption. Given a prior P_i, any reduced form σ_i, and any history h that is reachable under σ_i, a reduced form of the posterior over incomplete information strategies is simply σ_{ih}. This fact is the primary reason why working with reduced forms is so convenient.

2.10 Optimization

For the moment, fix player 1's parameter, θ_1. Given $\varepsilon \geq 0$, the repeated game behavior strategy $\sigma_1 \in \Sigma_1$ is an *ε best response at* θ_1 (or σ_1 is *ε optimal* at θ_1) iff, for any reduced form σ_2 of player 1's belief,

$$V_1(\sigma_1, \sigma_2, \theta_1) + \varepsilon \geq \max_{s_1 \in S_1} V_1(s_1, \sigma_2, \theta_1).$$

A similar definition holds for player 2. Even if σ_1 is ε optimal at θ_1, the induced repeated game continuation strategy may be ε *sub*optimal at θ_1 in subgames that are far in the future or that the player views as possible but unlikely. Should such a subgame be reached, the player would, presumably, deviate from his repeated game pure strategy. The following stronger version of ε optimization eliminates this problem.

Definition 1. *Fix $\varepsilon \geq 0$, $\theta_1 \in \Theta$, and a belief P_2. The repeated game behavior strategy $\sigma_1 \in \Sigma_1$ is a* uniform ε best response at θ_1 *(or σ_1 is* uniformly ε optimal at θ_1*) iff, for any reduced form σ_2 of P_2 and any h that is reachable under* (σ_1, σ_2),

$$V_1(\sigma_{1h}, \sigma_{2h}, \theta_1) + \varepsilon \geq \max_{s_1 \in S_1} V_1(s_1, \sigma_{2h}, \theta_1).$$

A similar definition holds for player 2.

Definition 2. *Fix a belief P_2. τ_1^* is a* uniform ε best response almost everywhere *(or τ_1^* is* uniformly ε optimal almost everywhere*) iff, for ρ_1 almost every θ_1, $\tau_1^*(\theta_1)$ is a uniform ε best response at θ_1. Write $\tau_1^* \in \mathrm{BR}_1^{\varepsilon}(P_2)$. Similar definitions hold for player 2.*

Remark 2. If $\delta > 0$ then it is easy to see that if s_i is $\varepsilon = 0$ optimal at θ_i then s_i is *uniformly* $\varepsilon = 0$ optimal at θ_i. Bayesian learning is thus built into optimization when $\delta > 0$ and $\varepsilon = 0$. If either $\delta = 0$ (complete myopia) or $\varepsilon > 0$ then uniform ε optimization is a stronger requirement than ε optimization. □

Remark 3. The following argument establishes that, for any belief and any $\varepsilon > 0$, there exists a $\tau_i \in T_i$ that is uniformly ε optimal everywhere (not just almost everywhere). First, for each θ_i, a repeated game pure strategy uniform $\varepsilon = 0$ best response exists for any belief; this is a standard result for discounted repeated games. Subdivide the unit cube Θ into $k^{|A_1||A_2|}$ subcubes, where k is some positive integer. Construct τ_i by selecting, for each subcube, a θ_i in that subcube and setting τ_i to be constant on the interior of that subcube and equal to a repeated game pure strategy best response at θ_i. To each zero dimensional boundary (i.e. a corner) of a subcube, arbitrarily assign a repeated game strategy used in the interior of one of the subcubes that share this boundary. Proceed in this manner to assign repeated game strategies to the relative interiors of boundaries of higher dimension. τ_i is clearly measurable: Θ has been partitioned into a finite number of measurable sets, and τ_i is constant on each element of the partition. Moreover, for any $\varepsilon > 0$, there is a k sufficiently large that the τ_i so constructed is uniformly ε optimal at every θ_i. □

3 Prediction and prototheories

3.1 Prediction

Informally, player 1 learns to predict the path of play iff, eventually, he makes forecasts along the path of play that are almost as accurate as if he knew τ_2. This does *not* mean that player 1 learns τ_2; even if player 1 learns to predict the path of play, his forecast of what player 2's behavior would be off the path of play might be erroneous. And prediction does not mean that player 1 necessarily learns to forecast the path of play generated by the *realized* repeated game pure strategy, $\tau_2(\theta_2)$; see Remark 5, below.

Let $\mathbb{N}$ be the set of natural numbers (including zero) and consider a set $\mathbb{N}^\diamond \subset \mathbb{N}$. Say that $\mathbb{N}^\diamond$ has *density 1* iff

$$\lim_{n\to\infty} \frac{|\{1,2,\ldots,n\} \cap \mathbb{N}^\diamond|}{n} = 1.$$

Recall that $\pi(z,n)$ is the n-period initial segment of the path of play z, and that $C(h)$ is the cylinder of paths of play with initial segment h.

Definition 3. *Fix a belief P_2. Let $\sigma = (\sigma_1, \sigma_2)$ be any profile of repeated game behavior strategies. Player 1* weakly learns to predict the path of play generated by σ *iff, for any σ_2^P that is a reduced form of P_2, the following conditions hold.*

1. $\mu_\sigma(C(h)) > 0$ *implies* $\mu_{(\sigma_1,\sigma_2^P)}(C(h)) > 0$, *for any finite history* h.
2. *For any real number $\eta > 0$ and μ_σ almost every path of play z, there is a set $\mathbb{N}^P(\eta,z) \subset \mathbb{N}$ of density 1 such that, for any $n \in \mathbb{N}^P(\eta,z)$ and any $a_2 \in A_2$, letting $h = \pi(z,n)$,*

$$|\sigma_2(h)(a_2) - \sigma_2^P(h)(a_2)| < \eta.$$

The definition for player 2 is analogous.

Remark 4. Under Definition 3, prediction is weak in that forecasts are required to be accurate only at all but a sparse set of dates, rather than at all dates after some date $\bar{n}$. The reason for adopting the weaker definition stems from results and counterexamples in Lehrer and Smorodinsky [16]; see also the discussion in Nachbar [22]. □

Definition 4. *Fix a belief P_2. Player 1* weakly learns to predict the path of play generated by $\tau = (\tau_1, \tau_2)$ *iff, for any σ_2 that is a reduced from of τ_2 and for ρ_1 almost every $\theta_1 \in \Theta$, player 1 weakly learns to predict the path of play generated by $\sigma = (s_1, \sigma_2)$, where $s_1 = \tau_1(\theta_1)$. The definition for player 2 is analogous.*

Definition 5. *Fix a belief P_2 and a parameter profile $\theta = (\theta_1, \theta_2)$. Player 1* weakly learns to predict the *ex post* path of play generated by $\tau = (\tau_1, \tau_2)$ at θ *iff player 1 weakly learns to predict the path of play generated by $(\tau_1(\theta_1), \tau_2(\theta_2))$. The definition for player 2 is analogous.*

Remark 5. If the ρ_i are absolutely continuous with respect to Lebesgue measure then, even if player 1 weakly learns to predict the path of play generated by τ, player 1 may not weakly learn to predict the *ex post* path of play at ρ almost every θ. But prediction will imply *ex post* prediction at θ if $\rho(\theta) > 0$, or if the τ_i are constant on a neighborhood of θ. □

Remark 6. Foster and Young [6], building on Jordan [11, 13], establishes a conflict between *ex post* prediction and uniform $\varepsilon = 0$ optimization when the ρ_i are absolutely continuous with respect to Lebesgue measure and when the support of ρ is a neighborhood of a game, like matching pennies (exhibited in Fig. 2 in Sect. 4.1), where the unique equilibrium is in mixed actions. This particular conflict does not arise if (a) *ex post* prediction is weakened to prediction in the sense of Definition 4, or (b) players only ε optimize, or (c) the support of ρ is a neighborhood of a game with an equilibrium in pure actions. □

3.2 Prototheories

Definition 6. *P_2, player 1's belief about player 2, is* nondogmatic *with respect to a product set $\hat{T} = \hat{T}_1 \times \hat{T}_2 \subset T$ iff, for any $\tau \in \hat{T}$, player 1 weakly learns to predict the path of play generated by τ. An analogous definition holds for player 2's belief about player 1.*

Informally, if P_i is nondogmatic with respect to $\hat{T}$ then P_i does not rule out any incomplete information strategy in $\hat{T}_i$.[2]

Definition 7. *A* prototheory *of Bayesian learning is a nonempty product set $\hat{T} = \hat{T}_1 \times \hat{T}_2 \subset T$ such that for each i there is* some *belief P_i that is nondogmatic with respect to $\hat{T}$.*

A prototheory $\hat{T}$ becomes a theory of Bayesian learning once one specifies (a) beliefs that are nondogmatic with respect to $\hat{T}$, (b) an $\varepsilon \geq 0$, and (c) a rule

[2] I have erred on the side of choosing a definition that is, if anything, too weak in order to strengthen the negative results to follow. The weakness here is that player i is required to learn to predict the path of play only if he himself chooses an incomplete information strategy from $\hat{T}_i$. Conceivably, he might fail to learn to predict if he were to choose an incomplete information strategy outside of $\hat{T}_i$.

for selecting from the uniform ε best response correspondence. If $\hat{T}$ is a prototheory then nondogmatic beliefs do in fact exist. $\hat{T}_i$ is to be thought of, very loosely, as the support of the belief P_i.[3] By an argument that parallels one given in Nachbar [22], there is no belief that is nondogmatic with respect to $\hat{T} = T$, except in the trivial case in which the opponent has only a single stage game action. Therefore, a prototheory is necessarily a proper subset of $\hat{T}$.

Remark 7. By the "grain of truth" result of Kalai and Lehrer [14], any $\hat{T}$ that is countable is a prototheory. □

4 CS and consistency

In this section, I will define two properties that I would like prototheories to exhibit: CS and consistency. Theorem 1 in Sect. 5.1 will establish that these properties, although seemingly weak, are incompatible.

4.1 CS

Recall the definition of density 1 in Sect. 3.1. And recall that $\pi(z, n)$ is the history equal to the n-period initial segment of the path of play z.

Definition 8. *Let $\hat{T}$ be a prototheory. Consider $\theta^* = (\theta_1^*, \theta_2^*)$ and $\lambda \geq 0$ such that $\rho(\overline{N}_\lambda(\theta^*)) > 0$. $\hat{T}$* satisfies CS on *$\overline{N}_\lambda(\theta^*)$ iff the following conditions hold.*

1. *There is an $\xi \in (0,1)$ such that, for each i, the following is true. Consider any $\tau_i \in \hat{T}_i$ and any $\sigma_i \in \Sigma_i$ that is a reduced form of τ_i on $\overline{N}_\lambda(\theta_i^*)$. There is a $\tau_i^\xi \in \hat{T}_i$ and an $s_i \in S_i$ such that $\tau_i^\xi(\theta_i) = s_i$ for ρ_i almost every $\theta_i \in \overline{N}_\lambda(\theta_i^*)$ and, for any history h, if $s_i(h) = a_i$ then*

$$\sigma_i(h)(a_i) > \xi.$$

2. *Consider any $\tau_1 \in \hat{T}_1$ such that, for some $s_1 \in S_1$, $\tau_1(\theta_1) = s_1$ for ρ_1 almost every $\theta_1 \in \overline{N}_\lambda(\theta_1^*)$. For any function $\gamma_{12} : A_1 \to A_2$ there is a $\tau_2 \in \hat{T}_2$ and an $s_2 \in S_2$ such that the following is true.*
   (a) *$\tau_2(\theta_2) = s_2$ for ρ_2 almost every $\theta_2 \in \overline{N}_\lambda(\theta_2^*)$.*
   (b) *Let z be the path of play generated by (s_1, s_2). There is a set $\mathbb{N}^\gamma(z) \subset \mathbb{N}$ of density 1 such that for every $n \in \mathbb{N}^\gamma(z)$, letting $h = \pi(z, n)$,*

$$s_2(h) = \gamma_{12}(s_1(h)).$$

   And an analogous statement holds for any $\tau_2 \in \hat{T}_2$ that is constant on $\overline{N}_\lambda(\theta_2^*)$ and any $\gamma_{21} : A_2 \to A_1$.

[3] Because many different incomplete information strategies are asymptotically outcome equivalent, player 1 will typically be able to learn to predict incomplete information strategies outside of $\hat{T}_2$, but this is to be thought of as incidental to player 1's decision making.

If the theory is modified to allow players to play nonpure repeated game behavior strategies at θ for which $\rho(\theta) > 0$ then this definition is a strict generalization of the definition of CS given in Nachbar [22] for repeated games of complete information.

CS condition (1) is motivated by the idea that an incomplete information strategy τ_2 that is sensitive to infinitesimal changes in θ_2 is complicated. In particular, such a τ_2 must read the parameter θ_2 to infinite precision, even if θ_2 is not Turing computable.[4] This paper takes the position that a rational player should be cautious, and that one facet of caution is that if player 1 thinks that it is possible that player 2 might choose τ_2 then player 1 should think that it is also possible that player 2 might choose a variant of τ_2, say τ_2', that, for λ small, is constant on $\overline{N}_\lambda(\theta_2^*)$, even at the expense of some payoff loss (which is presumably compensated for by a reduction in unmodeled complexity costs).

I have four additional remarks. First, CS condition (1) will be satisfied on $\overline{N}_\lambda(\theta^*)$ if, for each τ_i in $\hat{T}_i$, $\hat{T}_i$ contains *one* locally constant variant of τ_i. The condition does not require that $\hat{T}_i$ contain *every* locally constant variant.

Second, one obvious way to construct a variant τ_2' of τ_2 that is constant on $\overline{N}_\lambda(\theta_2^*)$ is as follows. For each $\theta_2 \in \overline{N}_\lambda(\theta_2^*)$, set $\tau_2'(\theta_2)(h)$ equal to the action a_2 for which $\sigma_2(h)$ is maximal, where σ_2 is the reduced form of τ_2 on $\overline{N}_\lambda(\theta_2^*)$. Equivalently, choose the action a_2 for which the corresponding set of parameters in $\overline{N}_\lambda(\theta_2^*)$ has highest ρ_2 measure. (If there is a tie, break the tie arbitrarily.) Under this construction, if $\tau_2'(\theta_2)(h) = a_2$ then $\sigma_i(h)(a_2) \geq 1/|A_2|$. This construction is consistent with CS condition (1) for any $\xi < 1/|A_2|$.

Third, in the above motivation, player 1 was certain as to player 2's belief but believed that player 2 might only ε optimize (because of unmodeled complexity costs). Example 2 will illustrate that one can also motivate CS condition (1) by assuming that player 1 is *certain* that player 2 optimizes perfectly (i.e. $\varepsilon = 0$) but is uncertain as to player 2's belief.

Fourth, the force of CS condition (1) is not to rule *out* purification along the lines of Harsanyi [7] but to rule *in* the possibility that the opponent may not be purifying *perfectly*. This distinction will become clearer when I develop Example 2.

CS condition (2) requires that, even if each player knows that the true parameter profile was drawn from $\overline{N}_\lambda(\theta^*)$, each player is still uncertain, at least initially, as to what repeated game strategy his opponent will implement. CS condition (2) is motivated by considerations of both symmetry and caution.

- Symmetry. Aside from differences in the cardinality and labeling of their action sets, the mechanics of constructing repeated game and incomplete information strategies are fundamentally the same for both players. Symme-

[4] $\theta_2 \in \Theta \subset \mathbb{R}^{|A_1||A_2|}$ is *Turing computable* if it can be calculated to arbitrary precision by a Turing machine, which is a formal model of a digital computation with unbounded memory. The Turing computable parameters are a countable subset of Θ, which is uncountable.

try requires that this similarity be reflected in the prototheory. CS condition (2) captures a weak form of symmetry in the fact that s_2 is constructed from s_1 via the relabeling function γ_{12}.

- Caution. The requirement that CS condition (2) holds for all possible γ_{12} captures a form of strategic uncertainty. CS condition (2) requires that if some τ_2 happens to be included in the prototheory then certain variations on τ_2 must be included as well.

In the context of repeated games of complete information, Nachbar [22] explicitly defined primitive conditions called *weak caution* and *symmetry* and showed that CS is a consequence of these conditions. Weak caution condition (1) is actually identical to CS condition (1). One can formulate exact analogs of weak caution and symmetry in the incomplete information setting, as conditions on incomplete information strategies rather than on repeated game behavior strategies, and one can again derive CS as a consequence. But the analog of weak caution condition (2) may be overly strong in incomplete information settings; I will return to this issue in Example 1. The exact analogs of the other conditions do not appear to be problematic. In this paper I will use CS, which is weaker than weak caution and symmetry, as a primitive, rather than a derived, property.

I have one last remark before proceeding to the examples. The results of this paper will not require that the prototheory satisfy CS on *every* neighborhood of *every* parameter profile but only that it satisfy CS on *some* neighborhoods of *some* parameter profiles. For example, if $\delta = 0$ (players are myopic) then one can argue that CS condition (2) is unreasonable on neighborhoods of games in which some action is strictly dominant, since CS condition (2) is incompatible with mutual knowledge of ε rationality (for ε small) on such neighborhoods. But I would argue that CS condition (2) *is* reasonable on many other neighborhoods, and in particular on neighborhoods of stage games where every action is rationalizable; the examples below will illustrate this.

Example 1. Suppose that the stage game is 2×2 and that $A_i = \{H, T\}$ for each i. Suppose that ρ_1 puts probability 1 on a subset $\hat{\Theta}_1 \subset \Theta$ consisting of just three parameters: $\hat{\Theta}_1 = \{\theta_1^{\mathrm{H}}, \theta_1^{\mathrm{T}}, \theta_1^{\mathrm{C}}\}$. θ_1^{H} gives a payoff of 1 when player 1 plays H and a payoff of -1 when player 1 plays T, regardless of what player 2 plays. Thus, if the parameter is θ_1^{H} then player 1 finds it strictly dominant to play H regardless of δ. θ_1^{T} is analogous, but makes T rather than H strictly dominant. Finally, θ_1^{C} is the parameter corresponding to player 1's payoffs in the coordination stage game given in Fig. 1. The definition of

	H	T
H	$-1, -1$	$1, 1$
T	$1, 1$	$-1, -1$

Fig. 1. A coordination game

$\hat{\Theta}_2 = \{\theta_2^{\mathrm{H}}, \theta_2^{\mathrm{T}}, \theta_2^{\mathrm{C}}\}$ is analogous. Let $\tau^* = (\tau_1^*, \tau_2^*)$ be a Nash equilibrium profile for the incomplete information game. (More precisely, to handle the case $\delta = 0$, let τ_1^* induce a uniform $\varepsilon = 0$ best response to τ_2^* at each element of $\hat{\Theta}_1$, and similarly for player 2.)

If $\hat{T}_i = \{\tau_i^*\}$ for each i, so that the prototheory is just a Nash equilibrium profile of the incomplete information game, then the prototheory effectively assumes that there is no strategic uncertainty at $\theta^{\mathrm{C}} = (\theta_1^{\mathrm{C}}, \theta_2^{\mathrm{C}})$, even though the repeated coordination stage game has a continuum of pure equilibria. If CS condition (2) is to hold at θ^{C} then the prototheory must be augmented to incorporate a degree of strategic uncertainty.

To make the example more concrete, suppose that $\rho_i(\theta_i^{\mathrm{H}}) = 2/3$, $\rho_i(\theta_i^{\mathrm{T}}) = 2/9$, and $\rho_i(\theta_i^{\mathrm{C}}) = 1/9$. Suppose further that $\delta = 0$. If θ^{C} is the true state then one can readily verify that, under any equilibrium τ^*, both players play T in the first period, both play H in the second period, and that both are certain from period 3 onwards that θ^{C} is the true parameter profile. To make things as simple as possible, let h^* be the two-period history in which both players play T in the first period and both play H in the second. Consider a Nash equilibrium profile τ^* such that, following any history in which h^* is the initial segment, player 1 plays H and player 2 plays T.

To augment $\hat{T}$ so that CS condition (2) holds at θ^{C}, the simplest fix is to add *one* other incomplete information strategy, call it τ_i, to each $\hat{T}_i$. In the case of player 2, take τ_2 to be identical to τ_2^* except that, at θ_2^{C}, for any history of length two or more with initial segment h^*, τ_2 plays H rather than T. Define τ_1 analogously. One can verify that if the prototheory is augmented to include these τ_i then the augmented prototheory satisfies CS at θ^{C}, for $\lambda = 0$ and with $\mathbb{N}^{\gamma}(z) = \{3, 4, \ldots\}$ for each of the four relevant z (corresponding to the four repeated game profiles possible at θ^{C} under the augmented prototheory). Note that τ_2 is rationalizable; indeed (τ_1, τ_2) is a Nash equilibrium of the incomplete information game.

Two additional remarks. First, the strategic uncertainty required by CS condition (2) is, if anything, too mild. Under the augmented prototheory, players are certain that if the parameter profile is θ^{C} then all strategic uncertainty will be resolved by period four. One could argue that caution should imply that strategic uncertainty be resolved only asymptotically, but CS does not require this.

Second, the augmented prototheory satisfies CS at θ^{C} even though the augmented prototheory contains only *equilibrium* incomplete information strategies. But note that the implied $\mathbb{N}^{\gamma}(z)$ are proper subsets of $\mathbb{N}$. If the definition of CS were strengthened to require $\mathbb{N}^{\gamma}(z) = \mathbb{N}$ then, in this example, there would not exist any prototheory that, using only equilibrium incomplete information strategies, satisfies CS. In contrast, in a complete information setting, even if one requires $\mathbb{N}^{\gamma}(z) = \mathbb{N}$, it is possible to satisfy CS in the θ^{C} game using only equilibrium repeated game strategies. Thus, $\mathbb{N}^{\gamma}(z) = \mathbb{N}$ is a much stronger assumption in incomplete information settings than in complete information settings. It is partly for this reason that I find weak caution condition (2), which implies $\mathbb{N}^{\gamma}(z) = \mathbb{N}$, more compelling in

	H	T
H	$1, -1$	$-1, 1$
T	$-1, 1$	$1, -1$

Fig. 2. Matching pennies

complete information settings than in incomplete information settings. (Weak caution condition (2) was discussed above in connection with CS condition (2).) □

Example 2. Again, suppose that the stage game is 2×2 and that $A_i = \{H, T\}$. Consider matching pennies, given in Fig. 2. Let $\theta^{\mathrm{MP}} = (\theta_1^{\mathrm{MP}}, \theta_2^{\mathrm{MP}})$ be the parameter profile for this stage game. Let ρ be absolutely continuous with respect to Lebesgue measure and let $\overline{N}_\lambda(\theta^{\mathrm{MP}})$ be the smallest neighborhood for which $\rho(\overline{N}_\lambda(\theta^{\mathrm{MP}})) = 1$. Informally, I will assume that $\lambda > 0$ is small, so that all games in $\overline{N}_\lambda(\theta^{\mathrm{MP}})$ are close to matching pennies. Let $\tau^* = (\tau_1^*, \tau_2^*)$ be a Nash equilibrium of the incomplete information game. For λ small, this equilibrium is a purification, along the lines of Harsanyi [7], of the equilibrium of repeated matching pennies. Suppose $\hat{T}$ is given by $\hat{T}_i = \{\tau_i^*\}$ for each i.

$\hat{T}$ violates CS condition (1) on $\overline{N}_\lambda(\theta^{\mathrm{MP}})$ no matter how small one takes $\lambda > 0$. The position taken here is that $\hat{T}$ is not a reasonable prototheory; for λ small, a reasonable prototheory should (as required by CS) include the incomplete information strategies "H always" (the strategy that executes H in every period, regardless of history, for any parameter) and its mirror image "T always."

In defense of this position, note that if player 2 were *certain* that player 1 were playing τ_1^* then the payoff loss from player 2 playing either "H always" or "T always" would be close to zero, for λ small. And executing "H always" or "T always" is simple, whereas executing τ_2^* requires, among other things, that player 2 read the true parameter θ_2 to infinite precision. Moreover, if player 2 is not certain that player 1 is playing τ_2^* then there may be no payoff loss at all from "H always" or "T always," depending on player 2's belief, since, while neither "H always" nor "T always" is part of any Nash equilibrium of the incomplete information game, both *are* rationalizable for λ sufficiently small.

One can verify that if "H always" and "T always" are added to the prototheories for both players then the augmented prototheory, with just three incomplete information strategies for each player, satisfies CS. One might argue that this augmented prototheory is still far too thin, but this merely underscores that CS is, if anything, too weak. □

Example 3. Suppose that, for each i, $\hat{T}_i$ consists of the set of strategies that can be implemented by Turing machines (see Footnote 4). A Turing machine that implements τ_i takes as input the parameter θ_i and the history h and produces as output $\tau_i(\theta_i)(h)$. I assume that each of the $|A_1||A_2|$ coordinates of θ_i is encoded as a binary string of length k; typically, any such encoding will only approximate θ_i. For each Turing machine, the precision k is fixed across all parameters, but k may vary from Turing machine to Turing machine. For each i, $\hat{T}_i$ is countable and hence $\hat{T}$ is a prototheory; see Remark 7. It can be verified that $\hat{T}$ satisfies CS at Lebesgue almost every parameter profile in $\Theta \times \Theta$.

One defect of this example is that, because the set of parameters considered by any given τ_i is effectively finite, prediction implies *ex post* prediction at every parameter profile. Because of this, the sort of purification considered in Example 2 will fail asymptotically, causing problems with existence of even ε equilibrium. One solution is to allow players to execute nonpure behavior strategies (e.g. allow players to use Turing machines with access to a finite set of, possibly biased, coin flippers). Another solution is to augment $\hat{T}$ by the inclusion of strategies that take as input an infinite precision description of the parameter, provided this is done in such a way that CS is preserved. For instance, if the prototheory originally considered in Example 2 is augmented by inclusion of the Turing implementable incomplete information strategies, the resulting augmented prototheory satisfies CS. □

4.2 Consistency

Definition 9. *A prototheory $\hat{T}$ is ε* consistent *iff there is a belief P_1, nondogmatic with respect to $\hat{T}$, such that there is an incomplete information strategy in $\hat{T}_2$ that is uniformly ε optimal almost everywhere, that is,* $\mathrm{BR}_2^{\varepsilon}(P_1) \cap \hat{T}_2 \neq \emptyset$*, and similarly* $\mathrm{BR}_1^{\varepsilon}(P_2) \cap \hat{T}_1 \neq \emptyset$.

Definition 10. *$\hat{T}$ is* consistent *iff it is ε consistent for every $\varepsilon > 0$*

Consistent prototheories exist. A trivial example is a Nash equilibrium profile. See Example 1 or Example 2. But, as those examples illustrate, a Nash equilibrium profile is not a satisfactory candidate for the prototheory of a learning theory.

To motivate consistency, consider the following informal introspective argument. Suppose that the prototheory is mutual knowledge, meaning that both players know $\hat{T}$. In particular, each player knows the support of his opponent's belief. Mutual knowledge of the prototheory does not imply mutual knowledge either of beliefs or of rationality. Suppose further that player 1 thinks that it is *possible* that player 2 is rational, and therefore that player 2 chooses an incomplete information strategy that is uniformly ε optimal almost everywhere with respect to some belief that is nondogmatic with respect to $\hat{T}$. Finally, suppose that player 1 thinks that the ε of a rational player 2 might be arbitrarily small. Then the prototheory must be consistent. Conversely, if the prototheory is not consistent but is mutual knowledge then, evidently, at least one of the players is *certain* that the other is not rational.

5 Impossibility

5.1 Statement of the main result

Given $\theta_1 \in \Theta$, an action $a_1 \in A_1$ is *weakly dominant* iff, for any $a_2 \in A_2$,

$$u_1(a_1, a_2, \theta_1) \geq \max_{a'_1 \in A_1} u_1(a'_1, a_2, \theta_1).$$

The definition for player 2 is similar.[5]

Definition 11. NWD *(the No Weak Dominance condition) holds at $\theta \in \Theta \times \Theta$ iff neither player has a weakly dominant action.*

Given a parameter θ_1, Player 1's *minmax* payoff is given by

$$m_1(\theta_1) = \min_{\alpha_2 \in \Delta(A_2)} \max_{\alpha_1 \in \Delta(A_1)} u_1(\alpha_1, \alpha_2, \theta_1).$$

Player 1's *pure action maxmin* payoff is given by

$$M_1(\theta_1) = \max_{a_1 \in A_1} \min_{a_2 \in A_2} u_1(a_1, a_2, \theta_1).$$

The definitions for player 2 are analogous.

Definition 12. MM *(the Maxmin/Minmax condition) holds at $\theta \in \Theta \times \Theta$ iff, for each player i, the pure action maxmin payoff is strictly less than the minmax payoff,*

$$M_i(\theta_i) < m_i(\theta_i).$$

Examples of stage games that satisfy MM are matching pennies, rock-scissors-paper, battle of the sexes, and many coordination games. Note that for these stage games the case for CS is particularly strong.

The main result of this paper is the following theorem, proved in Sect. 5.2 as a corollary of a stronger but more cumbersome result, Theorem 2.

Theorem 1. *Let $\hat{T}$ be a prototheory. Suppose that $\rho(\overline{N}_\lambda(\theta^*)) > 0$ for every $\lambda > 0$.*

1. *Suppose that NWD holds at θ^* and suppose that, for some $\lambda > 0$, $\hat{T}$ satisfies CS on $\overline{N}_\lambda(\theta^*)$. Then there is a $\bar{\delta} \in (0, 1]$ such that, for any $\delta \in [0, \bar{\delta})$, $\hat{T}$ is not consistent.*
2. *Suppose that MM holds at θ^* and suppose that, for some $\lambda > 0$, $\hat{T}$ satisfies CS on $\overline{N}_\lambda(\theta^*)$. Then, for any $\delta \in [0, 1)$, $\hat{T}$ is not consistent.*

If $\rho(\theta^*) > 0$ then these results hold also for $\lambda = 0$.

Referring to the discussion in Sect. 4.2, Theorem 1 implies that if the prototheory is mutual knowledge and if it satisfies CS on any neighborhood of

[5] This definition is somewhat weaker than the standard one in game theory in that I do not require strict inequality for any a_2.

any stage game at which MM holds then each player is evidently *certain* that the other player is not rational.

If $\rho(\theta^*) = 0$, as will be the case if ρ_i is absolutely continuous with respect to Lebesgue measure, then the conclusion of Theorem 1 is somewhat weak in that an incomplete information strategy τ_i that is not uniformly ε optimal almost everywhere could, in principle, still be uniformly ε optimal on a subset of Θ of large ρ_i measure. In Sect. 5.3, I derive a crude upper bound on the ρ_i measure of such sets.

On the other hand, if $\rho(\theta^*) > 0$ then the analysis essentially reduces to that of the complete information case covered by Nachbar [22]. In particular, if MM holds at θ^* and if CS holds at θ^* then, for sufficiently small ε, no $\tau_i \in \hat{T}_i$ will be uniformly ε optimal at θ^*. The ρ_i probability that the repeated game strategy induced by τ_i is uniformly ε optimal is thus at most $1 - \rho_i(\theta_i^*)$, which can be zero or arbitrarily close to zero.

Remark 8. A previous version of this paper, Nachbar [21], established a different but related inconsistency theorem. Suppose that it is mutual knowledge that players choose incomplete information strategies that are uniformly $\varepsilon = 0$ optimal almost everywhere, but suppose that the beliefs themselves are not mutual knowledge. Rather, each player thinks that his opponent might have any belief out of some set of beliefs. If the set of beliefs satisfies a certain property roughly in the same spirit as CS then the theory will be inconsistent in the sense that each player will choose an incomplete information strategy that is not, loosely speaking, in the support of his opponent's belief. □

5.2 Proofs

The following result is the analog of Theorem 2 in Nachbar [22].

Theorem 2. *Let $\hat{T}$ be a prototheory. Suppose that $\rho(\overline{N}_\lambda(\theta^*)) > 0$ for every $\lambda > 0$.*

1. *Suppose that NWD holds at θ^*. Then there is a $\bar{\lambda} > 0$ and a $\bar{\delta} \in (0, 1]$ such that, for any $\delta \in [0, \bar{\delta})$, there is an $\bar{\varepsilon}_\delta > 0$ such that, for any $\lambda \in (0, \bar{\lambda})$ and for any $\varepsilon \in [0, \bar{\varepsilon}_\delta)$, if $\hat{T}$ satisfies CS on $\overline{N}_\lambda(\theta^*)$ then $\hat{T}$ is not ε consistent.*
2. *Suppose that MM holds at θ^*. Then there is a $\bar{\lambda} > 0$ such that, for any $\delta \in [0, 1)$, there is an $\bar{\varepsilon}_\delta > 0$ such that, for any $\lambda \in (0, \bar{\lambda})$ and for any $\varepsilon \in [0, \bar{\varepsilon}_\delta)$, if $\hat{T}$ satisfies CS on $\overline{N}_\lambda(\theta^*)$ then $\hat{T}$ is not ε consistent.*

If $\rho(\theta^*) > 0$ then these results also hold for $\lambda = \bar{\lambda} = 0$.

Assuming for the moment that Theorem 2 is true, Theorem 1 then follows as an easy corollary.

Proof of Theorem 1. Given that $\rho(\overline{N}_\lambda(\theta^*)) > 0$ for every $\lambda > 0$, if $\hat{T}$ satisfies CS on $\overline{N}_\lambda(\theta^*)$ for some $\lambda > 0$ then, in particular, it satisfies CS on $\overline{N}_\lambda(\theta^*)$ for some $\lambda \in (0, \bar{\lambda})$. Theorem 1 then follows from Theorem 2 and the definition of consistency. ■

To prove Theorem 2, I will first prove the following intermediate result.

Lemma 1. *Let $\hat{T}$ be a prototheory. Suppose that $\rho(\overline{N}_\lambda(\theta^*)) > 0$ for every $\lambda > 0$.*

1. *Suppose that NWD holds at θ^*. There is a $\bar{\delta} \in (0, 1]$ and a $\lambda^\diamond > 0$ such that, for any $\delta \in [0, \bar{\delta})$, there is an $\varepsilon_\delta^\diamond$ such that, for any $\lambda \in (0, \lambda^\diamond)$, for any belief profile that is nondogmatic with respect to $\hat{T}$, for any i, for any $\tau_i \in \hat{T}_i$, and for any σ_i that is a reduced form of τ_i on $\overline{N}_\lambda(\theta_i^*)$, if $\hat{T}$ satisfies CS on $\overline{N}_\lambda(\theta^*)$ and if σ_i is uniformly ε optimal at every $\theta_i \in \overline{N}_\lambda(\theta_i^*)$ then $\varepsilon > \varepsilon_\delta^\diamond$.*
2. *Suppose that MM holds at θ^*. There is a $\lambda^\diamond > 0$ such that, for any $\delta \in [0, 1)$, there is an $\varepsilon_\delta^\diamond$ such that, for any $\lambda \in (0, \lambda^\diamond)$, for any belief profile that is nondogmatic with respect to $\hat{T}$, for any i, for any $\tau_i \in \hat{T}_i$, and for any σ_i that is a reduced form of τ_i on $\overline{N}_\lambda(\theta_i^*)$, if $\hat{T}$ satisfies CS on $\overline{N}_\lambda(\theta^*)$ and if σ_i is uniformly ε optimal at every $\theta_i \in \overline{N}_\lambda(\theta_i^*)$ then $\varepsilon > \varepsilon_\delta^\diamond$.*

If $\rho(\theta^*) > 0$ then these results also hold for $\lambda = \lambda^\diamond = 0$.

To prove Lemma 1, I introduce the concept of evil twins for repeated game strategies.

Definition 13. *A repeated game pure strategy $s_2 \in S_2$ is an ε* evil twin *of a repeated game pure strategy $s_1 \in S_1$ at the parameter $\theta_1 \in \Theta$ iff s_1 is not uniformly ε optimal at θ_1 for any belief such that player 1 weakly learns to predict the path of play generated by (s_1, s_2). A similar definition holds for evil twins of player 2's repeated game pure strategies.*

In repeated matching pennies, the stage game for which was exhibited in Fig. 2, an evil twin of s_1 is any best response to s_1. In the repeated coordination game whose stage game is exhibited in Fig. 1, an evil twin of s_1 is the identical twin defined by, for all h, $s_2(h) = s_1(h)$. Lemma 2 will identify more general conditions under which a strategy is an evil twin. Before stating Lemma 2, I must define some notation.

Define $a_2^M : A_1 \times \Theta \to A_2$ by, for any repeated game pure action $a_1 \in A_1$ and any parameter $\theta_1 \in \Theta$,

$$a_2^M(a_1, \theta_1) = \arg \min_{a_2 \in A_2} u_1(a_1, a_2, \theta_1).$$

If the right-hand side is not single valued, arbitrarily pick one of the values to be $a_2^M(a_1, \theta_1)$. The function a_1^M is defined similarly.

Given s_1 and θ_1, define $S_2^M(s_1, \theta_1) \subset S_2$ to be the set consisting of all $s_2 \in S_2$ for which there exists a set $\mathbb{N}^\diamond \subset \mathbb{N}$ of density 1 such that, for all $n \in \mathbb{N}^\diamond$, letting z denote the path of play generated by (s_1, s_2) and letting $h = \pi(z, n)$,

$$s_2(h) = a_2^M(s_1(h), \theta_1).$$

The definition of $S_1^M(s_2, \theta_2)$ is analogous.

Define the function $\tilde{a}_2 : A_1 \times \Theta \to A_2$ by, for any pure action $a_1 \in A_1$ and any parameter $\theta_1 \in \Theta$,

$$\tilde{a}_2(a_1, \theta_1) = \arg \max_{a_2 \in A_2} \left[\max_{a_1' \in A_1} u_1(a_1', a_2, \theta_1) - u_1(a_1, a_2, \theta_1) \right].$$

If the right-hand side is not single valued, arbitrarily pick one of the values to be $\tilde{a}_2(a_1, \theta_1)$. The function $\tilde{a}_1$ is defined similarly. Loosely, $\tilde{a}_2(a_1, \theta_1)$ is the action that gives player 1 maximal incentive *not* to play a_1. $\tilde{a}_2(a_1, \theta_1)$ does not necessarily minimize player 1's payoff from a_1. That is, it is not necessarily true that $\tilde{a}_2(a_1, \theta_1) = a_2^M(a_1, \theta_1)$.

Given s_1 and θ_1, define $\tilde{S}_2(s_1, \theta_1) \subset S_2$ to be the set consisting of all $s_2 \in S_2$ for which there exists a set $\mathbb{N}^\diamond \subset \mathbb{N}$ of density 1 such that, for all $n \in \mathbb{N}^\diamond$, letting z denote the path of play generated by (s_1, s_2) and letting $h = \pi(z, n)$,

$$s_2(h) = \tilde{a}_2(s_1(h), \theta_1)$$

The definition of $\tilde{S}_1(s_2, \theta_2)$ is analogous.

By virtue of the normalization of the parameters to the unit cube, the most that a player can receive in any period is

$$\bar{u} = \max_{a, \theta_i, i} u_i(a, \theta_i) = 1$$

and the least that a player can receive in any period is

$$\underline{u} = \min_{a, \theta_i, i} u_i(a, \theta_i) = -1.$$

This implies that, for any θ_i, any repeated game pure strategy s_i is uniformly

$$\varepsilon_\delta^{\max} = \frac{\bar{u} - \underline{u}}{1 - \delta} = \frac{2}{1 - \delta}$$

optimal.

The statement of Lemma 2 requires some additional definitions. Define $w_1 : \Theta \to \mathbb{R}$ by

$$w_1(\theta_1) = \min_{a_1 \in A_1} \left[\max_{a_1' \in A_1} u_1(a_1', \tilde{a}_2(a_1, \theta_1), \theta_1) - u_1(a_1, \tilde{a}_2(a_1, \theta_1), \theta_1) \right],$$

and similarly for w_2. Define $w : \Theta \times \Theta \to \mathbb{R}$ by

$$w(\theta) = \min_i w_i(\theta_i).$$

Since u_i is continuous in θ_i, w_i is continuous, hence w is continuous. If NWD holds at θ then $w_i(\theta_i) > 0$ for each i, hence $w(\theta) > 0$.

Lemma 2.

1. *Suppose that NWD holds at θ. Set*

$$\delta_\theta = \frac{w(\theta)}{w(\theta) + 2}$$

and

$$\varepsilon_\theta = w(\theta) - \frac{2\delta_\theta}{1 - \delta_\theta}.$$

Consider any $\delta \in [0, \delta_\theta)$ *and any* $\varepsilon \in [0, \varepsilon_\theta)$. *For any* $s_1 \in S_1$ *and any* $s_2 \in \tilde{S}_2(s_1, \theta_1)$, s_2 *is an* ε *evil twin of* s_1 *at* θ_1. *And an analogous statement holds for any* $s_2 \in S_2$.

2. *Suppose that MM holds at* θ. *Set*

$$\varepsilon_\theta = \min_i m_i(\theta_i) - M_i(\theta_i) > 0.$$

Consider any $\delta \in [0, 1)$ *and any* $\varepsilon \in [0, \varepsilon_\theta)$. *For any* $s_1 \in S_1$ *and any* $s_2 \in S_2^M(s_1, \theta_1)$, s_2 *is an* ε *evil twin of* s_1 *at* θ_1. *And an analogous statement holds for any* $s_2 \in S_2$.

Proof. Lemma 4 in Nachbar [22], which is a generalization of Proposition 1 in Nachbar [20], shows that the conclusion of Lemma 2 holds when NWD holds at θ provided

$$\varepsilon < w(\theta) - \frac{2\delta}{1 - \delta}.$$

The right-hand side of this inequality is strictly positive for any $\delta < \omega(\theta)/(\omega(\theta) + 2)$. Hence the inequality holds for any $\delta \in [0, \delta_\theta)$ and any $\varepsilon \in [0, \varepsilon_\theta)$. The argument for MM is analogous. ∎

The next lemma records that the repeated game pure strategies invoked in Lemma 2 will be incorporated into any prototheory that satisfies CS.

Lemma 3. *Consider a neighborhood* $\overline{N}_\lambda(\theta^*)$ *such that* $\rho(\overline{N}_\lambda(\theta^*)) > 0$ *and suppose that* $\hat{T}$ *satisfies CS condition (2) on* $\overline{N}_\lambda(\theta^*)$. *Consider any* $\tau_1 \in \hat{T}_1$ *such that, for some* $s_1 \in S_1$, $\tau_1(\theta_1) = s_1$ *for* ρ_1 *almost every* $\theta_1 \in \overline{N}_\lambda(\theta^*)$. *Consider any* $\theta_1^\circ \in \overline{N}_\lambda(\theta_1^*)$.

1. *There is a* $\tau_2 \in \hat{T}_2$ *and an* $s_2 \in \tilde{S}_2(s_1, \theta_1^\circ)$ *such that* $\tau_2(\theta_2) = s_2$ *for* ρ_2 *almost every* $\theta_2 \in \overline{N}_\lambda(\theta_2^*)$.
2. *There is a* $\tau_2 \in \hat{T}_2$ *and an* $s_2 \in S_2^M(s_1, \theta_1^\circ)$ *such that* $\tau_2(\theta_2) = s_2$ *for* ρ_2 *almost every* $\theta_2 \in \overline{N}_\lambda(\theta_2^*)$.

And similar statements hold for player 2.

Proof. Given τ_1 and s_1 as in the statement of the lemma, define $\gamma_{12} : A_1 \to A_2$ by $\gamma(a_1) = \tilde{a}_2(a_1, \theta_1^\circ)$. In view of the definition of $\tilde{S}_2(s_1, \theta_1^\circ)$, the conclusion then follows from CS condition (2), with $\mathbb{N}^\diamond = \mathbb{N}^\gamma(z)$. The proof for $S_2^M(s_1 \theta_1^\circ)$ is analogous, as is the proof for player 2. ∎

Proof of Lemma 1. Suppose that NWD holds at θ^*. Set $\lambda^\diamond > 0$ such that NWD holds at every $\theta \in \overline{N}_{\lambda^\diamond}(\theta^*)$. Because the function w employed in the statement of Lemma 2 is continuous, one can take $\lambda^\diamond$ sufficiently small that there is a $\bar{\delta} \in (0, 1]$ and an ε' such that the conclusion of Lemma 2 holds for every $\delta \in [0, \bar{\delta})$, every $\varepsilon \in [0, \varepsilon']$, every $\lambda \in (0, \lambda^\diamond)$, and every $\theta \in \overline{N}_{\lambda^\diamond}(\theta^*)$.

Consider any $\delta \in [0, \bar{\delta})$ and set

$$\varepsilon_\delta^\diamond = \varepsilon' \xi (1 - \delta) > 0,$$

where $\xi \in (0, 1)$ is as in the definition of CS condition (1).

Consider any $\lambda \in [0, \lambda^{\diamond})$ and suppose that $\hat{T}$ satisfies CS on $\overline{N}_\lambda(\theta^*)$. Note that, by hypothesis, $\rho(\overline{N}_\lambda(\theta^*)) > 0$ for any $\lambda > 0$. If $\lambda = 0$, one must assume separately that $\rho(\theta^*) > 0$.

Suppose that σ_1 is a reduced form of $\tau_1 \in \hat{T}_1$ on $\overline{N}_\lambda(\theta_1^*)$ and suppose that σ_1 is uniformly ε optimal at every $\theta_1 \in \overline{N}_\lambda(\theta_1^*)$. By CS condition (1), there is a $\tau_1^\xi \in \hat{T}_1$ and a repeated game pure strategy s_1 such that $\tau_1^\xi(\theta_1) = s_1$ for all $\theta_1 \in \overline{N}_\lambda(\theta_1^*)$ and such that for any $a_1 \in A_i$ and any h, if $s_1(h) = a_1$ then $\sigma_1(h)(a_1) > \xi$.

Choose any $\theta_1^\circ \in \overline{N}_\lambda(\theta_1^*)$. Since σ_1 is uniformly ε optimal at every $\theta_1 \in \overline{N}_\lambda(\theta_1^*)$, σ_1 is, in particular, uniformly ε optimal at θ_1°. It follows from Lemma 6 in Nachbar [22] that s_1 is uniformly $\varepsilon/[\xi(1-\delta)]$ optimal at θ_1°.

On the other hand, by Lemma 3 above, there is an incomplete information strategy $\tilde{\tau}_2 \in \hat{T}_2$ such that, for ρ_2 almost every $\theta_2 \in \overline{N}_\lambda(\theta_2^*)$, $\tau_2(\theta_2) = s_2 \in \tilde{S}_2(s_1, \theta_1^\circ)$. Since $\tau_1(\theta_1) = s_1$ for ρ_1 almost every $\theta_1 \in \overline{N}_\lambda(\theta_1^*)$ and $\tau_2(\theta_2) = s_2$ for ρ_2 almost every $\theta_2 \in \overline{N}_\lambda(\theta_2^*)$, since $\rho(\overline{N}_\lambda(\theta^*)) > 0$, and since, by hypothesis, player 1 weakly learns to predict the path of play generated by (τ_1, τ_2), it follows that player 1 weakly learns to predict the path of play generated by (s_1, s_2). Since s_1 is uniformly $\varepsilon/[\xi(1-\delta)]$ optimal at θ_1°, it then follows from Lemma 2 and the construction of ε' that

$$\varepsilon/[\xi(1-\delta)] > \varepsilon',$$

or

$$\varepsilon > \varepsilon'\xi(1-\delta) = \varepsilon_\delta^\diamond,$$

as was to be shown. The proofs for MM and for player 2 are similar. ■

Proof of Theorem 2. Assume that either NWD or MM holds at θ^*. Let $\bar{\delta}$ and $\lambda^\diamond$ be as in the statement of Lemma 1. (If MM holds, set $\bar{\delta} = 1$.) Consider any $\delta \in [0, \bar{\delta})$. Let $\varepsilon_\delta^\diamond$ be as in the statement of Lemma 1. By the proof of Lemma 1, for any θ, $\varepsilon_\delta^\diamond < \varepsilon_\theta < \varepsilon_\delta^{\max}$, where ε_θ is defined in Lemma 2 and $\varepsilon_\delta^{\max}$ is defined just prior to the statement of Lemma 2.

Because u_i is continuous in θ_i, there is a function $e_\delta : [0, \varepsilon_\delta^{\max}] \times \mathbb{R}_+ \to [0, \varepsilon_\delta^{\max}]$ with the following properties.

1. For each i, if s_i is uniformly ε optimal at *any* $\theta_i \in \overline{N}_\lambda(\theta_i^*)$ then s_i is uniformly $e_\delta(\varepsilon, \lambda)$ optimal at *every* $\theta_i \in \overline{N}_\lambda(\theta_i^*)$.
2. e_δ is continuous and nondecreasing in both arguments.
3. $e_\delta(\varepsilon, 0) = \varepsilon$ for any $\varepsilon \in [0, \varepsilon_\delta^{\max}]$.

By properties (2) and (3), there is a $\bar{\lambda} \in (0, \lambda^\diamond]$ such that $e_\delta(0, \lambda) < \varepsilon_\delta^\diamond$ for all $\lambda \in (0, \bar{\lambda}]$. By property (2), there is an $\bar{\varepsilon}_\delta \in (0, \varepsilon_\delta^\diamond]$ such that $e_\delta(\varepsilon, \lambda) < \varepsilon_\delta^\diamond$ for all $\varepsilon \in [0, \bar{\varepsilon}_\delta)$ and all $\lambda \in (0, \bar{\lambda})$.

Now consider any $\tau_i \in \hat{T}_i$ and suppose that τ_i is uniformly ε optimal almost everywhere. Then for ρ_i almost every θ_i', $\tau_i(\theta_i')$ is uniformly ε optimal at θ_i'. Hence, for any $\lambda > 0$, for ρ_i almost every $\theta_i' \in \overline{N}_\lambda(\theta_i^*)$, $\tau_i(\theta_i')$ is uniformly $e_\delta(\varepsilon, \lambda)$ optimal at *every* $\theta_i \in \overline{N}_\lambda(\theta_i^*)$. Therefore, if σ_i is a reduced form of τ_i on $\overline{N}_\lambda(\theta_i^*)$ then σ_i is uniformly $e_\delta(\varepsilon, \lambda)$ optimal at every $\theta_i \in \overline{N}_\lambda(\theta_i^*)$.

Consider any $\lambda \in (0, \overline{\lambda})$ and suppose that $\hat{T}$ satisfies CS on $\overline{N}_\lambda(\theta^*)$. It then follows from Lemma 1 that

$$e_\delta(\varepsilon, \lambda) > \varepsilon_\delta^\diamond,$$

and hence that

$$\varepsilon \geq \overline{\varepsilon}_\delta.$$

Therefore, $\tau_i \in \hat{T}_i$ is not uniformly ε optimal almost everywhere for any $\varepsilon \in [0, \overline{\varepsilon}_\delta)$. Since T_i was arbitrary, $\hat{T}$ is not ε consistent for any $\varepsilon \in [0, \overline{\varepsilon}_\delta)$. ■

5.3 Bounding the probability that $\tau_i(\theta_i)$ is not uniformly ε optimal

Suppose that the conditions of Theorem 1 are satisfied. Specifically, let NWD or MM hold at θ^* and suppose that $\rho(\overline{N}_\lambda(\theta^*)) > 0$ for any $\lambda > 0$. Let $\overline{\delta} \in (0, 1]$ and $\overline{\lambda} > 0$ be as in Theorem 2 (if MM holds set $\overline{\delta} = 1$). Choose any $\delta \in [0, \overline{\delta})$ and let $\varepsilon_\delta^\diamond$ be as in Lemma 1. As noted in the proof of Theorem 2, $\varepsilon_\delta^\diamond < \varepsilon_\delta^{\max}$, where the latter was defined prior to the statement of Lemma 2. Consider any $\lambda \in (0, \overline{\lambda})$ and suppose that $\hat{T}$ satisfies CS on $\overline{N}_\lambda(\theta^*)$. Finally, consider any belief profile that is nondogmatic with respect to $\hat{T}$.

For any ε and any τ_i, define D_i to be the subset of $\overline{N}_\lambda(\theta_i^*)$ on which τ_i is uniformly ε optimal.

$$D_i = \{\theta_i \in \overline{N}_\lambda(\theta_i^*) : \tau_i(\theta_i) \text{ is uniformly } \varepsilon \text{ optimal at } \theta_i\}.$$

D_i is measurable since τ_i is measurable. Let

$$\eta_i = \frac{\rho_i(D_i)}{\rho_i(\overline{N}_\lambda(\theta_i^*))}.$$

η_i is the ρ_i probability, conditional on $\theta_i \in \overline{N}_\lambda(\theta_i^*)$, that $\tau_i(\theta_i)$ is uniformly ε optimal at θ_i.

Let the function e_δ be as in the proof of Theorem 2. For any $\theta_i' \in D_i$, since $\tau_i(\theta_i')$ is uniformly ε optimal at θ_i', $\tau_i(\theta_i')$ is uniformly $e_\delta(\varepsilon, \lambda)$ optimal at *every* $\theta_i \in \overline{N}_\lambda(\theta_i^*)$.

On the other hand, for *every* $\theta_i' \in \overline{N}_\lambda(\theta_i^*)$, $\tau_i(\theta_i')$ is uniformly $\varepsilon_\delta^{\max}$ optimal at every $\theta_i \in \overline{N}_\lambda(\theta_i^*)$.

Let σ_i be a reduced form of τ_i on $\overline{N}_\lambda(\theta_i^*)$. It follows that σ_i is uniformly

$$\eta_i e_\delta(\varepsilon, \lambda) + (1 - \eta_i)\varepsilon_\delta^{\max}$$

optimal at every $\theta_i \in \overline{N}_\lambda(\theta_i^*)$. By Lemma 1

$$\eta_i e_\delta(\varepsilon, \lambda) + (1 - \eta_i)\varepsilon_\delta^{\max} > \varepsilon_\delta^\diamond.$$

For ε and λ sufficiently small, $e_\delta(\varepsilon, \lambda) < \varepsilon_\delta^\diamond < \varepsilon_\delta^{\max}$, and hence

$$\eta_i < \frac{\varepsilon_\delta^{\max} - \varepsilon_\delta^\diamond}{\varepsilon_\delta^{\max} - e_\delta(\varepsilon, \lambda)},$$

which is independent of τ_i. For ε and λ sufficiently small, e_δ is close to zero and so the right-hand side is approximately

$$1 - \frac{\varepsilon_\delta^\diamond}{\varepsilon_\delta^{\max}}.$$

Thus, if ε and λ are small and if $\rho(\overline{N}_\lambda(\theta^*))$ is close to 1 (if players are confident that something like θ^* is the true stage game) then, for any $\tau_i \in \hat{T}$, the ρ_i *unconditional* probability that $\tau_i(\theta_i)$ is uniformly ε optimal is bounded above by, approximately $1 - \varepsilon_\delta^\diamond/\varepsilon_\delta^{\max}$. In particular, for any $\tau_i \in \hat{T}_i$, the ρ_i probability that $\tau_i(\theta_i)$ is ε optimal at θ_i is bounded away from 1.

As already remarked in the discussion following the statement of Theorem 1, if $\rho(\theta^*) > 0$ then, for ε small enough, the unconditional probability that $\tau_i(\theta_i)$ is uniformly ε optimal is bounded above by $1 - \rho_i(\theta_i^*)$, which can be zero or arbitrarily close to zero.

References

[1] Aumann R (1964) Mixed and behavior strategies in infinite extensive games. In: Dresher M, Shapley LS, Tucker AW (eds.) Advances in game theory (Annals of mathematics studies 52) Princeton University Press, Princeton

[2] Böge W, Eisele T (1979) On solutions of bayesian games. Int J Game Theory 8: 193–215

[3] Brandenburger A, Dekel E (1993) Hierarchies of beliefs and common knowledge. J Econ Theory 59: 189–198

[4] Conlon JR (1999) Hope springs eternal: Learning and the stability of cooperation in short horizon repeated games. University of Mississippi

[5] Dekel W, Gul F (1997) Rationality and knowledge in game theory. In: Kreps DM, Wallis KF (eds.) Advances in Economics and Econometrics: Theory and Applications 1. Cambridge University Press, Cambridge

[6] Foster D, Young P (1999) When rational learning fails. Wharton School, University of Pennsylvania

[7] Harsanyi J (1973) Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. Int J Game Theory 2: 1–23

[8] Jackson M, Kalai E (1997) Social learning in recurring games. Games Econ Behav 21: 102–134

[9] Jackson M, Kalai E (1999) False reputation in a society of players. J Econ Theory 88(1): 40–59

[10] Jordan JS (1991) Bayesian learning in normal form games. Games Econ Behav 3: 60–81

[11] Jordan JS (1993) Three problems in learning mixed-strategy Nash equilibria. Games Econ Behav 5: 368–386

[12] Jordan JS (1995) Bayesian learning in repeated games. Games Econ Behav 8: 8–20

[13] Jordan JS (1997) Bayesian learning in games: A non-Bayesian perspective. In: Bicchieri C, Jeffrey R, Skyrms B, (eds.) The dynamics of norms. Cambridge University Press, Cambridge

[14] Kalai E, Lehrer E (1993) Rational learning leads to Nash equilibrium. Econometrica 61: 1019–1045

[15] Koutsougeras LC, Yannelis NC (1994) Convergence and approximation results for non-cooperative Bayesian games: Learning theorems. Econ Theory 4: 843–857

[16] Lehrer E, Smorodinsky E (1996) Compatible measures and merging. Math Oper Res 21: 697–706

[17] Lehrer E, Smorodinsky R (1997) Repeated large games with incomplete information. Games Econ Beh 18: 116–134
[18] Mertens JF, Zamir S (1985) Formulation of Bayesian analysis for games with incomplete information. Int J Game Theory 14: 1–29
[19] Milgrom P, Weber R (1985) Distributional strategies for games with incomplete information. Math Oper Res 10: 619–632
[20] Nachbar JH (1997a) Prediction, optimization, and learning in repeated games. Econometrica 65: 275–309
[21] Nachbar JH (1997b) Evil twins and Bayesian learning in repeated games of incomplete information. Washington University, St. Louis
[22] Nachbar JH (1999) Rational Bayesian learning in repeated games. Washington University, St. Louis
[23] Nyarko Y (1994) Bayesian learning leads to correlated equilibria in normal form games. Econ Theory 4: 821–841
[24] Nyarko Y (1997) Savage-Bayesian agents play a repeated game. In: Bicchieri C, Jeffrey R, Skyrms B, (eds.) The dynamics of norms. Cambridge University Press, Cambridge
[25] Nyarko Y (1998) Bayesian learning and convergence to Nash equilibria without common priors. Econ Theory 11: 643-655

Soc Choice Welfare (2001) 18: 327–386

Social Choice and Welfare

On the axiomatic method and its recent applications to game theory and resource allocation

William Thomson

Department of Economics, Harkness Hall, University of Rochester, Rochester, NY 14627-0156, USA (e-mail: wth2@troi.cc.rochester.edu)

Received: 29 September 1998/Accepted: 29 January 2000

Abstract. This is a study of the axiomatic method and its recent applications to game theory and resource allocation. It begins with a user's guide. This guide first describes the components of an axiomatic study, discusses the logical and conceptual independence of the axioms in a characterization, exposes mistakes that are often made in the formulation of axioms, and emphasizes the importance of seeing each axiomatic study from the perspective of the axiomatic program. It closes with a schematic presentation of this program. The second part of this study discusses the scope of the axiomatic method and briefly presents a number of models where its use have been particularly successful. It presents alternatives to the axiomatic method and answers criticisms often addressed at the axiomatic method. It delimits the scope of the method and illustrates its relevance to the study of resource allocation and the study of strategic interaction. Finally, it provides extensive illustrations of the considerable recent success that the method has met in the study of a number of new models.

1 Introduction

Until recently the axiomatic method[1] had been the primary method of investigation in a few branches of economics and game theory, such as abstract

Support from NSF, under grant SES 9212557, and the comments of Lars Ehlers, Marc Fleurbaey, Toru Hokari, Bettina Klaus, Yves Sprumont, and John Weymark are gratefully acknowledged. I am particularly indebted to Walter Bossert, Laurence Kranich, James Schummer, and two referees for their detailed remarks.

[1] A point of language needs to be clarified at the outset so as to delimit the scope of this essay. The axiomatic method has been used at different levels of formal analysis.

social choice, inequality measurement, and utility theory, but in the last ten to twenty years its use has considerably expanded. This has certainly been the case for two important domains of game theory where it had been applied at the very beginning. One of them is bargaining theory, which is concerned with the selection of a payoff vector from some feasible set (see Thomson 1999a, for a survey). The other is the theory of coalitional games with transferable utility, which deals with the determination of players' rewards as a function of the profitability of the arrangements they can make in groups (see Peleg 1988, for a detailed treatment). More remarkably, a number of models for which the axiomatic method has proved extremely fruitful have recently been identified. Axiomatic studies of these models have shed new light on well-known solutions, and sometimes led to the discovery of new solutions. The models concern the subjects listed below. In each case, I give a few representative references; general presentations of the literature can be found in Moulin (1988, 1995), Young (1994), Fleurbaey (1996), Roemer (1996), and Thomson (1999a,b).

- Apportionment: how should representatives in Congress be allocated to States as a function of their populations, when proportionality is desired but exact proportionality is not possible? (See Balinski and Young 1982, for a comprehensive treatment.)
- Bankruptcy and taxation: how should the liquidation value of a bankrupt firm be divided among its creditors? When money has to be raised to cover the cost of a public project, what fraction of his income should each taxpayer be assessed? (O'Neill 1982; Aumann and Maschler 1985; Chun 1988; Dagan 1996b; see Thomson 1995b, for a survey.)
- Quasi-linear social choice problems: given a finite set of public projects, and assuming that utility can be freely transferred between any two agents at a one-to-one rate, which project should be chosen and what share of the cost (or monetary compensation) should each agent be charged (or receive)? (Moulin 1985a,b; Chun 1986.)
- Fair allocation in economic contexts: the general question is whether efficiency can be reconciled with equity, but equity is a multifaceted concept, and a myriad of specific issues can be raised. Many have now been resolved for a wide range of models. (See the surveys by Moulin 1995; Thomson 1996a,c; and Moulin and Thomson 1997; Kolm 1997.)

I will not discuss its role in ancient mathematics (Euclidean geometry) and modern mathematics (*e.g.* the construction of number systems). Debreu's (1959) subtitle to the *Theory of value*, "An axiomatic analysis of economic equilibrium," reflects his objective of giving equilibrium analysis solid mathematical foundations, and to develop a theory whose internal coherence could be evaluated independently of the (economic) interpretation given to the variables. At a second level, we find the axiomatic foundations of utility theory and individual decision making. I will not discuss these two levels, limiting myself to a third level, which concerns the search for *solutions* to classes of *multi-agent interaction* (formal definitions of these terms appear below).

- Cost allocation: given a list of quantities demanded by a set of agents for a good, and given the cost of producing the good at various levels, how should the cost of satisfying aggregate demand be divided among the agents? (Tauman 1988; Moulin 1996; Moulin and Shenker 1991, 1992, 1994; Kolpin 1994, 1996; Aadland and Kolpin 1998.)
- Coalitional games without transferable utility: given a set of feasible utility vectors for each group or "coalition" of agents, how should agents' payoffs be chosen? (Aumann 1985a; Hart 1985; Peleg 1985; see Peleg 1988, for a survey.)
- Matching: given two sets of agents, each agent in each set being equipped with a preference relation over the members of the other set, how should they be paired? This problem and variants had been the object of a number of strategic analyses (Roth and Sotomayor 1990), but their axiomatic analysis has recently expanded in a variety of new directions (Sasaki and Toda 1992; Sasaki 1995; Kara and Sönmez 1996, 1997; Toda 1991, 1995, 1996; Sönmez 1995, 1999).
- Measurement of the freedom of choice: given two sets of possible choices, when can one say that one set offers greater freedom of choice than the other? This literature, initiated by Pattanaik and Xu (1990), is a recent entry into the field but it is developing fast (Bossert et al. 1994; Klemisch-Ahlert 1993; Kranich and Ok 1994; Puppe 1995; Kranich 1996, 1997; Bossert 1997).
- Equal opportunities: given a group of agents with different talents or handicaps, how should resources be distributed among them? Here, the literature is also very new (Bossert 1994; Fleurbaey 1994, 1995; Iturbe and Nieto 1996; Maniquet 1994; Bossert et al. 1996).
- Allocation by means of lottery mechanisms: given a group of alternatives and a group of agents with von-Neuman preferences over these alternatives and lotteries over these alternatives, what lottery should be selected? A number of models have recently been expanded to accommodate such mechanisms (Bogomolnaia and Moulin 1999; Abdulkadiroglu and Sönmez 1999; Ehlers 1999; Ehlers et al. 1999).

It may be timely to look at these various developments in a unified way and to assess the methodology on which they are based. I have two main goals. The first one is to explain how an axiomatic study should be conducted and, taking a broader view, how the axiomatic program envisioned. The second one is to give an idea of the recent progress that has been permitted by the use of the axiomatic method, in particular with regards to concretely specified models of resource allocation. For that reason, many of the examples that I take to illustrate points of pedagogy belong to this area. I also draw extensively from the theory of cooperative games. I certainly do not attempt to give a complete presentation of the axiomatic literature, and in particular, I take almost no example from the considerable theory of Arrovian social choice. On this subject, a number of other works are available (Sen 1970; Kelly 1978; Fishburn 1987).

This study has grown much beyond what I had planned, and a guide appears necessary. Part I is a users' guide. It is composed of seven sections.

Section 2 introduces the basic notions of a problem and a solution. Section 3 describes the components of an axiomatic study, its starting point, its goals, and the sort of results that we should expect from it. Section 4 discusses the issue of independence of the axioms in a characterization, by which I mean both their logical independence but also their conceptual independence. Section 5 presents typical errors made in the formulation of axioms. Section 6 widens the scope of the discussion and explains how each axiomatic study should be seen within the framework of what I call the axiomatic program. It describes the goals of this program. Section 7 is a schematic summary of Part I.

Part II discusses the scope of the axiomatic method and evaluates it in comparison to other methods. Section 8 presents the alternatives to the axiomatic method and shows their connections to it. Section 9 responds to a number of criticisms that have been raised against the axiomatic method. Section 10 presents and assesses the commonly held position that the scope of the axiomatic method is limited to abstract models, and to cooperative situations. Section 11 discusses the relevance of the axiomatic method to the study of resource allocation. It introduces the distinction between abstract and concrete models and discusses the limitations and the merits of abstract models. Section 12 evaluates the relevance of the axiomatic method to the study of strategic interaction. It points out that the opposition that is often made between the axiomatic and the strategic approaches in game theory is conceptually flawed. Finally, it argues in favor of an integrated approach in which the axiomatic method is given a wider role.

Part I: A user's guide

2 Basic set-up: Problems and solutions

Before describing the axiomatic method, I introduce the basic terminology that I will use, in particular the concepts of "problems" and "solutions".

2.1 Problems

An axiomatic study of multi-person interaction starts with the specification of a class of ***problems***. A problem is given by specifying data pertaining to the alternatives available and data pertaining to the agents (players, consumers, firms, generations ...). Usually included are the preferences of the agents over the alternatives.

Problems can be described in varying degrees of detail. To illustrate the wide range of possibilities, an "Arrovian" social choice problem (Arrow 1963; Sen 1970) simply consists of a usually unstructured set of feasible alternatives, together with the preferences of the agents over this set. Bargaining problems and coalitional games consist only of sets of attainable utility vectors. For normal form games, a set of actions is specified for each agent, along with the utility vector associated with each profile of actions. For extensive form games, sequences of actions are given together with the utility vector associated with each profile of sequences of actions. These models are already more concrete, as they include information on how utilities result from individual choices, even more so of course when a description of the sequential structure of the actions is added. For allocation problems in economic environments, the precise physical structure of alternatives is included. These problems stand at the opposite end of the spectrum from abstract social choice problems.

In what follows, I frequently take them as illustrations and I assume some familiarity with the basic definitions, and if not with all of the axioms that have been considered in their study (a list appears in Subsect. 9.1), with at least the general principles underlying the central axioms, and with the main solutions. The Appendix contains short descriptions of the models.

2.2 Solutions

Given a class of problems, $\mathcal{D}$, a ***solution***[2] ***on*** $\mathcal{D}$ is a correspondence that associates with every $D \in \mathcal{D}$ a non-empty[3] set of alternatives in the feasible set of D.[4] My generic notation is F for solutions, X for the universal space of alternatives to which the alternatives that are feasible for D belong, and $X(D)$ for this set of feasible alternatives of D. Altogether then, a solution is a correspondence $F : \mathcal{D} \to X$ such that $\phi \neq F(D) \subseteq X(D)$. The aim of the investigation is to identify "good" solutions, good in the sense that they provide either an accurate description of the way problems are resolved in the real world, or a recommendation that an impartial arbitrator or judge could or should make.

Solutions are allowed to be multivalued in some models, and required to be *singlevalued* in others. Whether the objective is descriptive or prescriptive, *singlevaluedness* is of course desirable: a solution that makes precise predictions or recommendations is more likely to be useful. However, *singlevaluedness* is often a very strong requirement and for many models, the search has been for multivalued solutions.

In bargaining theory *singlevaluedness* has been imposed in almost all cases. In the theory of coalitional games with transferable utility, a number of *singlevalued* solutions exist but several important ones are multivalued. When utility is not transferable, *singlevaluedness* is very demanding. In the theory of

---

[2] A variety of other terms are used, such as "rule", "mechanism", "solution function", "solution concept", and "correspondence".

[3] The non-emptiness requirement is not universally imposed. Whether it should be is discussed in Subsect. 4.4.

[4] Note that I do not consider here the problem of deriving a ranking of the set of feasible alternatives, the central objective of the Arrovian social choice literature.

resource allocation, multivaluedness is usually permitted. Here too, *single-valuedness* would in general be an unreasonably strong requirement. However, in some special cases, (examples are bankruptcy and taxation models, and one-dimensional models with single-peaked preferences both in the private good case and in the public good case), it is met by a number of interesting solutions.

3 The components of an axiomatic study

An axiomatic study often begins by noting that for a given domain of problems several intuitively appealing solutions exist, and that some means should be found of distinguishing between them. Alternatively, it may start with the observation that there appears to be only one natural candidate solution for the domain, and be motivated by the desire to find out whether other solutions may be available after all. Yet, for other domains, no well-behaved solution is known, and the axiomatic approach is a good way of finally uncovering at least one such solution, or identifying how close solutions can get to meeting various criteria of good behavior. An axiomatic study has the following components:

1. It begins with the specification of a domain of problems, and the formulation of a list of desirable properties of solutions for the domain.
2. It ends with (as complete as possible) descriptions of the families of solutions satisfying various combinations of the properties.

It should also offer

3. An analysis of the logical relations between the properties;
4. a discussion of whether plausible alternative specifications of the domain would affect the conclusions, and if so, how;
5. a discussion of the implications of substituting for the properties natural variants of them.

Studying the logical relations between the axioms is an effective way to assess their relative power. *Understanding the implications of alternative specifications of the domain* is important too since it is frequently the case that other choices could have been made that are almost as natural. The robustness of our conclusions with respect to these choices should be tested. *Formulating and exploring variants of the axioms* is equally useful as it is not rare that the general ideas that inspire them could have been given slightly different and almost as appealing mathematical forms. We need to know the extent to which our conclusions are sensitive to choices between these various forms, given that the differences between them may have limited conceptual significance.

An axiomatic study often results in *characterization theorems*. They are theorems identifying a particular solution or perhaps a family of solutions, as the only solution or family of solutions, satisfying a given list of axioms. A characterization is the most useful if it offers an explicit description of the solution(s); in the case of a family, a formula specifying it as a function of some parameter belonging to a space of small mathematical complexity (say a finite dimensional Euclidean space) is of greatest practical value.[5] The format of a characterization is as follows[6]:

Theorem 1. (Characterization Theorem)*: A solution* $F : \mathscr{D} \to X$ *satisfies axioms* $A_1, \ldots, A_k$ *if and only if it is solution* F^* *(alternatively, if and only if it belongs to the family* $\mathscr{F}^*$*.)*

An axiomatic study may also produce *impossibility theorems*, stating the incompatibility of a certain list of axioms on a certain domain.

3.1 The objective of an axiomatic study should not in general be the characterization of a particular solution

In the previous section, I stated that the objective of an axiomatic study should be to understand and to describe as completely as possible the implications of lists of properties of interest. Instead, authors often start by stating that their objective is to characterize a particular solution. Apart from two classes of exceptions discussed below, I do not consider this to be a legitimate goal.[7] Whatever reasons we have of being interested in a particular solution, and some of them may be quite justified, does not usually make a characterization of the solution a valid objective.

A first reason for such an interest is that the definition of the solution is intuitively appealing. But this does not suffice to warrant the exclusive focus on the solution because there may be other solutions with appealing definitions.

Another reason may be that the solution seems to give the right answers in particular situations about which, once again, intuition appears to be a reliable guide. But here too, other solutions may be equally successful for these examples. Moreover, for us to infer from the examples that good behavior is to be expected from the solution in general, they should be representative of sufficiently wide classes of situations. This observation suggests that the class of situations that each example illustrates be formally identified, that the requirement on a solution that it behave in a desirable way for that class be formulated as an axiom, and that the implications of this axiom be investigated. I will discuss this program in detail in Subsect. 8.2.

---

[5] Of course, it is not up to the investigator whether such a formula exists.

[6] An analogy with particle physics may not be totally out of place. There, the search is for the minimal list of elementary particles in terms of which all other particles can be described. These elementary constituents are the "atoms" of the theory. Similarly, an axiomatic characterization can be seen as the "decomposition" of a solution into elementary properties. One important difference though is that a given solution can sometimes be characterized in several alternative ways.

[7] What motivates the analysis should not in principle affect the analysis itself, but in fact it often does, and a number of errors commonly made can be traced to this unjustified objective.

3.2 The objective of characterizing a particular solution is legitimate in some situations

A first type of exceptions to the principle stated above, that the objective of an axiomatic study should not be the characterization of a particular solution, is when the solution happens to be widely used in practice. A second type is when the solution has played an important role in theoretical literature. We may be able to discover through an axiomatization why the solution has emerged in the real world or in theoretical studies.

1. Important examples of the first type can be found in the contexts of resource allocation and abstract social choice. A primary one is the Walrasian solution. It is quite remarkable that this solution has guided production and allocation decisions in so many different historical contexts, and very natural to infer that it must have special properties that no other solution satisfies: identifying the properties characterizing it becomes a legitimate exercise. (In the last two decades, some answers have been found to this question. Indeed, although its informational merits had been noted and given intuitive descriptions for a number of years, it is only relatively recently that precise notions of informational efficiency have been formulated, and characterizations of the solution on the basis of these properties developed: under certain assumptions it is "best" from that viewpoint; Hurwicz 1977; under related assumptions, it is "uniquely best"; Jordan 1982).[8]

Majority rule and Borda's rule are examples of voting rules that are frequently applied in practice, and again, it is proper to ask: What are the properties that these solutions must enjoy, and others not, that have led to such wide use? (Here too, characterizations due to May 1952; Young 1974; Ching 1995 and others, have thrown considerable light on the issue.)

2. Examples of the second type are formulas or algorithms that are sometimes suggested. We are often drawn to "simple" or "elegant" formulas, or formulas that can be given a simple interpretation. Similarly, certain algorithms or procedures may appeal to our intuition. It is quite justified to be

[8] Clearly, if the objective is to understand what features of the Walrasian solution have made it an almost universal means of exchanging goods, this search for an axiomatization should proceed under an additional constraint, namely that the axioms be pertinent to the "spontaneous" development of institutions. In that respect, explanations based on considerations of informational simplicity are the most likely to be the "right" ones, whereas it is doubtful that the variable population considerations such as *consistency* that recently have led to the Walrasian solution have much relevance (more on this later). Of course, this does not mean that wanting to figure out the implications of *consistency* is not worthwhile, and it is of great interest that the Walrasian solution should have emerged from such considerations as well. To summarize, I would say that wondering whether certain properties of informational simplicity characterize the Walrasian solution is legitimate, but it is the characterization of the class of solutions satisfying *consistency* that we should be after, whether or not the Walrasian solution belongs to it, and not the characterization of this particular solution on the basis of a condition of this kind.

curious about whether the intellectual appeal of a formula or algorithm is due to their embodying properties of general interest.

A solution for coalitional games with transferable utility defined by means of an attractive algorithm is the nucleolus (Schmeidler 1969; Kohlberg 1971). It is mainly this intuitive appeal that had made this solution a frequent point of reference in the game theory literature and it was natural to wonder whether a formal justification for it could be found. Such a justification, based on an idea of *consistency*[9], was eventually discovered for a variant known as the prenucleolus (Sobolev 1975; see below for a discussion).

But note that whether the goal is to understand why a solution is used in practice or where its intellectual appeal resides, if characterizations are possible, it is the properties on which they are based that should take center stage in further research on the subject.

To pursue our last example, the focus of the literature that followed Sobolev's work on the prenucleolus has indeed been on identifying the implications of various *consistency* notions.

When we need to simply understand, perhaps not to characterize, a particular solution, because the solution has already merited our attention by enjoying some central properties and we would like to know more about it, I claim that the axiomatic method can be of great help, and I propose a protocol for its use below.

3.3 The characterization of a unique solution is not necessarily preferable to the characterization of a family of solutions

A characterization theorem has the merit of completely describing the implications of a list of properties, and that is why we should be striving for such results. *Although many authors prefer that a single solution be identified in a characterization, presumably because the class of problems under study has then been given a unique resolution, I will also challenge this view and say that such a characterization is actually not as good news as the characterization of a family of solutions.*

Indeed experience tells us that, more often than we would like, impossibilities are precipitated by relatively short lists of properties. Typically, if we have shown that a certain list of properties are satisfied by an entire family of solutions, we will be eager to take advantage of the opportunity this multiplicity gives us and impose additional requirements. Some of them may be met by several members of the family, and our next task will be to find out exactly which they are. Starting with the property that we consider the most important, we should then identify the subfamily satisfying it. If this subfamily still contains more than one element, we should bring to bear the property that we consider to be the second most important and so on, and we might very well proceed until a single solution remains.

[9] *Consistency* being often mentioned in these pages, we remind the reader that a precise statement of the property is given in Sect. 9.1.

More likely however, since we rarely have in mind a strict priority of properties, the analysis will branch off in several directions, depending on the order in which we impose the additional properties, each branch possibly ending with the characterization of a unique solution. This sort of tree structure of our findings is typical of an axiomatic study. Certainly, at a stage when several solutions are still acceptable, it is natural to want to know if they should really be thought of as equivalent, or whether they can be distinguished on the basis of additional properties of interest. Then, the objective of characterizing the various solutions "from each other" becomes legitimate.

We will probably want to conclude an axiomatic study with characterizations of *particular solutions,* because such theorems indicate that we have then reached the boundary of the feasible. However, the number of these individual characterizations, and therefore the scope of our study will be all the greater if our first findings are characterizations of *families of solutions*, that is, if we are successful in describing the implications of lists of properties that indeed are not strong enough to force uniqueness.

3.4 For practical reasons, the analysis itself may have to begin from solutions

Although properties come first conceptually, it is certainly useful in practice, and in some cases very useful, to have at our disposal several examples of solutions when starting an axiomatic study. In fact, we are more likely to achieve our goal if we have available a wide repertory of them. The examples can be used in assessing the strength of axioms, testing conjectures concerning the compatibility of axioms, and the independence of axioms in characterizations. This issue is discussed next.[10]

4 Independence of axioms in a characterization

Here, I develop the view that the study of the independence of the axioms in a characterization should be part and parcel of the analysis. By the term independence, we usually understand "logical" independence, but I also discuss what can be called the "conceptual" independence of the axioms. I argue that although axioms should be logically and conceptually independent, they should be compatible in their spirit. Finally, I clarify a logical issue concerning the way in which a characterization is affected by expanding or contracting the domain of problems under consideration.

[10] Until recently, it was actually unusual for a new solution to emerge for the first time in an axiomatic study. For most domains, the solutions that had been found the most valuable had been given intuitive definitions first and axiomatic justifications were found later. As the program expands, studies of new models more frequently take axioms as their point of departure, and it is becoming common for solutions to be introduced in the process of such analysis. Also, as existing models are probed more deeply, variants of existing solutions that do not have their simple form have often been uncovered by axiomatic work.

4.1 In a characterization, the axioms should be logically independent

Recall the "if and only if" format of a characterization. The issue of independence pertains to the statement: "If a solution satisfies a certain list of axioms, it is solution F^*." This is the "uniqueness part", the other direction being the "existence part". The axioms are ***independent*** if by deleting any one of them, it is not true that the solution F^* remains the only admissible one. Verifying that the solution identified in the theorem satisfies the axioms is usually easy, principally because the work can be divided into separate steps, one for each of the axioms, whereas the uniqueness part has to do with the way they interact.

4.1.1 A first reason to establish independence of axioms is to ensure that our results are stated in the most general form

The obvious argument in favor of independence concerns the generality of our conclusions: *if one of the axioms is redundant, we widen the scope of the result by deleting it.*

The interest of many researchers in characterizations lies in the mathematical appeal of results "packaged" as "if and only if" theorems. However, we often know more than what such a theorem says. In the course of our analysis, we may have discovered that if some of the axioms were weakened in certain ways, the solution that is characterized would remain the only acceptable one (in other words, we know more than what the uniqueness part says). We may also have learned that the solution actually satisfies stronger versions of some of the axioms. Consequently, the "if and only if" format is a little dangerous: it conceals some of the information that we have uncovered. In particular, it may result in a uniqueness part in which the axioms are not independent.

If we have shown that the uniqueness part holds without a certain axiom, we should write the characterization without it, but remark separately that the solution does satisfy it. If uniqueness does not hold without the axiom but does with a weaker but natural version of it,[11] it is the weaker version that should appear in the characterization and here we should also point out that the solution happens to satisfy the stronger version. If the solution satisfies much stronger versions of the axioms than the ones used in the uniqueness part, we should probably not present our findings as an "if and only if" theorem.

[11] In proofs, we do not need to invoke the axioms in all of the situations to which they apply, but only in selected situations. Therefore the weaker conditions obtained by limiting their scope to these situations will certainly suffice for the uniqueness proof, but working with these conditions will not necessarily give us a "better" theorem. The weaker condition is less natural. For instance, if a requirement of efficiency is imposed for all economies, and in some proof the requirement is invoked for an economy in which agents have Leontieff preferences, the result could be stated with the weaker but artificial requirement that the rule be efficient for Leontieff economies.

4.1.2 A second, practical, reason to establish independence of the axioms is to discover more general results

A *practical reason* for checking independence has to do with research strategy: it is a way of exploring the "neighborhood" of the characterization. The better we know this neighborhood, the more confident we will be about the correctness of our results. This exploration may also help us discover other techniques of proof for the characterization, or simplifications of the proof that we have.

4.1.3 How to establish logical independence of axioms

In order to establish the independence of axiom A_1, say, from the other axioms in Theorem 1, it suffices to exhibit one solution different from F^* and satisfying $A_2, \ldots, A_k$, but not A_1. However, we should not be satisfied with just one or any example of a solution, for several reasons.

1. First, *the examples should be as "natural" as possible*; ideally, they should be solutions that we might have been tempted to use on other grounds, such as solutions that we know enjoy other properties of interest, or solutions that have been the object of particular attention in the literature. Establishing independence in this way will provide a direct explanation of why these potentially worthwhile solutions are disqualified given our objectives.

In the context of bargaining theory, in order to prove that *contraction independence* is independent of *Pareto-optimality*, *symmetry*, and *scale invariance*, four axioms that characterize the Nash solution, it is best to bring up a solution such as the Kalai-Smorodinsky solution, because – this is the lesson that one can draw from the literature – it should probably be thought of as the major competitor to the Nash solution, instead of less prominent solutions or solutions constructed for that specific purpose.

2. *Second, to be really useful, the examples may very well have to satisfy properties that are not given in the original list.* A property that we consider basic may not appear explicitly in the characterization because it is implied by the list of axioms $A_1, \ldots, A_k$ that are the focus of the study, but it may not be implied by the shorter list obtained by dropping A_1. Then, the independence of A_1 from $A_2, \ldots, A_k$ should also be investigated under the additional assumption that the solution satisfies the property.

An example of such a property, for many models, is *continuity*. This property being quite desirable, we will want to know whether A_1 is independent from $A_2, \ldots, A_k$ *together* with *continuity*. If not, A_1 can be replaced by *continuity* in the characterization, and this might in fact be a more interesting uniqueness part (of course we should not forget to note then that the solution satisfies A_1, and perhaps also state the characterization with A_1).

3. Finally, *we should look for as wide a class of counterexamples as possible.* Indeed, we might be able in the process to identify all of the solutions satisfying $A_2, \ldots, A_k$. From the characterization of the class of solutions satisfying $A_2, \ldots, A_k$, it will typically be easy to deduce how the class would be further restricted by adding either A_1, or one of several conditions that are

reasonable alternatives to A_1.[12] The more general characterization is not necessarily the result that we will write up though, since its proof will probably be more complex. If we judge that the cost of the additional technical developments is too high in relation to the increased generality of the theorem, we should retain the simpler and less general result, but inform our readers of what we know, in a remark, a footnote, or an appendix, with a degree of detail that depends on our intended audience.

For example, in the context of bargaining theory, *symmetry* can be shown to be independent of the other three conditions that we listed earlier as characterizing the Nash solution, by simply producing the solution defined by maximizing the product of player 1's utility and the square of player 2's utility. However, the whole class of solutions satisfying these three conditions can essentially be obtained by noting that maximizing *any* product of weighted utilities would also work, and it is much more informative to exhibit this class.[13] The resulting class of "weighted Nash solutions" has indeed been found of great interest in theory and applications.

4.2 In a characterization, the axioms should express conceptually distinct ideas

Although in a given characterization, several axioms may be motivated by the same general principle (such as a principle of fairness, or a principle of incentive-compatibility), *each axiom should preferably embody only one specific aspect of the general idea.*

I write "preferably" because, like most of the other rules formulated here, this recommendation should not be followed too rigidly. I now give three reasons for that.

1. *A first reason for a given axiom to incorporate distinct conceptual considerations is when it has a simple and direct procedural interpretation.*

In bargaining theory, the axiom of *midpoint domination*, which says that the solution outcome should dominate the average of the agents' most preferred alternatives, is an illustration. It does embody partial notions of *efficiency* (since the outcome should be sufficiently close to the boundary of the problem for this domination to be possible), *symmetry* (the function that associates with each problem the point that is to be dominated satisfies *sym-*

[12] It is not entirely true that given any two lists of axioms related by inclusion, characterizing the implications of the shorter list is necessarily more difficult. For instance, the classes of solutions satisfying only *Pareto-optimality*, or only *symmetry*, are of course very simple to describe. It is probably more accurate to say that up to a point, the difficulty increases. Then, it starts decreasing. I am not making a formal point here, but this statement describes fairly accurately most situations with which I have some familiarity. A main reason is that the basic axioms that we tend to impose first are one-problem axioms whose implications are usually much easier to determine than those of "multi-problem" axioms. The distinction is discussed in detail below.

[13] The qualification "essentially" is because when violations of *symmetry* are extreme, certain dictatorial solutions and lexicographic extensions of them are also admissible.

metry), and *scale invariance* (the function is *scale invariant*). However, it implies none of these three axioms.[14] Moreover, it is descriptive of an intuitively appealing scheme that agents often use: the midpoint corresponds to the vector of utility levels that they reach when they randomize with equal probabilities between their preferred outcomes. A closely related example, taken from the theory of coalitional games with transferable utility, is the requirement on a solution that for the two-person case, it coincides with the so-called "standard solution" (Hart and Mas-Colell 1989), the solution that picks the alternative at which the surplus above the individual rationality utility levels is split equally. Again, this requirement embodies partial notions of *efficiency and symmetry*, but it does so in a way that is very intuitive. It too corresponds to the flipping of the coin to which agents often resort in practice.

I will also note two difficulties – and these are the other two reasons to which I alluded above – in following the recommendation not to incorporate in an axiom distinct conceptual considerations. They should warn us against being too dogmatic in putting it in practice:

2. *Our judgment whether a given axiom does mix ideas that would better be kept separate may well depend on the perspective taken.*

In the theory of coalitional games with transferable utility, and for the fixed population models in which it is typically used, the core can certainly be taken as a primitive notion. However, when the scope of the analysis widens so as to permit variations in populations, and axioms are introduced in order to relate the recommendations made by solutions in response to such variations, the core can be decomposed in terms of *individual rationality* and *consistency* (Peleg, 1985, 1986). In the context of resource allocation, the notion of an envy-free allocation is another example that is intuitively appealing from a normative perspective, and it is difficult to conceive of more basic ones from which it could be derived. Yet, when the perspective shifts from uniquely normative considerations and strategic concerns are addressed in addition, no-envy can be derived under very mild domain assumptions from the much more elementary fairness condition of *equal treatment of equals* and the implementability condition of *Maskin-monotonicity* (Geanakoplos and Nalebuff 1988; Moulin 1993a; Fleurbaey and Maniquet 1997).[15] For an example taken from the theory of non-cooperative games, to which I return below, Nash equilibrium can be decomposed – this decomposition is exact – in terms of *individual rationality*, *consistency*, and *converse consistency* (Peleg and Tijs 1996).

3. The final reason is that *in the process of gaining a deeper understanding of a subject, our judgment about possible formal decompositions of an axiom into more elementary ones may change.* As we discover links between notions

[14] A very simple characterization of the Nash solution can be obtained by means of this axiom and Nash's *contraction independence* (Moulin 1988).
[15] This is not an exact decomposition, since these two axioms together only imply no-envy; they are not equivalent to it.

that we previously perceived as distinct, the way in which we partition and structure the "conceptual field" into individual conditions sometimes evolves.

On a variety of domains, *monotonicity* and *consistency* conditions are traditionally thought of as being unrelated, and they are stated separately. However, in some situations such as the allocation of private goods, they can actually be understood as "conditional" versions of a general "replacement principle", a strong requirement of solidarity, which says that a change in the environment in which agents find themselves should affect all of their welfares in the same direction. It pertains to situations in which agents are not "responsible" for the change when it is socially undesirable, nor deserve any "credit" for it when it is socially desirable. If the principle is applied to the departure of some of the agents, the issue is whether they leave empty-handed or with their components of what the solution has assigned to them. When imposed together with *efficiency*, we therefore obtain either a monotonicity condition or a consistency condition. (This point is developed in Thomson 1995a.)

Similarly, we could argue that for the problem of fair division, the standard forms of the monotonicity conditions such as *resource-monotonicity*, which states that an increase in the social endowment, population being kept fixed, should benefit everyone, or *population-monotonicity*, which states that an increase in the population, resources being kept fixed, should penalize every agent initially present, make sense only in the presence of *efficiency*. Since *efficiency* will indeed typically be required, the demand that all "relevant" agents be affected in the same direction if the parameter (resources or population) increases or decreases – this too is a requirement of solidarity – may be judged more natural (Thomson 1995a).

Finally, in private ownership economies, an axiom such as *individual-endowment monotonicity*, which states that if an agent's endowment increases, he should not be made worse off, can be interpreted from the normative viewpoint, as reflecting the desire that the agent should benefit from resources on which we feel that he has legitimate rights, as he may have obtained them through an inheritance or thanks to his hard work. Alternatively, it may be seen from the strategic viewpoint, as providing him the incentive never to destroy the resources he controls, as this would result in a socially inefficient outcome.

4.3 In a characterization, the axioms should be conceptually compatible

Although it is important that axioms be logically independent and that they express distinct ideas, it is equally important that they be conceptually compatible: the intuition underlying the formulation of one axiom should not be violated by the others. This point seems clear enough but nevertheless deserves to be made.

I will give an example from the theory of bargaining that has to do with the joint use of *continuity* and *consistency*. The most commonly used topological notion (Hausdorff topology) in that theory ignores subproblems involving subsets of the players. On the other hand, *consistency* is motivated by the desire to link recommendations across cardinalities, and certain subproblems appear explicitly in its statement. When this condition is imposed, it is therefore natural to use a continuity notion based on a topology that recognizes the importance of subproblems too. (Such a topology is used in Lensberg 1985, and Thomson 1985.)

The position could be adopted that in the formulation of each axiom we should take into account the essential ideas underlying the others. I illustrate the position with several examples, and for the reader who is concerned that its implementation creates a tension with the objective expressed in the previous subsection – I propose a less radical choice.

The first example again has to do with *efficiency* and *symmetry*, two properties that have been imposed together in a wide range of studies. In this application, an extreme form of the position stated in the previous paragraph is that if *efficiency* is imposed, the axiom of *symmetry* should be written so as to apply to problems from which it is only required that their Pareto-optimal boundary be symmetric (as opposed to problems that are fully symmetric). Such a formulation reflects a strong view that *efficiency* should be given precedence. For another illustration of this viewpoint in the context of Arrovian social choice in economic environments, see Donaldson and Weymark (1988). A somewhat more flexible formulation is to require that two problems with the same Pareto set be solved at the same point[16] and to keep the other axioms including *symmetry* in their usual forms.

To take another example, if *individual rationality* is one of the requirements, it may make sense in the formulation of monotonicity conditions to focus on the subset of the feasible set at which the individual rationality conditions are met. Here too, I would suggest instead that an axiom of *independence of non-individually rational alternatives* be used in conjunction with the others – such an axiom has indeed appeared in the literature (Peters 1986).

4.4 Evaluating characterizations by the number of axioms on which they are based

The opinion is sometimes heard that a characterization of a solution or a family of solutions that makes use of "few" axioms is superior to a characterization involving "many" axioms. Before evaluating the validity of this position, which I will challenge, a "counting problem" needs to be confronted.

1. First, some requirements may be incorporated in the definition of what is meant by the term solution, instead of being imposed separately as axioms on solutions. If we believe that certain requirements are minimal, "non-negotiable", whereas our position concerning the others is more flexible, this way of proceeding may seem justified.

A central example here is *non-emptiness*: some authors require solutions to associate with each admissible problem at least one feasible outcome (as I

[16] Such a condition could be called *independence of non-Pareto optimal alternatives.*

have done above), whereas others state *non-emptiness* as an axiom. Other conditions that are often taken as part of the definition of a solution are *efficiency and symmetry*.

The choice to write a given condition as a separate axiom may depend on how restrictive the condition is for the domain under consideration.

For bargaining problems, existence is almost never an issue, whereas for coalitional games without transferable utility, it often is. It is therefore safe to incorporate *non-emptiness* in the definition of a solution to the bargaining problem, and prudent to impose it as an axiom in the study of coalitional games.

However, I believe that even for requirements that we consider basic, the analysis always benefits from including a discussion of the extra freedom gained by deleting or weakening them, and for that reason, it is best to have them listed as separate axioms.

2. A second reason for the counting problem mentioned above is that it is of course always technically feasible to combine several axioms into one. By so doing, we decrease the number of axioms but not the demands on the solutions. I argued earlier that axioms should embody conceptually distinct desiderata, and this difficulty should in principle not occur, but practice is sometimes a different matter. I gave several reasons why in the previous section.

This counting problem being clarified, and contrary to the view stated above, *my position here is that the fact that an entire family of solutions rather than a unique solution has come out of an characterization involving a large number of axioms must be seen as good news, provided, once again, that they are logically independent and express conceptually distinct ideas, as they should.* This is because, for the class of problems under study, a solution or family of solutions exists that is well-behaved from a variety of perspectives.[17]

On the other hand, and to emphasize a position that I expressed earlier, *we should in general be striving for theorems describing the implications of few properties together.* These are better theorems since the implications of additional properties will typically be easily obtained from them as corollaries. In order to take advantage of such theorems, we should of course thoroughly explore the possible derivation of such corollaries. This argument will take its full force below when I discuss the importance of seeing each axiomatic study from the perspective of what I refer to as "the axiomatic program".

4.5 A logical issue: how enlarging or restricting the domain affects a characterization

It is important to understand how a characterization is affected by enlarging or restricting the domain of problems under consideration. Here, I discuss some

[17] I find the argument that a characterization based on fewer axioms is more "elegant" to have no relevance to the program with which I am concerned here.

common misconceptions about this issue. One of them is that the characterization of a given solution on a larger domain is a weaker theorem. Is this a legitimate view?

The first point to make is that it is not actually meaningful to speak of the "same" solution as having been characterized on two different domains. Formally, a solution is a triple consisting of a domain, a range, and "arrows" from every point in the domain to the range. By changing the domain, we change the solution and therefore we cannot characterize the same solution on two distinct domains. What causes much of the confusion here is that we often keep the same name for the mapping when we change the domain, and for a good reason: in most cases the solution is defined by means of the same formula, or the same algorithm, or the same set of equilibrium equations ... on the various domains.

In the theory of resource allocation, we use the phrase of "Walrasian solution" to designate the solution that selects the Walrasian allocations of each admissible economy, whether or not preferences are strictly monotonic or strictly convex and so on. It is certainly meaningful to apply the "Walrasian definition", or the "Walrasian formula", on these various domains.

When we mainly care about "one-problem" properties of solutions we can safely think of a formula or algorithm as defining the "same" solution on various domains. However, as soon as properties involving comparisons of problems are brought in (axioms involving pairs, or triples, or sequences of problems), we risk making logical errors by not keeping in mind that applying the same definition on two different domains produces two different solutions.

To gain further understanding of the issue, think of a solution constructed by "combining" existing solutions as follows: arbitrarily divide the domain into two subdomains, and apply one or the other of two arbitrarily chosen solutions, depending upon which of the subdomains the problem to be solved belongs to.[18]

For instance, on the domain of private good economies, consider the solution that selects the Walrasian allocations when all agents have Cobb-Douglas preferences, and the core otherwise.

We tend to immediately reject such hybrid solutions, but why? Is it because we feel that they are unlikely to meet any criteria of good behavior? Perhaps, but whether this is true really depends on which criteria we have in mind. If we care only about one-problem properties for instance and these properties happen to be met by each of the component solutions, there is nothing wrong with the hybrid solution, except perhaps for the inconvenience of having to check which of the two cases applies. We suspect however that for many other criteria, the hybrid solution would be disqualified. The axiomatic method can help us formally identify what these criteria are.

As a further illustration of the difficulty of deciding what a legitimate solution is, consider a domain of problems involving variable populations, each

---

[18] Such constructions are common in the abstract Arrovian theory of social choice.

economy being obtained by first drawing a finite group of agents from an infinite population of "potential" agents. A solution defined on such a domain associates with each group of agents and each specification of the data describing them (such as their preferences, their endowments, their production skills and so on), a set of allocations. Imagine now a solution constructed by switching back and forth between several existing solutions according to how many agents are involved. Again, our first reaction, when confronted with such a solution, is to reject it as "artificial". In the paragraphs to follow, I will try to find out whether and to what extent this view is valid.

A concrete example, for private good economies, is the solution obtained by selecting the Walrasian allocations when the number of agents is even and the core when it is odd. An objection to this solution is that it is "unnatural" to alternate between Walrasian notions and core notions: we should make up our mind and pick Walrasian allocations for all cardinalities or the core for all cardinalities. This seems convincing enough but what are the formal arguments to support the objection? In what sense does the Walrasian definition for even numbers "go together" or "fit" with the Walrasian definition for odd numbers, or the core for even numbers fit with the core for odd numbers?

In general, what is wrong with going back and forth between different existing notions in defining solutions? A possible answer is that our choice then cannot be described in terms of a single and simple formula. However, "compactness" of a definition does not seem much of an argument in its favor. First, alternating between two notions may not be a major technical complication. Second, and more importantly, arguments of simplicity of definitions should not take precedence over substantive economic considerations like efficiency, fairness, monotonicity, consistency and so on. The simplicity argument is of course not completely irrelevant because solutions passing the single-and-simple-formula test are more likely to satisfy invariance or independence properties of the kind that have played an important role in axiomatic analysis. But if that is the underlying reason, these properties should be formally identified and the analysis should focus on them.

Moreover, the single-and-simple-formula test is not in general well defined because on a certain domain a given solution may be described in several distinct ways, each of which suggesting a different extension to larger domains. For solutions defined on classes of problems that may involve any number of agents, this difficulty often occurs because solutions that are distinct when the number of agents is greater than two may coincide for the two-person case.

To illustrate this point in the context of resource allocation, consider on the one hand the solution that selects the core for all economies, and on the other hand the solution that selects the individually rational and efficient allocations for all economies. These two solutions happen to coincide in the two-agent case, so how is one to say that the extension of what we choose for two-person economies to economies with more agents should be the core or the individual rationality and Pareto solution?

How to extend a certain definition from the two-person case to the general case is in fact an issue that game theorists have had to confront on many occasions. Similar issues have been how to pass from classes of bargaining problems to classes of coalitional games, or from classes of coalitional games with transferable utility to classes of games without transferable utility. For instance, extending the Shapley value (1953) from coalitional games with transferable utility to the non-transferable utility case has been a central issue in the literature. In addition to Shapley 1969's proposal, we now know of several solutions to games without transferable utility that coincide with his 1953 value when restricted to the transferable utility case.

In addition to simplicity, a second argument in favor of using solutions defined by means of a single-and-simple formula is that whatever considerations would lead us to choosing a certain definition to solve problems involving a given number of agents should have led us to choosing the same definition to solve problems involving any other number of agents.

I agree with this view but only in so far as we do make the effort of uncovering what these considerations might be. This is precisely the role of axiomatic analysis to help us in this task, as they are certainly not given to us when we are presented with the definitions.

To the extent that a characterization of a solution holds independently of the number of agents, and many theorems of this kind are available, we may have a reason not to switch formulas as we move across the domain. However, it seems more productive to explicitly address the issue of how components of solutions should be linked across cardinalities. *Consistency* or *population monotonicity* are two such principles that have provided arguments in favor of using the same definition for all cardinalities. But note that *consistency* would not eliminate the solution that selects the core from equal division for two-person economies and the Walrasian allocations from equal division for economies of greater cardinalities. Yet, it eliminates the solution that selects the core from equal division for all cardinalities, a solution that certainly passes the single-and-simple-formula test. I argued earlier that this test is not always well-defined nor necessary; this example shows that it is not sufficient either.

Let us now return to the issue of how the choice of domains affects the generality of a characterization. On the one hand it is sometimes claimed that the result pertaining to the larger domain is stronger. The opposite view, that by enlarging the domain, we facilitate and therefore weaken the uniqueness part of a characterization is also often heard. The argument here is that since there are "more" situations to which the axioms apply, we give them greater power.

To better evaluate these views let us rewrite the Characterization Theorem in the form of two separate lemmas.

Lemma 1. *If a solution* $F : \mathscr{D} \to X$ *satisfies axioms* $A_1 - A_k$, *then it is* F^*.

Lemma 2. *The solution* $F^* : \mathscr{D} \to X$ *satisfies axioms* $A_1 - A_k$.

Suppose that instead we have established the following two lemmas pertaining to a superdomain D' of D and a solution F'^* defined on D' and whose restriction to D is F^* (in practice, the same names might be used to designate both F^* and F'^*):

Lemma 3. *If a solution* $F : \mathscr{D}' \to X$ *satisfies axioms* $A_1 - A_k$*, then it is* F'^*.

Lemma 4. *The solution* $F'^* : \mathscr{D}' \to X$ *satisfies axioms* $A_1 - A_k$.

Although it is clear that Lemma 4 is stronger than Lemma 2 – and to that extent the view that enlarging the domain provides a stronger result has some validity – there is in fact no logical relation between Lemmas 1 and 3. Indeed, in the proof of Lemma 3, it could very well be that in order to conclude that the solution coincides with F on $\mathscr{D}$, we use (and need) the fact that it satisfies the axioms on $\mathscr{D}' \backslash \mathscr{D}$. This is a sense in which working on the larger domain weakens the uniqueness lemma. On the other hand, precisely because the conclusion of Lemma 3 holds on a wider domain than that of Lemma 1, the two lemmas are in fact not comparable.[19]

If the uniqueness result obtained on the larger domain is not logically weaker than its counterpart for the smaller domain, it may of course be more vulnerable to criticism: by working on a larger domain, we increase the chance that situations exist for which the axioms are not as convincing.[20]

5 Common mistakes in the formulation of axioms

Here, I discuss two mistakes commonly made in the formulation of axioms: tailoring them to a particular solution and losing sight of the fact that priority should be given to their economic meaning.

5.1 Axioms tailored to a particular solution and lacking general appeal

A frequent and unfortunate consequence of wanting to arrive at a particular solution, a goal whose legitimacy I questioned above, is *formulating axioms tailored to that solution and lacking general appeal.* (For a discussion of this point in the context of the search for inequality indices, see Foster 1994.) By targeting a solution we could of course be led to the discovery and the formulation of properties it has that are of independent interest, but this is often not what happens. The common outcome is a characterization that simply amounts to restating the definition of the solution in a slightly different form. Of course, having at our disposal several equivalent definitions of a given solution may be useful. However, the axiom being typically satisfied only by

[19] There could be several solutions satisfying the axioms on the larger domain that all coincide on the smaller domain.

[20] Although we should not expect of any axiom that it be equally appealing in all situations in which it applies, it is important however that the proof not rely precisely on applications to situations where the axiom is less desirable, a situation that is unfortunately not uncommon.

the solution that the investigator intended to characterize (the tell-tale sign[21]), the result does not come as much of a surprise.[22]

5.2 Technical axioms

Avoiding technical axioms is generally desirable since what motivates our work are economically meaningful objectives, not mathematical ones. Unfortunately, this is not always completely feasible: sometimes we are able to determine the implications of a condition of primary concern to us only in the presence of several auxiliary conditions, some of which may be of mainly technical interest. Note however that frequently an axiom appears technical at first, but when we look into it a little more closely, we discover that it does have economic content.

For instance, in the study of bargaining problems and coalitional games without transferable utility, smoothness of boundaries, which is one of the restrictions imposed on problems in the formulation of a number of axioms, is often thought of as a technical detail, but in fact it has economic significance. Indeed the rates at which utility can be transferred between players are meaningful information, and the fact that when moving along the boundary of a feasible set, they may suddenly change is quite relevant when selecting a payoff vector. Perhaps an even more striking example is *continuity*. It is now well understood that in intertemporal models, the topologies on which such notions are based can be interpreted in terms of the agents' impatience, an economically meaningful concept (on this point, see Bewley 1972, and Brown and Lewis 1981).

6 Axiomatic studies and the axiomatic "program"

We should not make too much of an axiomatic study in isolation and of the fact that a particular solution has come out as the best behaved from a certain viewpoint. By changing perspectives, some other solution might very well emerge.

6.1 The axiomatic program

That different studies may lead to different solutions has been seen as a difficulty with the axiomatic method, but the opposite would be surprising. In

[21] We should not necessarily worry about this however. For instance, the fact that the Shapley value is essentially the only solution to games in coalition form to have a potential (Hart and Mas-Colell 1989) does not make this characterization a less valuable result. Considerations of potential are so far removed from any previous consideration that had been brought to bear in the study of these games, and the proof so unlike any previous one, that the result is indeed very illuminating.

[22] One could argue that no result that is fully understood is a surprise, but clearly there are degrees to which the conclusion can be guessed from the hypotheses.

fact, the possibility that recommendations conflict should probably be expected, and it should be confronted. Each axiomatic study should be evaluated in the light of other studies, in the wider context of the *axiomatic program*.

The objective of the axiomatic program is to give as detailed as possible a description of the implications of properties of interest, singly or in combinations, and in particular to trace out the boundary that separates combinations of properties that are compatible from combinations of properties that are not.

Characterization theorems are landmarks on the boundary. One additional property is either redundant, or it takes us into the realm of the infeasible.

6.2 Establishing priorities between axioms

When different solutions result from different axiomatic considerations, the axiomatic program is essentially silent on which axiom to emphasize, and therefore on which solution to recommend. Deciding which axioms should be given priority is up to the "consumer" of the theory. No metatheory exists to help us. I will only state the obvious here, and observe that *since many of the critical axioms that are commonly imposed pertain to changes in some parameter entering the description of the problems, the plausibility of these changes should be a primary consideration.*

In stable economic environments, resources are fixed and in the short run, so are populations. Then, "variable resource" and "variable population" axioms are not relevant. On the other hand, if frequent shocks occur in supplies, variable resource axioms may be important. Since in the long run, population is more likely to vary than in the short run, variable population axioms could be considered then. In teams, we do not have to worry about agents' misrepresenting the information they hold privately, but in more competitive situations, "implementability" requirements may be needed.

6.3 Formulating discrete weakenings of axioms

When an axiom of interest is shown to be incompatible with other important axioms, discrete weakenings of it can sometimes be identified and studied.

For the fair division of private goods, the requirement that no agent receives a bundle that dominates commodity by commodity that of any other agent – this condition is known as *no-domination* – is one such example, as a weakening of no-envy.

These weaker versions of the properties that were our starting point may of course not be as universally applicable however, as they are more likely to be domain-specific.

No-domination, as a weakening of no-envy, is meaningful only in situations where the space of alternatives is endowed with an order structure and preferences are monotonic with respect to that order (this is why it is indeed a weakening of no-envy), whereas no-envy is a meaningful condition even when no such structure is present.

6.4 Formulating parameterizations of axioms

Moreover, when a basic axiom is found not to be compatible with others, it is sometimes possible to formulate parameterized versions of it, with the parameter indicating the partial "degree" to which the axiom is satisfied. Then, we can attempt to identify the range of values of the parameter for which compatibility holds.

An illustration of this approach can be found in a study of the problem of fair division due to Moulin and Thomson (1988). There, the *equal division lower bound* (an allocation meets this bound if every agent finds his bundle at least as desirable as an equal share of the social endowment) is shown to be incompatible with *efficiency* and *resource-monotonicity*. When the *equal division lower bound* is not imposed, a possibility was known to exist, so that the question was open where the line between possibilities and impossibilities had to be drawn. To answer it, Moulin and Thomson introduce a parameter in the interval $[0, 1]$ that turns the discrete requirement that the *equal division lower bound* be met into a continuum of "graduated" conditions of increasing restrictiveness: when the parameter is 0, the condition is vacuously satisfied and when it is 1, the condition is the *equal division lower bound* itself. The result is that for all positive values of the parameter, that is, no matter how much one weakens the *equal division lower bound*, the incompatibility with *efficiency* and *resource-monotonicity* persists. Thanks to the parameterization, the possibility was shown to be the rare case, and the impossibility the norm.

6.5 Establishing functional relations between parameterized axioms

It is possible to go further. *When several properties are given parameterized forms, it becomes in principle possible to describe the tradeoffs between them by means of a functional relation between the parameters. Then the identification of this relation becomes a natural next step in our research program.* A concern for several properties that are incompatible when imposed in full can be partially accommodated by an appropriate selection of the parameters. Instead of having to give up one or the other, we can decide on the importance we would like to give to each and choose the parameters accordingly.

In a series of papers, Campbell and Kelly (see for instance Campbell and Kelly 1993, 1994a,b), have very completely described tradeoffs between efficiency and equity in the context of abstract social choice, in terms of proportions of profiles for which difficulties occur.

An example for resource allocation is given in Thomson (1987a) where a functional relation is established between a parameter measuring the extent to which a certain distributional requirement is met and another parameter measuring the extent to which *resource-monotonicity* is satisfied.

7 A schematic representation of the objectives of the axiomatic program

Figures 1 and 2, which give schematic representations of the objectives of the axiomatic program, summarize a number of the ideas discussed so far.

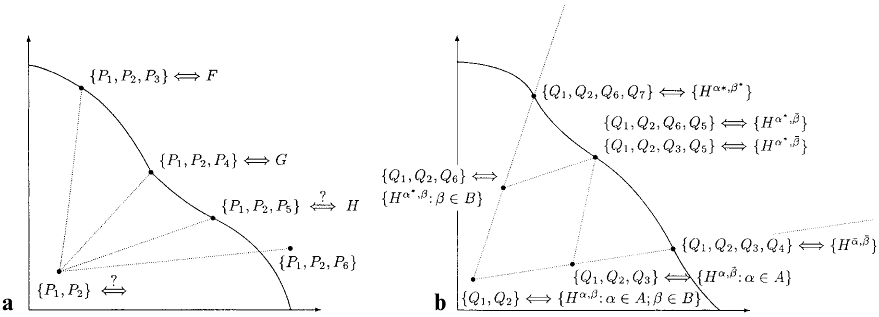

Fig. 1a,b The objectives of the axiomatic program. **(a)** An illustration of the trade-offs between properties P_3 and P_4. In the presence of P_1 and P_2, we cannot have both. **(b)** The scope of a theorem identifying a list of properties that do not force uniqueness, such as the pair $\{Q_1, Q_2\}$, is illustrated by the various corollaries derived from it by imposing additional properties. By adding Q_3, we obtain a one-parameter family, and by adding Q_4, only one member of the family remains acceptable. Alternatively, we could add Q_3 and then Q_4, or Q_6 and then $Q_5 \ldots$

Each point in the plane is interpreted as a combination of properties. The downward sloping line is the boundary between combinations of properties that are compatible and combinations that are not. Think of the northeasterly direction as indicating lists of increasing lengths. Close to the origin are short lists that are likely to be satisfied by large classes of solutions. As we progress in a northeasterly direction, fewer and fewer solutions are acceptable. Eventually, we reach the boundary and the realm of the infeasible. Our goal is to trace out with as much detail as possible this boundary, and for combinations of properties that are compatible, to give complete descriptions of the class of solution(s) satisfying them all. To illustrate notation, a characterization theorem identifying a family of solutions $\{H^\alpha : \alpha \in A\}$ as being the only solutions satisfying axioms P_1 and P_2 is written as "$\{P_1, P_2\} \Leftrightarrow \{H^\alpha : \alpha \in A\}$".

1. Tradeoffs between properties (Fig. 1a). A typical tradeoff between two properties is illustrated by the points $\{P_1, P_2, P_3\} \Leftrightarrow F$ and $\{P_1, P_2, P_4\} \Leftrightarrow G$. They both lie on the boundary and therefore represent combinations of properties that can be met together but in a unique way, by solutions F and G. In the presence of P_1 and P_2, only one of P_3 or P_4 can be met.

We may not have a good understanding of the implications of P_1 and P_2 together, as indicated by the point marked "$\{P_1, P_2\} \overset{?}{\Leftrightarrow}$", but a theorem spelling out the implications of these properties would be very desirable. Most likely, the characterizations of F and G would be obtained as simple corollaries. Also, the implications of alternative properties such as P_5 might be easily obtained (perhaps to give another point of the boundary), and the fact that some other properties, such as P_6, are incompatible with P_1 and P_2 may also come out. This possibility is developed in the next paragraph. I have indicated these potential implications by question marks.

2. The scope of a theorem establishing the characterization of a family of solutions (Figure 1b). Suppose that we have shown that the solutions satisfying

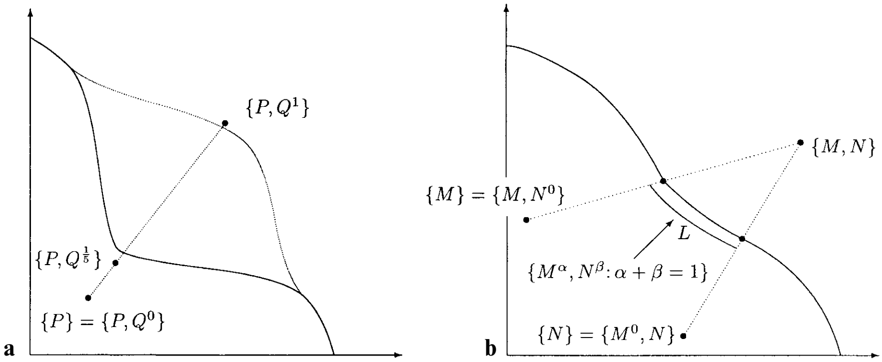

Fig. 2a,b The objectives of the axiomatic program. **(a)** The parameterization of a property may allow us to determine the partial extent to which the property can be satisfied. **(b)** When several properties are parameterized, the trade-offs between them can sometimes be given the form of a functional relation

Q_1 and Q_2 constitute a two-parameter family, a result represented by the point marked "Theorem 1: $\{Q_1, Q_2\} \Leftrightarrow \{H^{\alpha,\beta} : \alpha \in A; \beta \in B\}$." Such a theorem is very useful because from it, we can often quite easily determine the implications of additional properties. By adding Q_3, we reach a smaller family $\{H^{\alpha,\bar{\beta}} : \alpha \in A\}$, and then by adding either Q_4 or Q_5, we reach the boundary, at the points $H^{\bar{\alpha},\bar{\beta}}$ and $H^{\alpha^*,\bar{\beta}}$ respectively. Alternatively, starting from $\{Q_1, Q_2\}$, we could have added Q_6 first, to obtain the family $\{H^{\alpha^*,\beta} : \beta \in B\}$, and then added Q_5 (which perhaps would have taken us back to $H^{\alpha^*,\bar{\beta}}$), and so on. All of these corollaries indicate the "scope" of Theorem 1, which is symbolically indicated by the cone C whose vertex is the point labelled Theorem 1. The cone spans a whole section of the feasible region and of the boundary.[23]

3. Getting close to the boundary (Figure 2a). Suppose that we have established that P can be met but the pair $\{P, Q\}$ cannot, so that the boundary passes between the points $\{P\}$ and $\{P, Q\}$. This raises the question of where exactly it lies. Does it pass "close" to $\{P\}$ (the solid line) or "close" to $\{P, Q\}$ (the dashed line)? Properties are discrete concepts and the question does not seem very meaningful. Yet, it is sometimes possible to formulate parameterized versions of them, with the parameters indicating the partial extent to which they can be satisfied. Suppose that indeed we have a family $\{Q^\lambda : \lambda \in [0, 1]\}$ of graduated conditions of increasing strength such that Q^0 is vacuously satisfied and $Q^1 = Q$. In the figure, we have schematically indicated that only a weak version of Q is compatible with P because the boundary passes close to P.

4. Identifying a functional relation between parameterized axioms permitting to approach the boundary (Figure 2b.) When each of two properties is

[23] Think of it as a cone of light emanating from Theorem 1, its source.

feasible but their combination is not, we can sometimes establish trade-offs between partial, parameterized versions of the properties. Here, the two properties M and N have been parameterized as $\{M^{\alpha} : \alpha \in [0,1]\}$ and $\{N^{\beta} : \beta \in [0,1]\}$. For any pair of values of α and β such that $\alpha + \beta \leq 1$, the properties M^{α} and N^{β} are compatible. This is indicated by the curvilinear segment L, which represents pairs of values of the two parameters permitting compatibility.

Part II: Scope of the axiomatic method, alternatives to it, and recent achievement

In Part II of this essay, I discuss alternative to the axiomatic method, evaluates its scope, and its relevance to the study of allocation problems and strategic interaction.

8 Alternatives to the axiomatic method

So far, I have focused on a presentation of the axiomatic method, without discussing other approaches. *In what follows, I describe these alternatives, and show that not only they are compatible with the axiomatic method, but that in fact, they often naturally lead to it; at the very least, they are very usefully complemented by it.*

8.1 Basing solutions on the "intuitive" appeal of their definitions

For some authors, a solution may be so intuitive that it does not require an axiomatic justification. The position here is that *the appeal of a definition is a substitute for an axiomatic justification.*

For instance, Peleg (1985) opens his study of the *consistency* of solutions to coalitional games, in which he provides the first characterization of the core, by stating that this solution is so natural that there is little need to characterize it.

There should of course be no objection to relying on intuition since intuition underlies the formulation of the axioms too. I submit that the view just expressed is in complete agreement with the position developed in these pages, provided terms are properly defined. Indeed, we have seen a number of axioms that pertain to only one problem at a time in the domain of definition. Let us refer to them as *one-problem axioms*. When such an axiom actually applies to *every* problem in the domain – let us say that it has *full coverage* – it automatically defines a solution.

For most classes of problems, the concept of *Pareto-optimality* can be used either to define an axiom imposed on solutions, or to define a solution, simply the solution that selects for each problem its set of Pareto-optimal outcomes.[24] Similarly, notions such as *individual rationality* and *no-envy* can be used either as axioms or solutions. By contrast, *symmetry* (two agents with identical characteristics should be treated in the same way) is a one-problem axiom that does not have full coverage, since there are problems in which no two agents have identical characteristics. In fact, most problems are of this kind, so that we cannot define a solution on the basis of considerations of symmetry alone.[25]

To the extent that a solution is intended to provide as precise a prediction or recommendation as possible, it may be natural to focus on the axiom interpretation of a test if many alternatives pass it, and on the solution interpretation if the opposite holds.

If this language is adopted, and returning to our earlier examples, *Pareto-optimality* and *individual rationality* would be called axioms because for most economies many allocations pass either test, whereas we would speak of the Walrasian solution since there are typically few Walrasian allocations. The core is somewhere in between; depending upon the model and the number of agents, there may be few core allocations (think of a large exchange economy), or a large set of them (convex games are an example).

Alternatively, we could think of the solution that associates with each problem the set of its feasible outcomes satisfying some basic set of properties as a "presolution", the term suggesting that further restrictions need to be imposed on outcomes.

In the theory of resource allocation, the correspondence that selects for each economy its set of Pareto-optimal allocations, or the correspondence that selects for each economy its set of individually rational allocations, are examples of presolutions. In the theory of coalitional games, the notion of an imputation, an efficient payoff vector meeting the individual rationality constraints, can also be understood, as providing a first reduction of the set of payoff vectors worth considering and we could speak of the "imputation presolution".

8.2 Justifying solutions on the basis of the recommendations they make for test problems

Another approach consists in simply producing solutions, and evaluating them by verifying that they give appropriate answers in situations in which we feel that intuition is a reliable guide. This "direct" approach is the most frequently

[24] This is under the proviso that Pareto-optimal outcomes always exist, since I have required solutions always to be non-empty valued. For most classes of problems – all of the models discussed in this paper are included – the existence of Pareto-optimal outcomes is guaranteed.

[25] Except perhaps in the following trivial way: for each economy to which *symmetry* applies, only select allocations recommended by the axiom; for each other problem, select the whole feasible set, or some arbitrary subset of it.

taken.[26] Here, solutions are assessed by applying them to examples. They are promoted when they provide intuitively correct recommendations or predictions for the examples, and criticized when they do not.

Consider the extension of the Shapley value, known as the λ-transfer value, from coalitional games with transferable utility to games without transferable utility, and to resource allocation problems. (i) It had of course been known for a long time that on the subclass of coalitional games with transferable utility whose core is non-empty, the Shapley value may select payoff vectors outside of the core. Given the compelling definition of the core, this had been seen as a problem. (ii) Examples of games without transferable utility illustrating additional difficulties with the λ-transfer value were developed by Roth (1980). (iii) Shafer (1980) constructed an exchange economy in which the λ-transfer value assigns a positive part of society's resources to an agent whose endowment is zero. The Shafer and Roth examples were then the object of an extensive literature. (See Aumann 1985a; Roth 1986; Scafuri and Yannelis 1984; Yannelis 1982.)

In exchange economies, much has been made of certain "paradoxical" behaviors of the Walrasian solution. For instance, there exist (i) economies in which it allocates all of the gains from trade to only one of the agents; or (ii) economies in which an agent's welfare decreases when his endowment increases; or (iii) economies where an agent's welfare increases when he transfers some of his endowment to another agent whereas the recipient's welfare decreases (this is the well-known "transfer problem"). (iv) The Walrasian solution is also manipulable by misrepresentation of preferences.

Evaluating solutions by means of examples is a useful way to proceed but the lessons to be learned by examining examples are often not drawn with sufficient care. A few examples for which a solution does not make what appears to be the right choice are not a sufficient reason to reject the solution. First, it should not come as a surprise that any given solution would on occasion not make the right recommendation. More importantly, instead of serving as an indictment of the solutions in the study of which they were developed, the examples should instead be used in a constructive way to establish a new vista from which to consider the field. The axiomatic method suggest that the following protocol be set in motion.

1. *We should formally identify the class of situations that the examples illustrate.* The examples will be informative only if they are representative of a sufficiently wide class of cases. *The identification of this class should then inspire the formulation of a general property that can be incorporated as an axiom in the analysis*: the axiom simply specifies how the solution should behave on

[26] This is illustrated by the following list of examples of solutions that were introduced in this way: for bargaining problems, the Raiffa solution (1953); for coalitional games with transferable utility, the core (Gillies 1959); for normal form games, the Nash equilibrium solution (1951); for extensive form games, the subgame perfect equilibrium solution (Selten 1975); for exchange economies, the Walrasian solution; and for economies with single-peaked preferences, the uniform rule (Bennassy 1982).

the class. This process is not meant to be a substitute for intuition – the intuition we have about the examples – but instead as a way of articulating this intuition into operationally useful conditions pertaining to an entire class of cases, the cases illustrated by the examples. The questions can then be asked: How restrictive is the axiom? Which ones of the standard solutions satisfy it? Which other properties is it compatible with? Which combinations of properties is it compatible with? Which maximal combinations of properties is it compatible with?

Possible requirements on a solution suggested by the examples presented above are as follows. (i) In the context of coalitional games, a solution should be a subsolution of the core. (ii) In the context of resource allocation, a solution should not attribute to an agent more of every good that he owned initially; (iii) it should assign to an agent a welfare level that is monotonic with respect to his endowment; (iv) it should be immune to the "transfer problem". (v) In the context of the problem of fair division, a solution should assign to each agent a welfare level that is monotonic with respect to the social endowment. (vi) In the context of a wide variety of resource allocation problems, a solution should be immune to manipulation by misrepresentation of preferences. Of course, none of these requirements should be blindly accepted in all applications; each has its own range of relevance.

How appealing each requirement is will certainly depend on the intended application but in light of the debate that the examples have generated, it is clear that understanding their implications will be of great value. Incidentally, general theorems describing the limited extent to which such requirements are compatible with other appealing ones have now been established, largely exonerating the λ-transfer value and the Walrasian solution from the limitations that the examples had illustrated. These difficulties are now understood to be widely shared, and largely unavoidable on classical domains, although as we will see, quite a few interesting non-classical domains have been identified where they do not occur.

2. *Once the axioms have been formulated, and when the goal is to understand the merits of a particular solution, we can turn to the identification of subdomains of problems on which the solution does provide the right answer.* If it is relatively large, we might be willing to accept undesirable behavior of the solution on the complementary subdomain.[27]

In bargaining theory and in the theory of coalitional games without transferable utility, a number of conditions are satisfied by some of the central solutions under the assumption of strict comprehensiveness of problems[28] but violated if that assumption is not made. Violations only occur on the "boundary" of the domain.

[27] When probabilistic information is available about the likelihood of the various problems in the domain, this information can be used to quantify the severity of the problem.

[28] This is the assumption that the undominated boundary contain no non-degenerate subset parallel to a coordinate subspace.

In economic models of resource allocation, strengthening monotonicity assumptions on preferences has similar consequences: when we go from weakly monotonic preferences to strictly monotonic preferences, we find that a number of properties hold that cannot be satisfied otherwise.

3. *If the subdomain over which the violations of an axiom by a particular solution occur is large enough, we may need to restrict the domain of definition of the solution to the complementary subdomain.*[29]

The Shapley value, when applied on the domain of convex coalitional games with transferable utility, and when used as a solution to resource allocation problems, enjoys properties (core selection, various monotonicities), that it does not satisfy in general (Moulin 1992). In exchange economies, and under the assumption of gross substitutability of preferences, the Walrasian solution satisfies many properties (stability, uniqueness, various monotonicities) that it violates on standard domains (Polterovich and Spivak 1983; Moulin and Thomson 1988). Other restrictions on preferences, such as homotheticity, normality, and quasi-linearity imply better behavior of the Walrasian solution (and others) than on standard domains.

4. *Alternatively, we may keep the same domain of definition for the solution but limit the application of the axiom to a subdomain.*

In formulating the properties of *feasible set monotonicity* and *population-monotonicity* of bargaining solutions, we can restrict attention to strictly comprehensive problems. It is quite useful to know that on this large subdomain, the lexicographic extension of the egalitarian solution satisfies the properties (this is because it coincides there with the egalitarian solution, a solution that enjoys them in general).

5. *Another option is to weaken the conclusion of the axiom, provided we do not lose too much of the essential idea of its initial formulation.*

The egalitarian bargaining solution is not *consistent* but it so happens that the solution outcome of a reduced problem always Pareto-dominates the restriction of the original solution outcome to the subspace pertaining to the agents involved in the reduction, (instead of coinciding with that restriction as required by *consistency*; Thomson 1984). For most problems however, *consistency* and this property are equivalent. Still in bargaining theory, applying the axioms only when certain smoothness conditions are satisfied, when corner situations do not occur, or when the feasible set is strictly comprehensive, are other typical ways in which useful reformulations are obtained. In exchange economies, smoothness of preferences and interiority of allocations often play a role too.

6. *Finally, we may redefine the solution altogether.* Of course, the price of working with such redefinitions may be that some previously satisfied property will now be violated.

[29] Of course, restricting the domain is not always an option. The pathological examples may be ones for which it is particularly important that we be able to make recommendations.

In bargaining theory, the egalitarian solution only satisfies *weak Pareto-optimality*, and in order to obtain *Pareto-optimality*, it can be replaced by its lexicographic extension (Imai 1983).[30] In the process however *continuity* is lost as well as a number of monotonicity properties.

9 Common criticisms addressed at the axiomatic method

The criticism is sometimes levelled against the axiomatic method that the studies that have made use of it too often consist in the formulation of a large number of axioms, and in the analysis of their logical relations – a sterile exercise for some critics – only to end in some impossibility result. Another criticism is that when these studies do not end in impossibilities, the recommendations they make often conflict one with the other. Also, that characterizations are obtained on "too large" a domain. Finally, axioms are often criticized for not being descriptive of behavior. I take up each of these criticisms in turn and draw on the theory of cooperative games and on the theory of resource allocation to show that they are unfounded.

9.1 Too many axioms

Considering first the claimed multitude of axioms, I assert to the contrary that *in spite of the great variety of models that have now been the object of axiomatic analysis, and the apparently large number of axioms that have been used in these analyses, all of these axioms are expressions for each model of just a handful of elementary principles with wide appeal and relevance.* They are the following:

1. Efficiency. The principle of *efficiency*, or *Pareto-optimality* (and weaker versions such as *weak Pareto-optimality* and *unanimity*), is of course the most prominent one.

2. Symmetry. Many studies also involve some form of symmetry requirement. An example is *equal treatment of equals*, which says that identical agents should be treated identically (at each chosen alternative, or globally). A related condition is *anonymity*, which states that the solution should be invariant under "permutations" of agents.

3. Invariance and covariance. *Invariance principles* with respect to certain choices of utility functions play an important role in models where utility information is used (d'Aspremont and Gevers 1977; Sen 1977).

The general principles described next have underlaid a great number of recent developments.

4. Consistency and its converse. The *consistency principle* states the independence of a solution with respect to the departure of some of the agents with their assigned payoffs. It allows us to deduce, from the desirability of an

[30] On the domain of strategic games, either in normal form or in sequential form, Nash equilibrium can be replaced by undominated Nash equilibrium, or subgame perfection respectively. See below for a further discussion.

outcome for some problem faced by some group, the desirability of each restriction of the outcome to each subgroup for the problem obtained by imagining that the members of the complementary subgroup leave with their assigned payoffs and reevaluating the situation from the viewpoint of the remaining agents; these are the associated "reduced problems". The *converse* of this principle permits us to infer the desirability of an outcome for the problem faced by some group from the desirability of the restrictions of the outcome to all two-person subgroups in the associated reduced problems (see Driessen 1991, and Thomson 1996a, for surveys.)

5. Monotonicity. Consider now problems that can be described in terms of a parameter that belongs to a space endowed with an economically meaningful order structure (feasible set in utility space, technological opportunities in commodity space, population size). The *monotonicity principle* requires the welfares of all relevant agents (perhaps the entire set of agents or some particular subset of them) to be affected in a specific direction by changes in parameters that can be evaluated according to that order (see Thomson 1995b, for a survey of the applications of the principle to variations in populations).

6. Replacement. The *replacement principle* asserts that any change in some parameter entering the description of the problem under consideration, whether or not the change can be evaluated in some order, should affect the welfares of all relevant agents (again, who the relevant agents are depends on the application) in the same direction (Thomson 1990a). A primary example of such a parameter is preferences.

Both the *monotonicity* and *replacement* principles are formalizations of the central idea of solidarity, with the latter expressing the strongest demands.[31]

7. Informational simplicity. Principles of *informational simplicity* have also been considered. They express in various ways the idea that solutions should only depend on the essential features of each problem, either to facilitate calculations, or to help guarantee that the agents will have a good understanding of the situation (examples are *contraction independence* of Nash 1950; *local independence* of Nagahisa 1991, 1994, and Nagahisa and Suh 1995; see also Diamantaras 1992). These conditions turn out to have considerable relevance to strategic issues, discussed next.

8. Implementability. Finally, we have principles pertaining to the strategic behavior of the agents. *Strategy-proofness* states that it should always be in an agent's best interest to tell the truth about his characteristics, typically his preferences, but also the resources he controls (endowments of physical goods, knowledge of technologies, of likelihood of uncertain events ...) (see Barberà 1996, for a perspective, and Sprumont 1995, for a survey). *Implementability* says that there should be a game form such that for each economy, the set of equilibrium outcomes of the induced game coincides with the set of outcomes that the solution would have selected on the basis of truthful information (see

[31] In some models, the *monotonicity* and *consistency* principles can actually be seen as "conditional" forms of the *replacement* principle (Thomson 1995b).

Maskin 1985, Postlewaite 1985; Moore 1992, for surveys, and Corchón 1996, for a comprehensive treatment; see also Jackson 1999).

It occasionally takes time to discover that a single principle underlies developments in distinct areas. But once the principle has been recognized and given a general formulation, it can serve as a very useful link across models, providing conceptual unity and common elements of proof techniques.

A striking example illuminating this phenomenon is the *consistency* principle which I have mentioned repeatedly. The principle, which likely underlies a method of adjudicating conflicting claims suggested in the Talmud, a body of Jewish laws and commentaries that is over 2,000 years old (O'Neill 1982; Aumann and Maschler 1985), made a first explicit appearance in early studies of the bargaining problem (Harsanyi 1959) and in the theory of coalitional games with transferable utility (Davis and Maschler 1965). After a twenty-year lull, researchers returned to it, and its implications have now been very fully explored in a wide variety of areas: apportionment (Balinski and Young 1982), coalitional games with transferable utility (Sobolev 1975; Peleg 1986), bargaining (Lensberg 1985, 1988), various models of fair allocation (Tadenuma and Thomson 1991, 1993; Thomson 1988, 1994b), coalitional games without transferable utility (Peleg 1985; Tadenuma 1992), quasi-linear cost allocation (Moulin 1985a; Chun 1986), and bankruptcy and taxation (Young 1987, 1988; Dagan and Volij 1997), each time under a different name[32]. In the late 80's, it was recognized as a general principle, and the terminology settled on *consistency*.

It is true that some minimal adaptation of a general principle to each specific domain is usually necessary, so that the principle ends up giving rise to a constellation of specific properties.

Pursuing the theme of *consistency*, a variety of formulations have been considered depending upon whether the model is discrete, the decision to be made pertains to utility levels or to physical goods, all subgroups or only selected ones are allowed to leave (small groups or groups belonging to a class endowed with a particular structure), the agents who leave are guaranteed the payoffs originally promised to them or payoffs that are only required to be at least as large as these original payoffs.

However, in most cases, this adaptation is a fairly straightforward operation.[33] What is important is to understand the essential logic of, and motivation for, the principle behind its various avatars.

9.2 Too many impossibilities

Turning now to the claim that axiomatic analysis has too often resulted in impossibilities, it too has little merit. First, *impossibilities do not invalidate*

---

[32] The following names have been used: ("uniformity", "stability", "stability under arbitrary formations of subgroups", the "reduced game property", "bilateral equilibrium", "separability").

[33] For instance, a property such as *strategy-proofness* always takes the same form independently of which model is being considered.

axiomatic analysis: they simply reflect mathematical truths that cannot and should not be ignored. Moreover, an impossibility is often a characterization with one axiom too many, and it is a matter of presentation whether the focus is on the characterization or the impossibility. If we have the expectation or the hope that a certain list of desirable properties are compatible but in fact they are not, our conclusion will take the form of an impossibility theorem and the tone will be disappointment. In abstract social choice, this is undoubtedly the conclusion to be drawn from Arrow's work and much of the literature that followed it, a conclusion that may have been at the origin of this criticism.

However, and this is my second response to this criticism, it is now well-understood that the impossibility theorems of Arrovian social choice are mainly due to the analysis being conducted on unstructured domains of alternatives, and to the search being for general methods satisfying a restrictive independence condition. By focusing on concretely specified models and not insisting on the independence condition, a large number of meaningful positive results have now been uncovered, as we will see in the remaining pages of this essay.

9.3 Too many conflicting recommendations

Concerning the claim that when axiomatic analysis has not led to impossibilities, it has too often produced conflicting recommendations, I will first point out that whenever this has been the case, the axiomatic method should not be blamed for results that may not fulfill our hopes. To the contrary, it should be credited for having led to their discovery and thereby helped clarify the relative merits of *a priori* reasonable solutions. Moreover, *for several important domains, just a few solutions have in fact been identified as being clearly more deserving of our attention than other candidates*, as now illustrated:

1. Bargaining problems. I have already noted that in spite of the multiplicity of the solutions that had been proposed for bargaining problems, only three (and natural variants), have come up again and again in the literature. They are Nash's (1950) original solution, the Kalai-Smorodinsky solution (1975), and the egalitarian solution (Kalai 1977). The other solutions have played a role on rare occasions in axiometric analysis, or never. The Nash solution has usually come up in connection with some independence property, and the Kalai-Smorodinsky and egalitarian solutions when some monotonicity property is required. The egalitarian solution requires interpersonal comparisons of utility, and in contexts where for conceptual or practical reasons such comparisons are deemed unacceptable, we are left with just two principal contenders! (see Roth 1979; Peters 1992; Thomson and Lensberg 1989; Thomson 1999a, for surveys of this literature).

2. Coalitional games with transferable utility. Similarly, a great many solutions have been proposed in the theory of coalitional games with transferable utility, but one has been derived in numerous axiomatic analyses, namely the Shapley value (see Aumann 1985b, who emphasizes this point). Together with the core and the nucleolus – the latter has been important in recent developments – we only have three solutions on which to mainly focus. Further relevant criteria to rank them may be existence – recall that non-emptiness of the core is far from being always guaranteed – and *singlevaluedness* – when non-empty, the core often selects multiple allocations.

3. Standard resource allocation. In the study of allocation of private goods, it is also true that no single solution has always been shown superior to the others, but we can with a large degree of confidence eliminate from contention all but a few. The Walrasian solution has come out of axiomatic analyses on several occasions, and the egalitarian-equivalence solution and various selections from it have played an important role in the last few years. The Walrasian solution has been derived primarily from considerations of *informational efficiency* (Hurwicz 1977; Jordan 1982), *implementability* (Hurwicz 1979; Gevers 1986), or *consistency* (Thomson 1988; Thomson and Zhou 1993). Selections from the egalitarian-equivalence solution (Pazner and Schmeidler 1978) have emerged from considerations of *monotonicity*, with respect to endowments or technology (Thomson 1987b; Moulin 1987), or considerations of *welfare domination* pertaining to simultaneous changes in preferences and populations (Sprumont 1998; Sprumont and Zhou 1999).

The final examples pertain to somewhat narrower domains but for them, an even sharper focus on a small number of solutions and sometimes a single solution, has been obtained.

4. Allocation of a private good when preferences are single-peaked. For the allocation of a single infinitely divisible good when preferences are single-peaked, the same solution, the uniform rule, has come up in virtually all cases. Whether *strategy-proofness* (Sprumont 1991; Ching 1992, 1994; Barberà and Jackson 1994), *implementability*, *monotonicity* with respect to *resources* or with respect to *population*, *welfare-domination under preference-replacement*, or *consistency*, are imposed, (Thomson 1990b, 1994a, 1994b, 1995a, 1997; Dagan 1996a; Moreno 1995; Klaus et al. 1997, 1998), the uniform rule has emerged as the most important solution.

5. Auctioning a single indivisible good. For the allocation of a single indivisible good when monetary transfers are possible, the solution that selects for each economy the envy-free allocation at which the winner of the indivisible good is indifferent between his bundle and the common bundle of the losers has come up on several occasions. Considerations of *consistency, population-monotonicity* (Tadenuma and Thomson 1993, 1995), and *welfare-domination under preference-replacement* (Thomson 1998), have all led to that solution.[34]

6. Public choice when preferences are single-peaked. Finally, for the problem of choosing the level of a public good from an interval when preferences are single-peaked, a family of solutions, the generalized Condorcet solutions, and various subfamilies, have been characterized in several ways. Characterizations of these families have been obtained from considerations of *strategy-*

[34] This is the primary solution for this domain. Virtually all other solutions coincide with it.

proofness (Moulin 1980, 1984; Barberà and Jackson 1994; Ching 1997), *consistency* (Moulin 1984), *population-monotonicity* (Ching and Thomson 1992), and *welfare-domination under preference-replacement* (Thomson 1993; Vohra 1999).

These examples certainly do not guarantee that the same phenomenon will always occur but they do show that for several models, some very useful priorities among solutions are obtained by applying the axiomatic method.

Incidentally, note that if the objective of an axiomatic study were taken to be the characterization of a particular solution (a position challenged in Sect. 3), the fact that the solution has been characterized in earlier work might diminish the interest of the result. On the other hand, if we do not lose sight of the objective of the axiomatic *program*, which I have argued should be to identify as completely as possible which combinations of desirable properties are compatible, and how, then the fact that a certain solution comes up once again in a characterization should be celebrated: this may give us the hope that the class of problems under study has only one reasonable solution, or at least only a few such solutions. When it comes to actually making a choice, a consensus will then be much more likely.

9.4 Too large a domain

A concern that is sometimes expressed is that for axioms to be effective in proofs, the domain of problems under consideration has to be "large", even "too large". Characterizations depend too much on solutions being defined for a wide range of problems, including ones that are not likely to occur frequently, or even stand at the limit of what is plausible. Sometimes, crucial steps in proofs are made possible only by drawing on these problems that lie at the "boundary of the domain".

This criticism is unjustified. An axiomatic study properly conceived begins with the proper mathematical specification of the range of economic situations to be covered. If the domain has not been specified correctly, then of course our conclusions will not be useful. Admittedly in practice, some flexibility is sometimes available in specifying the domain, which is why I argued earlier that studying the sensitivity of our conclusions to the choice of domains should be part of our analysis. If we find that particular problems carry much of the burden of the proofs, then it is critical to make sure that they should be included.

For instance, in the study of resource allocation, we often allow preferences exhibiting an arbitrarily large degree of substitutability between goods or an arbitrarily large degree of complementarity (linear preferences and Leontieff preferences). Moreover, these preferences are often used in proofs. If in the particular class of situations that we have in mind, natural (upper or lower) bounds on degrees of substitutability between goods are justified, then these bounds should be imposed. There are however interesting situations where no such bounds exist, where for instance certain goods may essentially be undistinguishable, so that allowing for perfect substitutability is then quite legitimate. The domain should include these preferences, and there is nothing wrong if they appear in proofs.

We often start working with a standard domain, not knowing how much of a role its size will play in the analysis, but as results accumulate, we typically gain insights into the issue. For certain properties, we now have a deep understanding of it, an understanding that should be our goal in general. The development of the literature on *strategy-proofness* illustrates well how concerns about largeness of domains can be completely alleviated as a field evolves.

(i) The first studies of the property were conducted on abstract domains of Arrovian social choice, in which the set of alternatives is unstructured and preferences are unrestricted. The central result of that literature, the Gibbard (1973)-Satterthwaite (1975) theorem, essentially states that on such a domain, a solution is *strategy-proof* if and only if it is dictatorial (one agent is chosen beforehand and an outcome that is best for his announced preferences is selected). Could this theorem, proved on such a large domain, have any relevance to concretely specified economic models, models in which the set of alternatives is endowed with a variety of mathematical structures and preferences are correspondingly restricted?

(ii) Major progress in answering this question was achieved in the early 90's by Barberà and Peleg (1990). They derived the dictatorship conclusion for a model in which the space of alternatives is given a topological structure and preferences are required to be continuous. However, they imposed no convexity assumption on preferences. Moreover, in their proofs, they used preferences having several local maxima. Such preferences are usually excluded from our economic models.

(iii) However, Zhou (1991) imposed all of the classical assumptions and still derived dictatorship: for preferences of the kind typically considered in our microeconomic textbooks, dictatorship cannot be escaped.

(iv) Schummer (1997) further narrowed the class of admissible preferences and showed that under further restrictions such as homotheticity and even linearity, dictatorship still holds.

(v) Finally – but this is not quite the end of this journey since the latter results only apply to two-person economies – in the case of linear preferences, Schummer (1997) was able to exactly calculate how large the number of possible preferences had to be to force dictatorship. Remarkably, *only four* suffice.

For economies with indivisible goods and economies with public goods, Schummer (1996, 1999) has similarly shown that extremely narrow classes of problems lead to dictatorship.

After the initial results of Gibbard and Satterthwaite, one could legitimately entertain doubts about the relevance of their conclusion to concretely specified models of resource allocation. Thanks to these recent developments, we now know that dictatorship is essentially inescapable.

The attention that has been lavished on *strategy-proofness* is unequaled however. For other properties, and other classes of problems, we often do not know how sensitive to largeness of domains our conclusions are. Such analysis will have to be part of the axiomatic program as it develops further.

There is of course no fundamental reason why progress should only be in the direction of progressive narrowing of domains. Sometimes, starting from a domain on which certain properties are known to be compatible, we may be curious about how much and in what direction the domain can be widened without existence being lost. And we will want to determine how the class of admissible solutions will narrow in the process.

With regards to *strategy-proofness*, Alcade and Barberà (1994) have explored this issue for matching theory, Barberà et al. (1991) for the election of a committee, Ching and Serizawa (1998) for the allocation of a private good when preferences are single-peaked, Berga and Serizawa (1996) for public decision, again when preferences are single-peaked, and Ehlers (1999) for the assignment of indivisible objects. In each of these studies, the authors have been able to answer precisely the question whether a characterization obtained on a certain domain would persist when the domain is extended at all.[35]

9.5 Axioms are not descriptive of behavior

An additional criticism often addressed at the axiomatic method is that "people do not behave according to the axioms". Here the issue has to do with the scope of the axiomatic method, an issue discussed at length in the next section. The answer is that axiomatic studies are not necessarily concerned with behavior, but nothing prevents the axiomatic method from being used in addressing these issues. I will in particular discuss its usefulness in the study of equilibrium in games. There, the axioms are meant to formalize "components" of behavior. For instance, is it plausible to think that players discard dominated strategies? If yes, we may consider writing this down as one of the axioms that will compose their behavioral portrait.

On the other hand, in the normative analysis of allocation problems, the axioms are not intended to reflect behavior but rather social values. In formulating the rules according to which goods will be produced or exchanged, should we care about efficiency? Should we care about how gains made possible by improvements in technologies are distributed? Should we care about the impact of population changes on existing populations? These are essentially normative, not descriptive, issues.

10 The scope of the axiomatic program

In this section, I discuss the scope of the axiomatic method. Its relevance is wider than generally thought, and in particular it is not limited to abstract models and problems of cooperation.

---

[35] When no extension is possible, the domain is "maximal" for the list of properties that are being investigated.

10.1 Is the axiomatic method mainly suited to the analysis of abstract models?

Axiomatic studies of the abstract models of social choice, bargaining, and coalitional games are quite numerous, whereas until recently the number of axiomatic studies of concretely specified classes of resource allocation problems had been rather limited. This may suggest that the axiomatic method is mainly suited to the study of abstract domains. I believe otherwise, for the following reasons:

1. *First, enough evidence has accumulated in the last ten years to make a convincing case that the axiomatic method is not only conceptually compatible with concrete formulations but also operationally useful*; it does offer a workable and productive way of analyzing concretely specified economic models. The conceptual apparatus that has been elaborated, the proof techniques that have been developed, and the body of results that have been obtained, together provide what I consider compelling evidence in support of this position.

In addition to the examples used throughout this paper, see Young (1994); Moulin (1995); Thomson (1999b); or Moulin and Thomson (1997); for surveys of the literature on resource allocation; also see the various references of Subsect. 12.3 concerning strategic analysis.

2. Conversely, and with the possible exception of Arrovian social choice, the impression that *the theory of abstract models had progressed only, or principally, in the axiomatic mode, is greatly mistaken anyway.*

The historical record is clear. In the theory of bargaining, between Nash's publication of his classic article (1950) and the middle seventies, when the literature underwent a significant revival thanks to Kalai and Smorodinsky (1975) and Kalai (1977), only a handful of axiomatic studies of the bargaining problem appeared. In the theory of coalitional games with transferable utility, no axiomatization of solutions other than the Shapley value and variants of it was developed in almost thirty years following Shapley's classic 1953 paper. Apart from Sobolev's work (1975) on the prenucleolus (Schmeidler 1969), work that did not become known in the West for several years,[36] it is only in the early eighties that axiomatic analysis took a preeminent position in that branch of the literature. Then, axiomatic derivations of the core (Gillies 1959) and the prekernel (Davis and Maschler 1965) were obtained by Peleg (1986). At that time, characterizations of the Shapley value from new perspectives were also discovered (Young 1985; Hart and Mas-Colell 1989).

Nash's and Shapley's founding papers did give an axiomatic "tone" to the theory of bargaining and to the theory of coalitional games with transferable utility,[37] but as the above references indicate, these authors were essentially

[36] To this date, there is no published English translation of Sobolev's fundamental characterization of the prenucleolus, although several have been circulated.

[37] This may explain the mistaken view about the role played by the axiomatic method in the development of the theory of cooperative games described above, since no game theory textbook goes much beyond these two papers, and most students of the field obtain a flavor of the methodology through the abbreviated treatment that they find there.

not followed in their methodology until relatively recently, and in fact as far as the latter is concerned, quite recently.

An even more striking example is the theory of coalitional games without transferable utility. Until the late 1980's, that literature had been entirely non-axiomatic: none of the central solutions, the core, the λ-transfer value (Shapley 1969), the Harsanyi value (Harsanyi 1959, 1963), were given axiomatic justifications until twenty or thirty years after they were introduced. These characterizations are due to Peleg (1985) for the core, Aumann (1985a) for the λ-transfer value, and Hart (1985) for the Harsanyi value. Then, other solutions were also discovered in the course of axiomatic analysis – an example here is Kalai and Samet's (1985) egalitarian solution.

10.2 Is the axiomatic method mainly suited to the analysis of cooperative situations?

Another common perception is that the axiomatic method is mainly suited to the study of cooperative models. I argue below that this view is mistaken and I devote Sect. 12 to a discussion of the relevance of the axiomatic method to the study of strategic interaction.

11 On the relevance of the axiomatic method to the study of resource allocation

Here, I discuss the relevance of axiomatic studies of abstract models to the understanding of concrete resource allocation problems.

Instead of directly analyzing a class $\mathscr{C}$ of resource allocation problems specified with all of their physical details, a standard way of proceeding is to "reduce" them first so as to obtain abstract problems in a class $\mathscr{A}$ that we understand, and then to apply the conclusions derived in the analysis of $\mathscr{A}$.

1. *A first issue in evaluating the legitimacy of this approach is whether each concrete problem in $\mathscr{C}$ is mapped into one of the abstract problems in $\mathscr{A}$.* The answer is yes for several important classes.

Consider the problem of allocating private goods: under standard assumptions on preferences, endowments, and technologies, by taking the image in utility space of the set of feasible allocations (this is the reduction to which we just alluded), we obtain a problem satisfying the assumptions typically made in the theory of bargaining (non-degeneracy, convexity, compactness, and comprehensiveness).

If coalitions can form and preferences are quasi-linear, we can associate with each economy a coalitional game with transferable utility (by defining the worth of a coalition to be the maximal aggregate utility the coalition can achieve by redistributing among its members the resources it controls), and in fact this game satisfies the balancedness condition that has been central to the theory of these games (Shapley and Shubik 1969).

If general preferences are permitted, we end up with problems belonging to one of the classes that are standard in the theory of coalitional games without transferable utility.

2. *However, that each resource allocation problem in $\mathscr{C}$ maps to some problem in $\mathscr{A}$ is not sufficient to justify applying the results obtained in the study of $\mathscr{A}$. Since these results pertain to solutions defined on the whole of $\mathscr{A}$, we need to know whether conversely, each of the problems in $\mathscr{A}$ can be derived from some problem in $\mathscr{C}$.* We do have fairly general, and positive, answers to this kind of questions, at least when the class of concrete problems are exchange economies. Unfortunately, for other domains, not much is known.

Billera (1974) and Billera and Bixby (1973a, 1973b) have shown that if a bargaining problem satisfies the standard conditions mentioned in item 1, then indeed it is the image in utility space of some problem of distribution of private goods in which preferences satisfy standard assumptions. Similarly, Shapley and Shubik (1969) have shown that each totally balanced coalitional game with transferable utility can be derived from some economy satisfying commonly imposed assumptions. The main restriction in each of these studies has to do with the number of goods, which should be sufficiently large in relation to the number of agents. Sprumont (1997) has initiated the investigation of the conditions that a coalitional game with transferable utility has to satisfy in order to arise from some economy with public goods.

3. *Further, consider a requirement $P_{\mathscr{A}}$ involving pairs of abstract problems, and a requirement $P_{\mathscr{C}}$ involving pairs of concrete problems, such that the images in utility space of two concrete problems satisfying the hypotheses of $P_{\mathscr{C}}$ are two abstract problems satisfying the hypotheses of $P_{\mathscr{A}}$. Suppose that we have been able to determine the implications of $P_{\mathscr{A}}$. We would like to know whether we can deduce from this knowledge the implications of $P_{\mathscr{C}}$. To answer this, we need to know whether for each pair of problems satisfying the hypotheses of $P_{\mathscr{A}}$, there is a pair of concrete problems satisfying the hypotheses of $P_{\mathscr{C}}$ and whose images in utility space are the two abstract problems.*

This point is somewhat more subtle and the following illustration might be more revealing that the general statement. Suppose that the analysis of $\mathscr{A}$ has involved axioms pertaining to pairs of problems. In bargaining theory, an example is when two problems are related by inclusion, a situation to which the requirement of *strong monotonicity* pertains: it says that if the feasible set expands the payoffs of all agents should be at least as large as they were initially. It is often motivated by reference to an economic situation in which physical resources increase, and the desire to make all agents benefit from such increases. The implications of this requirement in bargaining theory are well understood: in the presence of efficiency, only the so-called monotone path solutions are acceptable (Kalai 1977; Thomson and Myerson 1980). The possibility of applying this result to economies hinges on whether, given two bargaining problems related by inclusion, there exist two economies that differ only in their endowments of resources – the endowment of one should dominate the endowment of the other – and such that their images in utility space coincide with the two bargaining problems. Slightly more formally, given a pair of problems in $\mathscr{A}$ related by inclusion (a situation to which we would like to apply the axiom of *strong monotonicity*), when are they the images of the two versions of a given problem in $\mathscr{C}$ resulting from two choices of the social endowment, one of which dominates the other, (a situation to which the axiom of *resource-monotonicity* applies)? When this operation is possible is useful information, but I am not aware of any general study of it. Certainly, we know from our previous discussion that a general positive answer should not be expected.

4. The operation may not always be critical however, for the following reason. In a characterization proof, not all possible problems or pairs of problems are used. Then, *the more limited question that needs to be asked is whether the pairs used in the proof of the characterization can be obtained from pairs of concrete problems satisfying the hypotheses of the axiom.*

In our example, not all pairs related by inclusion are used in deriving a characterization of the class of *strongly monotonic* solutions to the bargaining problem; in fact, much more restricted classes of such pairs are needed.

5. A limitation of the abstract model is that changes in the parameters as described in the hypotheses of an axiom may occur not only in the concrete circumstances motivating the condition but also in circumstances that are unrelated to them. The description of the model not being rich enough for the investigator to verify when the motivating situation applies, other situations may be "smuggled in" that were not intended, widening the scope of the condition too much. *To avoid this pitfall, it is important to directly study how a given solution defined on $\mathscr{A}$ and in which one may be interested behaves, when applied to the images of pairs of problems in $\mathscr{C}$.*

For such studies, see Roemer (1986a,b 1988, 1990, 1996) and Chun and Thomson (1988), who considered which monotonicity and consistency conditions are satisfied by solutions to the bargaining problem when they are used to define solutions to resource allocation problems.

In this regard, it is useful to note however that for a number of properties, as the number of commodities increases, what can be achieved enlarges considerably. In fact, as soon as the number of commodities is equal to two, the behavior of bargaining solutions when applied to economic problems is essentially what it is on abstract domains (Chun and Thomson 1988). These results show that the one-commodity case is quite special, putting into question the relevance of the numerous studies that have taken it as canonical example.

The advantage of working within a concretely specified model is that we can exactly identify the circumstances under which the possibility of an enlargement of the feasible set occurs, and decide case by case how the solution should respond. Altogether, and in the absence of complete answers to some of the questions just raised, it may be safer to work directly with concretely specified resource allocation models rather than abstract problems. The numerous references that I have given to recent studies of such models were intended to show that this position is not only methodologically sound but also operationally productive.

12 On the relevance of the axiomatic method to the study of strategic interaction

In this section, I discuss the application of the axiomatic method to the study of strategic models.

12.1 The conceptually flawed opposition between axiomatic game theory and non-cooperative game theory

As a preface to this discussion, I will note a frequent misunderstanding pertaining to the traditional division of game theory into its "cooperative" and "non-cooperative" branches. The former is thought of by many as the natural domain of application of the axiomatic method, and it is often referred to as "axiomatic game theory", non-cooperative games being the domain of "strategic" analysis. To illustrate, the axiomatic theory of bargaining is commonly opposed to its non-cooperative counterpart: axiomatic game theory is understood to be normative, that is, its objective is to recommend normatively appealing compromises; by contrast, non-cooperative game theory is supposed to be descriptive of the way a group of agents, each of them intent on promoting his own interest, would solve conflicts without outside interference.

My first observation is that *this opposition between the axiomatic approach and the non-cooperative approach is conceptually flawed.* Indeed the term "axiomatic" refers to the *methodology* of the investigator, who is outside of the game, and the term "non-cooperative" to the *behavior* of the agents involved in the game. Moreover, as discussed later, nothing prevents the axiomatic method to be applied to the study of non-cooperative games.

It may be more useful to distinguish between modes of analysis on the basis of the degree of concreteness with which we define the problems that we consider. It is this distinction that motivates the following sections.

12.2 Are abstract models of game theory more general, or less general, than concrete models?

Abstract models have been criticized for not providing adequate representations of the richness of actual conflicts. But they have also been praised for allowing a wider coverage: by discarding information about the concrete details of actual problems, we can handle within a single theory a much broader class of situations. Which viewpoint is the correct one?

1. In support of the first position, note that a game tree can be "collapsed" into a normal form game by ignoring all information about the tree structure and retaining only strategies and their associated payoffs, and a normal form game can in turn be collapsed into an abstract problem by ignoring all strategic information and retaining only the set of feasible payoffs. Therefore any solution defined on a class of abstract problems specified in utility space, can be mapped into a solution on a class of normal form games, and this solution can in turn be mapped into a solution on a class of extensive form games.

The conclusion is therefore mathematically unescapable that a possibly greater class of solutions is available for concretely specified models.

In support of the second position, I simply note that natural procedures can often be defined for associating with each normal form game an extensive form game, and for associating with each abstract problem a normal form game. Then, a solution to extensive form games can be mapped into a solution to normal form games. Similarly, a solution to normal form games can be mapped into a solution to abstract bargaining problems.

An operation of this latter kind was performed by Nash (1950) who suggested associating with each bargaining problem a certain strategic "game of demands". Another such procedure, a "game of solutions", was developed by van Damme (1986). Starting from a game specified in concrete terms, Stahl (1972) and Rubinstein (1982) have also proposed ways of associating with it a certain strategic game in extensive form, a "game of alternating offers". Gül (1989) and Hart and Mas-Colell (1996) have considered coalitional games and associated with each such game a sequential bargaining process.

2. *In actual conflicts, agents' actions are constrained in a variety of ways, due to tradition, laws, or historical accidents. It is often argued that it is these constraints that give each problem its specific character, and that without a realistic description of them, there is no hope of understanding how it will be solved.* Although the existence of such constraints cannot be denied, it is also true that considerable flexibility remains. Bargaining does not take place according to the rigid scenarios spelled out in most of our formal studies. The order in which agents move is quite variable; so is the time interval that separates an offer from a counter-offer; and the nature of these offers and counter-offers varies considerably.[38]

Of course, no mathematical model can possibly take into account all of this detail, and a focus on the central aspects of the negotiations is required. This is where the judgment of the modeler comes in, a judgment that only robustness analysis can test. If it is true that alternative modelings of a given bargaining situation essentially all lead to the same outcome, then a justification for the model has been obtained. A model of bargaining should be formulated so as to capture the essential elements of a class of relevant situations. The only way to become convinced of whether modeling has been successful is to perform this robustness analysis.

3. *A counter-argument is that situations where some flexibility seems to exist have been mis-specified.*

If the time at which bargaining has to be concluded is flexible, and is actually under the control of the players, then this flexibility should be incorporated into the analysis. If a certain issue may be part of the negotiations, the choice of the players to bring it up should also be put into the model. The possibilities of throwing away utility, being represented by a third party, extending the scope of negotiation to new issues, calling in an arbitrator, setting the agenda, …, can all in principle be incorporated in the game form or

[38] See Perry and Reny (1994), for an analysis where some flexibility is modeled.

the tree, strengthening the argument that there is never any need to consider anything more than the actual game form or the tree.

This argument is formally correct, but it actually begs the issue: until we understand well how these various changes in the game form or the tree affects the outcome, it is sterile to claim that only exactly specified game forms or trees should be analyzed. A successful negotiator is not one who only understands whatever explicit rules are given but rather one who knows how to manipulate the rules, that is, understands what could be called "the implicit game".

Political scientists, who have had to be concerned with procedures much more than economists, have contributed importantly to the understanding of how they affect the outcome of games. In some contexts, it has been shown that an appropriate choice of agenda could lead to any point in policy space (McKelvey 1976).

4. In order to be effective, the axiomatic method typically requires that the domain be "large enough," whereas players engaged in a particular conflict situation need not be concerned about other conflict situations. And indeed, why should they be? The answer to this criticism is two-fold: first, it is hard to imagine a player selecting a strategy in the particular game that he is facing today without drawing on his experience in previous situations of the same kind and attempting to formulate rules as to how he should play games in general. Minimally, he has to speculate about what his opponent(s) will do, so that his thinking should cover at least two game situations, not just one. Rationality on the part of a player does seem to require that he develops some theory of how to play games that extends beyond the particular game in which he is currently involved. Second, as analysts, and *even if the players are assumed to play only one game, we can feel confident about our conclusions only when we have understood how the solution that we are proposing behaves on a variety of games.* Our theory can only gain strength by being tested on a class of games.

When it comes to the recommendations that a judge or arbitrator should make, the need for a general procedure is also quite clear. Consider for instance the problem of dividing the liquidation value of a firm, say 12, between two claimants with claims 8 and 10. Without a general procedure for solving such bankruptcy problems, what should one think of the awards of 5 to claimant 1 and 7 to claimant 2? It is virtually impossible to evaluate such a recommendation in isolation, but by bringing within the scope of the exercise other situations of the same kind, one can begin to form an opinion. For instance, it is easier to evaluate the above recommendation *and* the awards of 5 to claimant 1 and 8 to claimant 2 when the liquidation value is 13, when these two situations are considered together. More generally, by extending the class of problems to be solved, we are better able to decide what to do for each of them. The various ways in which recommendations could or should be related as parameters change is what the axioms will express.

I also believe that the parties involved are much more likely to accept the decision of the judge or arbitrator if he provides reasons for his decision. Such reasons are most likely to refer to other similar situations. Once again, these reasons are the axioms of our theories.

12.3 Early achievements of the axiomatic method applied to strategic models

It is obvious that there is no intrinsic reason why abstract models should be analyzed only axiomatically, and conversely, as I have attempted to show, the axiomatic method can be profitably applied to concrete classes of resource allocation problems. I will now argue that *there is also no reason why strategic interaction should not be studied axiomatically. A number of axiomatic studies of strategic models have in fact been conducted, and they amply demonstrate the relevance and the usefulness of the approach.* Given the proliferation of solutions for strategic models that has occurred (van Damme 1991), the axiomatic method might in fact be quite welcome in sorting them out. I now give a list of contributions that are particularly significant in this regard.

1. Harsanyi and Selten (1988)'s book is a primary illustration. The authors consider normal form games and formulate a rich variety of conditions on solutions. Examples are the basic *invariance with respect to isomorphisms*, which says that two games that are the same up to a linear transformation of utilities and renaming of agents, should be solved in the same way up to that transformation; the self-explanatory *invariance with respect to payoff transformations that preserve the best reply structure*; *payoff monotonicity*, which says that if a pure strategy combination is chosen for some game and the payoff function is changed by increasing the payoffs at that strategy combination, then it should still be chosen for the new game; *cell consistency*, which says that the solution outcome of a game should agree with the solution outcomes of its cells; *truncation consistency*, which says that the solution outcomes of a truncated game should agree with the solution outcomes of the non-truncated game. Other axioms are *invariance with respect to sequential agent splitting, partial invariance with respect to inferior choices, partial invariance with respect to duplicates.* Harsanyi and Selten establish a large number of compatibility and incompatibility theorems. A related contribution is by Selten (1995).

2. Abreu and Pierce (1984) consider extensive form games and investigate the existence of solutions satisfying the following three axioms. *Normal form dependence*: two games having the same normal form are solved in the same way. *Dominance:* no dominated strategy is part of any solution outcome, and if $\hat{T}$ is obtained from T by eliminating a dominated choice, then the solution outcomes of $\hat{T}$ are the projections of the solution outcomes of T on $\hat{T}$. *Subgame replacement*: replacing a subgame which has a unique equilibrium outcome in pure strategies by the corresponding payoffs, gives a game whose solution outcomes are the restriction of the solution outcomes of the initial game on the new game. They show that no solution satisfies both *normal form dependence and subgame replacement*, and that no solution satisfies *dominance*.

3. Kohlberg and Mertens (1986) consider sequential games and formulate several requirements on a solution for such games: *existence*, *connectedness*, *backwards induction*, *invariance*, the requirement that two games with the same reduced normal form should be solved in the same way, *admissibility*, and *iterated dominance*. See also Mertens (1989, 1991, 1992).

4. Bernheim (1988) considers normal form games and formulates a number of axioms pertaining to a player's choice of an action to maximize his payoff subject to beliefs about his opponent's choices, under the assumption that players do not assign positive probability to choices of the other players that are judged "irrational". Under these assumptions, there remain the issues *whether priors are common or not,* and *whether the choices of the other players are perceived as independent random events or not.* The four combinations of the two axioms and their two negations characterize four equilibrium concepts, iterated dominance, correlated equilibrium, rationalizability, and Nash equilibrium. See also Brandenburger and Dekel (1987), and de Wolf and Forges (1995, 1998) and Bernheim (1998).

5. Peleg and Tijs (1996) derive most of the familiar equilibrium notions for games in strategic forms from considerations of *consistency* and various notions of *converse consistency*.[39] Additional axiomatic derivations of Nash equilibrium along these lines are obtained by Peleg et al. (1994), Peleg and Südholter (1994), Norde et al. (1993), and Shinotsuka (1994).

6. Jackson and Srivastava (1996) identify a general property of solutions (a property they call "direct breaking") that guarantees a certain kind of implementability.

7. Kaneko (1994) provides an axiomatic characterization of Nash equilibrium on the basis of epistemic considerations.

8. Peters and Vrieze (1994) derive a selection from the subset of the convex hull of the set of Nash equilibrium payoffs by translating the axioms used by Nash in deriving his solution to the bargaining problem in terms of the data entering the definition of normal form games.

9. Samet (1996) gives an axiomatization of operators describing the way agents formulate hypotheses about the way a game will be played.

10. Tan and Werlang (1988), Basu (1990), Salonen (1992), Ben-Porath and Dekel (1992), Börgers and Samuelson (1992), Tedeschi (1995), and Kaneko and Mao (1996) are other studies in which the axiomatic method is used, explicitly or implicitly.

12.4 On the interplay between the axiomatic and non-axiomatic modes of analysis

Instead of pitting the axiomatic approach to the study of conflict situations against non-axiomatic approaches, or abstract models against concrete mod-

---

[39] In this context, *consistency* says that if a strategy profile is selected by a solution for a game *G*, then in the "reduced game" obtained from *G* by imagining some of the agents playing their assigned components of the profile, and appropriately redefining the payoff function, the solution would still select the restriction of the original profile to the remaining agents. *Converse consistency* pertains to the opposite operation. When a strategy profile is such that its restrictions to subgroups of players are chosen by the solution for the associated reduced games, then it is selected by the solution for the large game.

els, a multifaceted approach seems the most promising. The merits of such an approach were certainly recognized by the founders of game theory. Nowadays, it is true however that game theorists have often fallen victims to the need for specialization that in the last two decades may have been a necessary accompaniment of the considerable expansion of the field. I will therefore conclude with further illustrations of the useful role that the axiomatic method can play in the study of strategic interaction.

12.4.1 The axiomatic and non-axiomatic approaches applied to game theoretic models have sometimes met in surprising and illuminating ways

In several interesting situations, axiomatic and non-axiomatic approaches have led to the same, or closely related conclusions. In such cases, each approach lends support to the other. I will give three illustrations, already mentioned earlier, taken from the theory of bargaining.

1. The first illustration is of course Nash's own work. Nash (1950) gives an axiomatic characterization of the Nash bargaining solution. In (1953) he also shows that the equilibria of a certain strategic game superimposed on his abstract model – in this game, strategies are utility levels – produce the very same outcomes.

2. Van Damme (1986) formulates a different game, in which players' demands have to be justified as resulting from the application of well-behaved bargaining solutions to the problem at hand, but the equilibria of its game also lead to the Nash outcomes.

3. Finally, Stahl (1972) and Rubinstein (1982) reformulate the process of bargaining by incorporating temporal elements in the negotiations. Their strategic game of alternating offers generates equilibrium outcomes that also coincide with the Nash outcome under an appropriate limit argument.

12.4.2 Axiomatic analysis provides the basis for understanding why different approaches may lead to the same outcomes

Axiomatic analysis can go further and sometimes offer general results helping us understand why different approaches lead to the same conclusions. I will give two illustrations.

Consider the following theorem, which is a variant of a result due to Hurwicz (1979): if a correspondence defined on a class of exchange economies satisfying standard assumptions (i) always selects Pareto-optimal and individually rational allocations, and (ii) when the initial allocation is Pareto-optimal, selects all individually rational allocations, and finally, (iii) is *Maskin-monotonic*,[40] then it contains the Walrasian solution.

This result teaches us a very general lesson about games. Indeed, since the (Nash) equilibrium correspondence of any game is necessarily *Maskin-*

[40] This says that if an allocation is chosen for some profile of preferences and preferences change in such a way that the allocation does not fall in anybody's preferences, then it is still chosen for the new profile.

monotonic, and often satisfies the first two conditions of the theorem, then for a large class of games, (games defined on classes of exchange economies), their equilibrium correspondences always include the Walrasian solution, a rather remarkable fact.

Recently, a number of authors have explicitly searched for principles underlying general results pertaining to strategic interaction. The potential of this approach is well illustrated by the success that it has met in connection with *consistency*.

1. Krishna and Serrano (1996) demonstrate how a strategic interpretation of the *consistency* condition shown by Lensberg (1988) to characterize the Nash bargaining solution in the context of a model with a variable population, would lead to the Nash solution. In his studies of non-cooperative models of bargaining and bankruptcy, Sonn (1994) finds the *monotonicity* and *consistency* conditions developed in the axiomatic theory of bargaining to be central to the derivation of the equilibrium equations. In a series of contributions, Serrano (1993, 1995, 1997) uses similar arguments to derive the nucleolus, the core, and the kernel.
2. Hart and Mas-Colell (1996) consider a non-cooperative bargaining process for coalitional games without transferable utility and identify a particular solution which is also one that comes out of certain axiomatic considerations. Here too, *consistency* plays an important role.
3. I have already discussed the characterizations of solutions to games in strategic form obtained by Peleg and Tijs (1996). These results are based on the application of notions of *consistency* and *converse consistency*, which until then had been exclusively seen from the normative angle. For other contributions on the subject, see Peleg, Potters, and Tijs (1994), Peleg and Südholter (1994), and Shinotsuka (1994).
4. Dagan et al. (1993) consider a strategic game for bankruptcy problems and exploit *consistency* ideas in order to characterize its equilibria.
5. Moldovanu (1990) similarly identify the equilibria of a game of offers in a model of assignment by drawing on the *consistency* of a certain solution.

12.4.3 The axiomatic method sometimes usefully complements strategic analysis

One of the central results in the theory of repeated games is the so-called "folk theorem," which states that any outcome Pareto-dominating the maximin point can be obtained at equilibrium. Therefore, the predictive power of strategic analysis is sometimes very low. In situations where an equilibrium results from preplay communication, the next obvious question is how players will ever agree on any one equilibrium. Selection of an equilibrium on the basis of normative considerations examined in the axiomatic mode may provide an answer.

12.5 Implementation theory as the domain par excellence of axiomatic analysis

Most importantly perhaps, and if one of our goals as social scientists is not only to understand the way conflicts are solved in the world, but also to discover and promote methods of conflict resolution that are more likely to result in good outcomes, the rules of the game should be an object of choice. Implementation theory is concerned with constructing games with the objective of identifying which social objectives are realistically achievable in the face of strategic behavior of the agents. This field is among those that have benefited the most from axiomatic analysis.

Indeed, the axiomatic method has assisted at all levels, in the determination of which normatively appealing social objectives are compatible, which equilibrium concepts are appropriate in the analysis of the games to which agents are confronted (Jackson and Srivastava 1996), and which solutions can be implemented with respect to each chosen equilibrium concept. More recently, much attention has been devoted to the characterization of which solutions can be implemented by means of games satisfying additional properties of interest, mainly intended to permit simplicity of the procedure; here too, the approach has been mainly axiomatic, with the axioms capturing notions of computational simplicity (Dutta et al. 1995; Saijo et al. 1993; Sjöström 1996).

13 Conclusion

In this essay, I described the axiomatic method and attempted to refute arguments against it. I also presented recent accomplishments, focusing on resource allocation in concretely specified economic models. I hope that these recent successes will motivate applications to yet other areas.

Appendix

This appendix contains short descriptions of the various models most often used as illustrations in the main body of the paper.

(1) A **bargaining problem** is a pair (B, d) of a non-empty, convex and compact subset of $\mathbb{R}^n_+$ and a point d in B. The set B is interpreted as a set of utility vectors attainable by the n agents if they reach a consensus on it, and d is interpreted as the alternative that will occur if they fail to reach any compromise.

(2) A **transferable utility game in coalitional form** is a vector v in $\mathbb{R}^{2^n-1}$. The coordinates of v are indexed by the non-empty subsets of the set of players. A coordinate is interpreted as the amount of "collective utility" that the members of the corresponding coalition can obtain.

(3) A **normal form game** is a pair (S, h) where $S = S_1 \times \cdots \times S_n$ and $h : S \to \mathbb{R}^n$. For each player i, S_i is a set of actions that he may take, and the function h gives the payoffs received by all the players for each profile of actions.

(4) An **extensive form game** (with no exogenous uncertainty) is a tree T, where each non-terminal node bears as label an element of $\{1, \ldots, n\}$, and each

terminal node bears as label a point in $\mathbb{R}^n$. As compared to the previous class of games, a sequential structure is added to the set of actions, and the nodes indicate times at which agents choose actions.

(5) An **exchange economy** is a list $(R_1, \ldots, R_n, \omega_1, \ldots, \omega_n)$ where each R_i is a continuous and monotonic preference relation defined on $\mathbb{R}^\ell_+$, and $\omega_i \in \mathbb{R}^\ell_+$ is agent i's endowment. The integer ℓ is the number of commodities.

(6) An **economy with single-peaked preferences** is a list $(R_1, \ldots, R_n, \Omega)$ where R_i is a single-peaked preference relation defined over the non-negative reals. The number Ω gives the amount of a non-disposable good to be divided among the n agents.

References

Aadland D, Kolpin V (1998) Shared irrigation costs: an empirical and axiomatic analysis. Math Soc Sci 35: 203–218

Abdulakadiroglu A, Sönmez T (1998) Random serial dictatorship and the core from random endowments in house allocation problems. Econometrica 66: 689–701

Abreu D, Pearce D (1984) On the inconsistency of certain axioms on solution concepts for non-cooperative games. J Econ Theory 34: 169–174

Alcade J, Barberà S (1994) Top dominance and the possibility of strategy-proof stable solutions to matching problems. Econ Theory 4: 417–435

Arrow K (1963) Social choice and individual values. Wiley, New York

d'Aspremont C, Gevers L (1977) Equity and the informational basis of social choice. Rev Econ Stud 44: 199–209

Aumann R (1985a) An axiomatization of the non-transferable utility value. Econometrica 53: 599–612

Aumann R (1985b) What is game theory trying to accomplish? In: Arrow K, Honkapohja S (eds.) Frontiers of economics, pp 28–76

Aumann R (1985c) On the non-transferable utility value: a comment on the Roth-Shafer examples. Econometrica 48: 667–677

Aumann R (1987) Correlated equilibrium as an expression of Bayesian rationality. Econometrica 55: 1–18

Aumann R, Maschler M (1985) Game theoretic analysis of a bankruptcy problem from the Talmud. J Econ Theory 36: 195–213

Balinski M, Young P (1982) Fair representation. Yale University Press, New Haven

Barberà S (1996) Notes on strategy-proof social choice functions. In: Arrow K, Sen A, Suzumura K (eds.) Social choice reexamined, Vol. 2. Cambridge University Press, Cambridge, pp 26–45

Barberà S, Jackson M (1994) A characterization of strategy-proof social choice functions for economies with pure public goods. Soc Choice Welfare 11: 241–252

Barberà S, Peleg B (1990) Strategy-proofness voting schemes with continuous preferences. Soc Choice Welfare 7: 31–38

Barberà S, Sonnenschein H, Zhou L (1991) Voting by committees. Econometrica 59: 595–609

Basu K (1990) On the non-existence of a rationality definition for extensive form games. Int J Game Theory 19: 33–44

Bennassy JP (1982) The economics of market disequilibrium. Academic Press, San Diego

Ben-Porath E, Dekel E (1992) Signaling future actions and the potential for self-sacrifice. J Econ Theory 57: 36–51

Berga D, Serizawa S (1996) Maximal domain for strategy-proof rules with one public good. Universitat Autònoma de Barcelona mimeo. J Econ Theory (forthcoming)
Bernheim D (1988) Axiomatic characterizations of rational choice in strategic environments. Scan J Econ 88: 473–488
Bernheim D (1998) Rational strategic choice revisited. Scan J Econ 100: 537–541
Bewley T (1972) Existence of equilibria in economies with infinitely many commodities. J Econ Theory 27: 514–540
Billera LJ (1974) On games without side-payments arising from a general class of markets. J Math Econ 1: 129–139
Billera LJ, Bixby RE (1973a) A characterization of Pareto surfaces. Proc Am Math Soc 41: 261–267
Billera LJ, Bixby RE (1973b) A characterization of polyhedral market games. Int J Game Theory 2: 253–261
Bogomolnaia A, Moulin H (1999) A new solution to the random assignment problem. Rice University mimeo
Börgers T, Samuelson L (1992) Cautious utility maximization and iterated weak dominance. Int J Game Theory 21: 13–25
Bossert W (1997) Opportunity sets and individual well-being. Soc Choice Welfare 14: 97–112
Bossert W, Fleurbaey M, van de Gaer D (1996) On second best compensation. mimeo
Bossert W, Pattanaik P, Xu Y (1994) Ranking opportunity sets: an axiomatic approach. J Econ Theory 63: 326–345
Brandenburger A, Dekel E (1987) Rationalizability and correlated equilibria. Econometrica 55: 1391–1402
Brown D, Lewis T (1981) Myopic economic agents. Econometrica 49: 359–368
Campbell D, Kelly J (1993) t or $t-1$, That is the trade-off. Econometrica 61: 1355–1365
Campbell D, Kelly J (1994a) Trade-off theory. Am Econ Rev 84: 422–426
Campbell D, Kelly J (1994b) Lebesgue measure and social choice trade-off. Econ Theory: 445–459
Ching S (1992) A simple characterization of the uniform rule. Econ Lett 40: 57–60
Ching S (1994) An alternative characterization of the uniform rule. Soc Choice Welfare 11: 131–136
Ching S (1995) What are the rationales for majority decisions? A characterization of pairwise majority rule. City University of Hong Kong mimeo
Ching S (1997) Strategy-proofness and "median voters". Int J Game Theory 26: 473–490
Ching S, Serizawa S (1998) A maximal domain for the existence of strategy-proof rules. J Econ Theory 78: 157–166
Ching S, Thomson W (1992) Population-monotonic solutions in public good economies with single-peaked preferences. University of Rochester mimeo. Soc Choice Welfare (forthcoming)
Chun Y (1986) The solidarity axiom for quasi-linear social choice problems. Soc Choice Welfare 3: 297–310
Chun Y (1988) The proportional solution for rights problems. Math Soc Sci 15: 231–246
Chun Y, Thomson W (1988) Monotonicity properties of bargaining solutions when applied to economies. Math Soc Sci 15: 11–27
Corchón L (1996) The theory of implementation of socially optimal decisions in economics. McMillan, London
Dagan N (1996a) A note on Thomson's characterization of the uniform rule. J Econ Theory 69: 255–261
Dagan N (1996b) New characterizations of old bankruptcy rules. Soc Choice Welfare 13: 51–59

Dagan N, Serrano R, Volij O (1993) A non-cooperative view of taxation methods in complete information environments. Hebrew University of Jerusalem
Dagan N, Volij O (1997) Bilateral compensations and consistent fair division rules in the context of bankruptcy problems. Int J Game Theory 26: 11–25
van Damme E (1986) The Nash bargaining solution is optimal. J Econ Theory 38: 78–100
van Damme E (1991) Stability and perfection of Nash equilibria, Springer, Berlin Heidelberg New York
Davis M, Maschler M (1965) The kernel of a cooperative game. Naval Res Logist Q 12: 223–259
Debreu G (1959) Theory of value. Yale University Press, New Haven
De Wolf O, Forges F (1995) On strategic equilibria and rational choice. CORE Discussion Paper
De Wolf O, Forges F (1998) Rational choice in strategic environments: further observations. Scan J Econ 100: 529–535
Diamantaras D (1992) An axiomatic study of economies with many private and public goods. Temple University (unpublished)
Donaldson D, Weymark J (1988) Social choice in economic environments. J Econ Theory 46: 291–308
Driessen T (1991) A survey of consistency properties in cooperative game theory. SIAM Rev 33: 43–59
Dutta B, Sen A, Vohra R (1995) Nash implementation through elementary mechanisms in economic environments. Econ Des 1: 173–203
Ehlers L (1998) Probabilistic allocation rules and single-dipped preferences. University of Maastricht mimeo
Ehlers L (1999) Coalitional strategy-proof house allocation. University of Maastricht, mimeo
Ehlers L, Klaus B (1999) Coalitional strategy-proofness, resource monotonicity, and separability for multiple assignment problems. mimeo
Ehlers L, Peters H, Storcken T (1999) Probabilistic decision schemes for one-dimensional single-peaked preferences. University of Maastricht
Fishburn PC (1987) Interprofile conditions and impossibility. Harwood Academic Publishers, New York
Fleurbaey M (1994) On fair compensation. Theory Decision 36: 277–307
Fleurbaey M (1996) Théories economiques de la justice. Economica, Paris
Fleurbaey M, Maniquet F (1997) Implementability and horizontal equity imply no-envy. Econometrica 65: 1215–1219
Foster J (1994) Normative measurement: is theory relevant? Am Econ Rev 84: 365–370
Geanakoplos J, Nalebuff B (1988) On a fundamental conflict between equity and efficiency. Discussion Paper
Gevers L (1986) Walrasian social choice: some simple axiomatic approaches. In: Heller W, Starr R, Starrett D (eds.) Social choice and public decision making: Essays in honor of K. Arrow, Vol. 1. pp 97–114
Gibbard A (1973) Manipulation of voting schemes. Econometrica 41: 587–601
Gillies DB (1959) Solutions to general non-zero sum games. In: Tucker AW, Luce RD (eds.) Contributions to the theory of games IV. Princeton University Press, Princeton, pp 47–85
Gül F (1989) Bargaining foundations of the Shapley value. Econometrica 57: 81–95
Harsanyi J (1959) A bargaining model for the cooperative *n*-person game. In: Tucker AW, Luce RD (eds.) Contributions to the theory of games IV. Princeton University Press, Princeton, pp 325–355
Harsanyi J (1963) A simplified bargaining model for the *n*-person cooperative game. Int Econ Rev 4: 194–220

Harsanyi J, Selten R (1988) A general theory of equilibrium selection in games. MIT Press, Cambridge
Hart S (1985) An axiomatization of Harsanyi's nontransferable utility solution. Econometrica 53: 1295–1313
Hart S, Mas-Colell A (1989) Potential, value and consistency. Econometrica 57: 589–614
Hart S, Mas-Colell A (1996) Bargaining and value. Econometrica 64: 357–380
Hurwicz L (1977) On the dimensional requirements of informationally decentralized Pareto-satisfactory processes. In: Arrow K, Hurwicz L (eds.) Studies in resource allocation processes. Cambridge University Press, Cambridge, pp 413–424
Hurwicz L (1979) On allocations attainable through Nash-equilibria. J Econ Theory 140–165
Imai H (1983) Individual monotonicity and lexicographic maxmin solution. Econometrica 51: 389–401
Iturbe I, Nieto J (1996) On fair allocations and monetary compensations. Econ Theory 7: 125–138
Jackson M (1999) Implementation: a crash course. Caltech mimeo (unpublished)
Jackson M, Srivastava S (1996) A characterization of game-theoretic solutions which lead to impossibility theorems. Rev Econ Stud 63: 23–38
Jordan J (1982) The competitive allocation process is informationally efficient uniquely. J Econ Theory 27: 1–18
Kalai E (1977) Proportional solutions to bargaining situations: interpersonal utility comparisons. Econometrica 45: 1623–1630
Kalai E, Samet D (1985) Monotonic solutions to general cooperative games. Econometrica 53: 307–327
Kalai E, Smorodinsky M (1975) Other solutions to Nash's bargaining problem. Econometrica 43: 513–518
Kaneko M (1994) Axiomatic characterizations of Nash equilibrium. VPI discussion paper
Kaneko M, Mao W (1996) *N*-person Nash bargaining with variable threats. Japanese Econ Rev 47: 235–250
Kara T, Sönmez T (1996) Nash implementation of matching rules. J Econ Theory 68: 425–439
Kara T, Sönmez T (1997) Implementation of college admission rules. Econ Theory 9: 197–218
Kelly JS (1978) Arrow impossibility theorems. Academic Press, New York
Klaus B, Peters H, Storcken T (1998) Strategy-proof division with single-peaked preferences and individual endowments. Soc Choice Welfare 15: 297–311
Klaus B, Peters H, Storcken T (1996) Reallocations of an infinitely divisible good. Maastricht University mimeo. Econ Theory (forthcoming)
Klemisch-Ahlert M (1993) Freedom of choice: a comparison of different rankings of opportunity sets. Soc Choice Welfare 10: 189–207
Kohlberg E (1971) On the nucleolus of a characteristic function game. SIAM J Appl Math 20: 62–66
Kohlberg E, Mertens JF (1986) On the strategic stability of equilibria. Econometrica 54: 1003–1037
Kolm S-C (1997) Modern theories of justice. MIT Press, Cambridge, MA
Kolpin V (1994) Coalition equitable cost sharing of multi-service facilities. University of Oregon mimeo
Kolpin V (1996) Multi-product serial cost sharing: an incompatibility with the additivity axiom. J Econ Theory 69: 227–233
Kranich L (1996) Equitable opportunities: an axiomatic approach. J Econ Theory 71: 131–147
Kranich L (1997) Equitable opportunities in economic environments. Soc Choice Welfare 14: 57–64

Kranich L, Ok E (1998) The measurement of opportunity inequality: a cardinality-based approach. Soc Choice Welfare 15: 263–287
Krishna V, Serrano R (1996) Multilateral bargaining. Rev Econ Stud 63: 61–80
Lensberg T (1985) Stability, collective choice and separable welfare. Ph.D. Dissertation, Norwegian School of Economics and Business Administration. Bergen, Norway
Lensberg T (1988) Stability and the Nash solution. J Econ Theory 45: 330–341
Maniquet F (1994) Fair allocation with unequal production skills: the equal right approach to compensation. mimeo
Maskin E (1977) Nash equilibrium and welfare optimality. MIT mimeo
Maskin E (1985) The theory of implementation in Nash equilibrium: a survey. In: Hurwicz L, Schmeidler D, Sonnenschein H (eds.) Social goals and social organization: Essays in memory of Elisha Pazner. Cambridge University Press, 173–204
May KO (1952) A set of independent necessary and sufficient conditions for simple majority rule. Econometrica 20: 680–684
McKelvey RD (1976) Intransitivity in multidimensional voting models and some implications for agenda control. J Econ Theory 12: 472–482
Mertens J-F (1989) Stable equilibria – a reformulation. Part I. Definitions and basic properties. Math Oper Res 14: 575–625
Mertens J-F (1991) Stable equilibria – a reformulation. Part II. Discussions of the definition and further results. Math Oper Res 16: 694–751
Mertens J-F (1992) The small worlds axiom for stable equilibria. Games Econ Beh 4: 553–564
Moldovanu B (1990) Stable bargained equilibria for assignment games without side-payments. Int J Game Theory 19: 171–190
Moore J (1992) Implementation in environments with complete information. Adv Econ Theory. In: Laffont J-J (ed.) Econometric society monograph. vol. I., chapt. 5, Cambridge University Press, 182–282
Moreno B (1995) Population-monotonicity in economic with individual endowments and single-peaked preferences. University of Alicante mimeo
Moulin H (1980) On strategy-proofness and single-peakedness. Pub. Choice 35: 437–455
Moulin H (1984) Generalized Condorcet-winners for single-peaked and single-plateaued preferences. Soc Choice Welfare 1: 127–147
Moulin H (1985a) The separability axiom and equal sharing methods. J Econ Theory 36: 120–148
Moulin H (1985b) Egalitarianism and utilitarianism in quasi linear bargaining. Econometrica 53: 49–67
Moulin H (1987) Egalitarian-equivalent cost sharing of a public good. Econometrica 55: 963–977
Moulin H (1988) Axioms of cooperative decision making. Cambridge University Press, Cambridge
Moulin H (1992) An application of the Shapley value to fair division with money. Econometrica 60: 1331–1349
Moulin H (1993a) On the fair and coalitionproof allocation of private goods. In: Binmore K, Kirman A, Tani P (eds.) Front Game Theory. MIT Press, pp 151–163
Moulin H (1996) Cost sharing under increasing returns: a comparison of simple mechanisms. Games Econ Beh 13: 225–251
Moulin H (1995) Cooperative microeconomics; a game-theoretic introduction. Princeton University Press, Princeton
Moulin H, Shenker S (1991) A non-cooperative and fair mechanism for cooperative production. Duke University mimeo, (unpublished)
Moulin H, Shenker S (1992) Serial cost sharing. Econometrica 60: 1009–1037
Moulin H, Shenker S (1994) Average cost pricing versus serial cost sharing: an axiomatic comparison. J Econ Theory 64: 178–207

Moulin H, Thomson W (1988) Can everyone benefit from growth? Two difficulties. J Math Econ 17: 339–345
Moulin H, Thomson W (1997) Axiomatic analysis of resource allocation. In: Arrow K, Sen A, Suzumura K (eds.) Social choice reexamined. Cambridge University Press, Cambridge, pp 101–120
Nagahisa R-I (1991) A local independence condition for characterization of Walrasian allocation rule. J Econ Theory 54: 106–123
Nagahisa R-I (1994) A necessary and sufficient condition for Walrasian social choice. J Econ Theory 62: 186–208
Nagahisa R-I, Suh S-C (1995) A characterization of the Walras rule. Soc Choice Welfare 12: 335–352
Nash J (1950) The bargaining problem. Econometrica 18: 155–162
Nash J (1951) Equilibrium points in *n*-person games. Econometrica 36: 48–49
Nash J (1953) Two-person cooperative games. Econometrica 21: 129–140
Norde B, Potters J, Reijnierse H, Vermeulen D (1993) Equilibrium selection and consistency. Report 9341, Department of Mathematics, University of Nijmegen
O'Neill B (1982) A problem of rights arbitration from the Talmud. Math Soc Sci 2: 345–371
Pattanaik P, Xu Y (1990) On ranking opportunity sets in terms of freedom of choice. Rech Econo Louvain 56: 383–390
Pazner E, Schmeidler D (1978) Egalitarian equivalent allocations: a new concept of economic equity. Q J Econ 92: 671–687
Peleg B (1985) An axiomatization of the core of cooperative games without side-payments. J Math Econ 14: 203–214
Peleg B (1986) On the reduced game property and its converse. Int J Game Theory 15: 187–200
Peleg B (1988) Introduction to the theory of cooperative games. Hebrew University of Jerusalem
Peleg B, Potters J, Tijs S (1994) Minimality of consistent solutions for strategic games, in particular for potential games. Tilburg University mimeo
Peleg B, Sudhölter P (1994) An axiomatization of Nash equilibrium in economic situations. CAE Working Paper, Cornell University
Peleg B, Tijs S (1996) The consistency principle for games in strategic form. Int J Game Theory 25: 13–34
Perry M, Reny P (1994) A noncooperative view of coalition formation and the core. Econometrica 62: 795–817
Peters H (1986) Characterizations of bargaining solutions by means of their status quo sets. University of Limburg, Maastricht, The Netherlands
Peters H (1992) Axiomatic bargaining game theory. Kluwer Academic, Dordrecht
Peters H, Vrieze O (1994) Nash refinement of equilibria. J Optimiz Theory Appl 83: 355–373
Polterovich V, Spivak V (1983) Gross substitutability of point to set correspondences. J Math Econ 11: 116–140
Postlewaite A (1985) Implementation via Nash equilibria in economic environments. In: Hurwicz L, Schmeidler D, Sonnenschein H (eds.) Social goals and social organization: essays in memory of E. Pazner. Cambridge University Press, Cambridge, pp 205–228
Puppe C (1995) Freedom of choice and rational decisions. Soc Choice Welfare 2: 137–153
Raiffa H (1953) Arbitration schemes for generalized two-person games. In: Kuhn HW, Tucker AW (eds.) Contributions to the theory of games. Princeton University Press, Princeton, pp 361–387
Roemer J (1986a) The mismarriage of bargaining theory and distributive justice. Ethics 97: 88–110
Roemer J (1986b) Equality of resources imply equality of welfare. Q J Econ 101: 751–784

Roemer J (1988) Axiomatic bargaining theory on economic environments. J Econ Theory 45: 1–31
Roemer J (1990) Welfare and axiomatic bargaining theory. Rech Econ Louvain 56: 287–301
Roemer J (1996) Theories of distributive justice. Harvard University Press, Cambridge
Roth AE (1979) Axiomatic models of bargaining. Springer, Berlin Heidelberg New York
Roth AE (1980) Values of games without side-payments: some difficulties with current concepts. Econometrica 48: 457–465
Roth AE (1986) On the non-transferable utility value: a reply to Aumann. Econometrica 54: 981–984
Roth AE, Sotomayor M (1990) Two-sided matching. Cambridge University Press, Cambridge
Rubinstein A (1982) Perfect equilibrium in a bargaining model. Econometrica 50: 97–109
Salonen H (1992) An axiomatic analysis of the Nash equilibrium concept. Theory Dec 33: 177–189
Saijo T, Tatamitami Y, Yamato T (1990) Toward natural implementation. Int Econ Rev 37: 949–980
Samet D (1996) Hypothetical knowledge and games with perfect information. Games Econ Beh 17: 230–251
Sasaki H (1995) Consistency and monotonicity in assignment problems. Int J Game Theory 24: 373–397
Sasaki H, Toda M (1992) Consistency and the characterization of the core of two-sided matching problems. J Econ Theory 56: 218–227
Satterthwaite M (1975) Strategy-proofness and Arrow's conditions: existence and correspondence theorem for voting procedures and social choice functions. J Econ Theory 10: 187–217
Scafuri AJ, Yannelis N (1984) Non-symmetric cardinal value allocations. Econometrica 52: 1365–1368
Schmeidler D (1969) The nucleolus of a characteristic function form game. SIAM J Appl Math 17: 1163–1170
Schummer J (1996) Eliciting preferences to assign positions and compensation. University of Rochester mimeo
Schummer J (1997) Strategy-Proofness versus efficiency with homothetic preferences in exchange economies. Soc Choice Welfare 14: 47–56
Schummer J (1999) Strategy-proofness versus efficiency for small domains of preferences over public goods. Econ Theory 13: 209–222
Selten R (1975) Reexamination of the perfectness concept for equilibrium points in extensive form. Int J Game Theory 4: 25–55
Selten R (1995) An axiomatic theory of a risk dominance measure for bipolar games with linear incentives. Games Econ Beh 8: 213–268
Sen A (1970) Collective choice and social welfare. Holden Day, San Francisco
Sen A (1977) On weights and measures: informational constraints in social choice analysis. Econometrica 45: 1539–1572
Serrano R (1993) Non-cooperative implementation of the nucleolus: the 3-person case. Int J Game Theory 22: 345–357
Serrano R (1995) Strategic bargaining, surplus sharing and the nucleolus. J Math Econ 24: 319–329
Serrano R (1997) Reinterpreting the kernel. J Econ Theory 77: 58–80
Shafer W (1980) On the existence and interpretation of the value allocation. Econometrica 48: 467–476
Shapley L (1953) A value for *n*-person games. In: Kuhn HW, Tucker NW (eds.) Contributions to the theory of games. Princeton University Press, Princeton, pp 307–317

Shapley L (1969) Utility comparisons and the theory of games. In: Guilbaud G (ed.) La décision. Editions du C.N.R.S., Paris
Shapley L, Shubik M (1969) On market games. J Econ Theory 1: 11–26
Shinotsuka T (1994) An axiomatization of Nash equilibria in economic situations. University of Rochester, mimeo
Sjöström T (1996) Implementation by demand games. Econ Design 1: 343–354
Sobolev AI (1975) The characterization of optimality principles in cooperative games by functional equations. Math Meth Soc Sci 6: 150–165 (in Russian)
Sönmez T (1999) Strategy-proofness and essentially single-valued cores. Econometrica 67: 677–689
Sönmez T (1999) Can pre-arranged matches be avoided in two-sided matching market? J Econ Theory 86: 148–156
Sonn S-Y (1994) Sequential bargaining for allocation problems. Dissertation, University of Rochester
Sprumont Y (1991) The division problem with single-peaked preferences: a characterization of the uniform allocation rule. Econometrica 59: 509–519
Sprumont Y (1995) Strategyproof collective choice in economic and political environments. Can J Economics 28: 68–107
Sprumont Y (1997) On the game-theoretic structure of public good economies. Int J Game Theory 26: 455–472
Sprumont Y (1998) Equal factor equivalence in economies with multiple public goods. Soc Choice Welfare 15: 543–558
Sprumont Y, Zhou L (1999) Pazner-Schmeidler rules in large economies. J Math Econ 31: 321–339
Stahl I (1972) Bargaining theory. Economic Research Institute at the Stockholm School of Economics, Stockholm
Tadenuma K (1992) Reduced games, consistency, and the core. Int J Game Theory 20: 325–334
Tadenuma K, Thomson W (1991) No-envy and consistency in economies with indivisible goods. Econometrica 59: 1755–1767
Tadenuma K, Thomson W (1993) The fair allocation of an indivisible good when monetary compensations are possible. Math Soc Sci 25: 117–132
Tadenuma K, Thomson W (1995) Refinements of the no-envy solution in economies with indivisible goods. Theory Decision 39: 189–206
Tan T, Werlang S (1988) The Bayesian foundations of solution concepts of games. J Econ Theory 45: 370–391
Tauman Y (1988) The Aumann-Shapley prices: a survey. In: Roth AE (ed.) The Shapley value: Essays in honor of Lloyd Shapley. Cambridge University Press, Cambridge, pp 279–304
Tedeschi P (1995) Bargained-correlated equilibria. Games Econ Beh 9: 205–221
Thomson W (1984) Monotonicity, stability, and egalitarianism. Math Soc Sci 8: 15–28
Thomson W (1985) A note on the Nash solution. University of Rochester, mimeo
Thomson W (1987a) Is growth compatible with justice: another impossibility. University of Rochester, mimeo
Thomson W (1987b) Monotonic allocation mechanisms. University of Rochester, mimeo
Thomson W (1988) A study of choice correspondence in economies with a variable number of agents. J Econ Theory 46: 237–254
Thomson W (1990a) A replacement principle. University of Rochester, mimeo
Thomson W (1990b) Manipulation and implementation of solutions to the problem of fair division when preferences are single-peaked. University of Rochester, mimeo
Thomson W (1993) The replacement principle in public good economies with single-peaked preferences. Econ Lett 42: 31–36

Thomson W (1994a) Resource-monotonic solutions to the problem of fair division when preferences are single-peaked. Soc Choice Welfare 11: 205–223
Thomson W (1994) Consistent solutions to the problem of fair division when preferences are single-peaked. J Econ Theory 63: 219–245
Thomson W (1995a) Population-monotonic solutions to the problem of fair division when preferences are single-peaked. Econ Theory 5: 229–246
Thomson W (1995b) Population-monotonic allocation rules. In: Barnett W, Moulin H, Salles M, Schofield N (eds.) Social Choice, Welfare, and Ethics. Cambridge University Press, Cambridge, chap. 4, 79–124
Thomson W (1995c) The axiomatic analysis of bankruptcy and taxation problems: a survey. University of Rochester
Thomson W (1996a) Consistent allocation rules. University of Rochester, mimeo
Thomson W (1996b) Fair allocation rules. University of Rochester, mimeo
Thomson W (1997) The replacement principle in economies with single-peaked preferences. J Econ Theory 76: 145–168
Thomson W (1998) The replacement principle in economies with indivisible goods. Soc Choice Welfare 15: 57–66
Thomson W (1999a) Bargaining theory: the axiomatic approach. Academic Press (forthcoming)
Thomson W (1999b) The theory of fair allocation. University of Rochester, mimeo
Thomson W, Lensberg T (1989) The axiomatic theory of bargaining with a variable number of agents. Cambridge University Press
Thomson W, Myerson R (1980) Monotonicity and independence axioms. Int J Game Theory 9: 37–49
Thomson W, Zhou L (1993) Consistent allocation rules in atomless economies. Econometrica 61: 575–587
Toda M (1991) On the fair allocations in the two-sided matching problems. Tokyo-Keizai University. Discussion Paper
Toda M (1995) Characterizations of the core of two-sided matching problems which allow self-matchings. Tokyo-Keizai University. Discussion Paper
Toda M (1996) Monotonicity properties and characterization of the core of matching problems. Tokyo Keizai University, mimeo
Vohra R (1999) The replacement principle and tree-structured preferences. Econ Lett 63: 175–180
Yannelis N (1982) A remark on the cardinal value equivalence theorem. mimeo
Young P (1974) An axiomatization of Borda's rule. J Econ Theory 9: 43–52
Young P (1985) Monotonic solutions for cooperative games. Int J Game Theory 14: 65–72
Young P (1987) On dividing an amount according to individual claims or liabilities. Math Oper Res 12: 398–414
Young P (1988) Taxation and bankruptcy. J Econ Theory 44: 321–335
Young P (1994) Equity: in theory and practice. Princeton University Press, Princeton
Zhou L (1991) Inefficiency of strategy-proof allocation mechanisms in pure exchange economies. Soc Choice Welfare 8: 247–254