ELSEVIER European Economic Review 42 (1998) 553–560

EUROPEAN ECONOMIC REVIEW

Definable preferences: An example[1]

Ariel Rubinstein*

Department of Economics, Tel Aviv University, Tel Aviv 69978, Israel

Abstract

A preference relation is definable in a language if there is a formula in this language which is satisfied precisely for those pairs which satisfy the relation. The paper suggests that definability is a natural category of requirements of preferences in economic models. To demonstrate the analytical possibilities, the paper studies the set of definable preferences in one context using Craig Lemma. © 1998 Elsevier Science B.V. All rights reserved.

JEL classification: C00; D00

Keywords: Preferences; Definability; Logics

1. Introduction

An economic agent enters into a standard economic model accompanied by a preference relation defined on a set of relevant consequences. The preferences

[1] Much of the material in this chapter is based on Rubinstein (1978), one of my early papers, which was never published; for a related paper, see Rubinstein (1984). An extended version of the current paper will consist of Chapter 1 in a forthcoming book, 'Language and Economics', based on my Churchill Lectures, delivered at Cambridge University, England, Spring 1996.

* Also at: Department of Economics, Princeton University, Princeton, NJ 08544, USA. E-mail: rariel@princeton.edu. Website: http://www.princeton.edu/~ariel.

PII S0014-2921(98)00008-7

are taken to be the basis for systematic description of his behavior as well as for welfare analysis. We usually assume that an economic agent is 'rational' in the sense that his choice, in any given situation, is an outcome of him maximizing his preferences. Given that we adopt the rational man paradigm, the other constraints imposed upon an economic agent's preferences are often weak: For example, in general equilibrium theory we usually just impose conditions of monotonicity, continuity and quasi-convexity. But in many economic studies, we do restrict attention to some family of preferences which have simple utility presentation. Several questions thus arise: When we do not restrict the set of preferences, why do we not do so, even though the restriction may prevent us from obtaining stronger results? When we restrict attention to a small family of preferences, what is the basis for the restriction? Why is it that the utility function $(\log(x_1 + 1))x_2$ in a two-commodity world lies within the scope of classic studies whereas lexicographic preferences do not? And, in general, what are the types of considerations that make us include or exclude preferences from the scope of the analysis?

One possible consideration is that some preferences may better fit empirical data; although I am ignorant about empirical economics, I doubt that this was a significant consideration in the choice of the restrictions imposed on the set of preferences in the economic theory literature. Another consideration is 'analytical convenience'; this is a serious consideration, which has to be taken with the necessary natural caution. From the point of view of 'bounded rationality', one may argue that some preferences are more plausible than others since they can be derived from plausible procedures of choice. Finding such derivations is one of the main targets of models of bounded rationality (for a discussion of this point see, for example, Chapter 2 in Rubinstein, 1997).

This paper, however, is concerned with a different consideration: the availability of a description of the preferences in a decision maker's language. The presumption of this paper is that when a decision maker is involved in an intentional choice, he, using his daily language, *describes* the choice to himself or to agents who operate on his behalf. Thus, 'My first priority is to get as many guns as possible and only secondarily do I worry about increasing the quantity of food' is a natural description of a preference relation. 'I spend 35% of my income on food and 65% on guns' is a natural description of a rule of behavior, one consistent with maximizing some Cobb–Douglas utility function. On the other hand, the function $(\log(x_1 + 1))x_2$ is a standard utility function that is expressed by a simple mathematical formula, but I do not know about any rule of behavior stated in everyday language which corresponds to this utility function.

When a decision maker is a collective (recall that decision makers in economics are often families, groups or organizations), the presumption that preferences are definable makes even more sense. In that case, a decision rule must be stated in words in order to be communicated among the individuals in the collective, in the deliberation and the implementation stages.

The aim of this paper is to demonstrate that the requirement that preferences be definable can be analyzed fruitfully in formal terms. More specifically, the paper provides an example of a formal investigation of the connection between a decision maker's language and the set of definable preferences. This may be a first step in a much more ambitious research plan, aimed at studying the interaction between economic agents taking into account the 'language' as a restriction on agents' behavior, institution design, communication etc. But, dreams aside, let us move to the essence of the paper which is quite modest.

2. Definable preferences on the basis of binary relations

The particular problem we will deal with here is the construction of a transitive binary relation on the basis of an array of K basic binary relations. Such a problem lies at the core of social choice theory, where the K basic preferences are interpreted as the preferences held by K members of a society and the social preferences are the collective preferences. If we replace the term 'individual i' with 'property i', social choice theory is transformed from a theory about social decisions into a theory about the formation of individualistic preferences on the basis of K relevant criteria, according to which alternatives can be compared. For example, the alternatives may be the K commodity bundles and each of the K criteria refers to a comparison of the quantities of one of the K commodities.

The assumption modeled and discussed in this paper is that the decision maker, when determining whether alternative x relates to alternative y, justifies his judgment in a language using the names of the K relations. We have to specify formally what we mean by the 'language of the decision maker' and we will take the most simple way of doing so, the decision maker is assumed to use a language of the calculus of propositions (for a full presentation of this term see for example, Boolos and Jeffrey (1989) or Crossley (1972)). The primitives of such a language is a set of *atomic propositions*. A *formula* is a string of symbols constructed inductively by the following rules: Any atomic proposition is a formula; if ϕ and ψ are formulae, then $(\neg\phi), (\phi \wedge \psi), (\phi \vee \psi), (\phi \rightarrow \psi)$ and $(\phi \leftrightarrow \psi)$ are formulae (we will usually omit the parentheses). The truth value of a formula ϕ, with the K atomic propositions $v_1, \ldots, v_K$ for an assignment of truth values $t_1, \ldots, t_K$, is denoted by $\phi(t_1, \ldots, t_K)$ and is defined, naturally and inductively, using the 'truth tables' of the connectives.

Definition. Given K binary relations $P_1, \ldots, P_K$ on a set A, a binary relation P on the set A is defined by a formula ϕ in the calculus of propositions with the atomic propositions $xP_1y, \ldots, xP_Ky$ if, for any a, $b \in A$, aPb if and only if $\phi(aP_1b, \ldots, aP_Kb) = T$. If a binary relation P is defined by some formula in the calculus of propositions with the atomic propositions $xP_1y, \ldots, xP_Ky$, we will say that P is *definable*.

Thus, for example, the Pareto relation is defined by the formula $\bigwedge_k xP_k y$, and the majority relation is defined by $\bigvee\{\bigwedge_{k\in M} xP_k y | M \subseteq \{1,\ldots,K\}$ containing at least $K/2$ elements$\}$.

Note that the condition that a relation P on a set A is definable by a formula ϕ with the atomic propositions $xP_1y,\ldots,xP_Ky$ is related to what, in social choice theory, is called the *neutrality* condition: For any two pairs of alternatives a, b, and c, d if, for any k, aP_kb if and only if cP_kd, then aPb if and only if cPd.

The binary relations on which the definition of the binary relation P is based are assumed to satisfy some properties – all have the form that for any x, y, and z, there is a dependency of the truth value of xP_kz on the truth values of xP_ky and yP_kz. It is assumed that the same restriction implies equally for all P_k. More precisely, for any k, let $T(P_k)$ be a formula with the variables xP_ky, yP_kz and xP_kz, which describes the restriction on P_k. The formula $T(P_k)$ is assumed to be a noncontradicting conjunction of formulae each of the structure $\delta_1 xP_ky \wedge \delta_2 yP_kz \rightarrow \delta_3 xP_kz$, where each $\delta_i \in \{-1, +1\}$ with $-1\phi \equiv \neg\phi$ and $+1\phi \equiv \phi$. We assume that the same restriction applies to all K relations, that is $T(P_k)$ is obtained from $T(P_1)$ by replacing the any variable xP_1y, yP_1z and xP_1z with xP_ky, yP_kz and xP_kz accordingly.

Examples.

- The formula $[(xP_yy \wedge yP_kz \rightarrow xP_kz)] \wedge [(\neg xP_yy \wedge \neg yP_kz \rightarrow \neg xP_kz)]$ fits the requirement that P_k is an *ordering*: Either the truth of both xP_ky and yP_kz, or the falseness of both, determines the truth value of xP_kz (positively and negatively, respectively).
- The formula $T(P_k) = [(xP_yy \wedge yP_kz \rightarrow xP_kz)] \wedge [(xP_yy \wedge \neg yP_kz \rightarrow \neg xP_kz)] \wedge [(\neg xP_yy \wedge yP_kz \rightarrow \neg xP_kz)]$ fits to the requirement that P_k is an *equivalence relation*: The truth value of xP_kz is determined by the truth values of xP_ky and yP_kz – unless both are false.

We say that the basic relations are *deterministic* if $T(P_k)$ is a conjunction of four formulae, one for each of the possible configurations of xP_yy and yP_kz. Otherwise, we say that the basic relations are *nondeterministic*.

We now come to the main definition of the paper. We wish to investigate the set of *transitive* binary relations *definable* in this language. It requires that the definition of the relation is such that, combined with the restrictions on the primitive relations (as expressed by the formulae $T(P_1),\ldots,T(P_K)$) *logically implies* the transitivity of the relation, that is, the transitivity is satisfied whatever truth values are assigned in the atomic propositions.

Definition. Given a restriction on the basic formula expressed by $T(P_k)$, the formula ϕ with the variables $xP_1y,\ldots,xP_Ky$ *defines a transitive relation* if the

formula $[\phi(x, y) \wedge \phi(y, z) / \bigwedge_k T(P_k)] \rightarrow \phi(x, z)$ is a tautology. (The formulae $\phi(x, z)$ and $\phi(y, z)$ are the formulae obtained from $\phi(x, y)$ after substituting each xP_ky with yP_kz and xP_kz, respectively).

The requirement that the formula $\psi = [\phi(x, y) \wedge \phi(y, z) / \bigwedge_k T(P_k)] \rightarrow \phi(x, z)$ is a tautology needs elaboration. The requirement does *not* refer to any set of alternatives. The transitivity of the defined relation is required to hold in respect to all configurations of the basic relations as long as each P_k satisfies the properties required by $T(P_k)$. If we do not impose any restrictions on the basic relations, ψ would become the formula $[\phi(x, y) \wedge \phi(y, z)] \rightarrow \phi(x, z)$. However, none of the atomic propositions $\{xP_ky, yP_kz\}_k$, which appear in $[\phi(x, y) \wedge \phi(y, z)]$ appear in $\phi(x, z)$. This makes it impossible for $[\phi(x, y) \wedge \phi(y, z)] \rightarrow \phi(x, z)$ to be a tautology unless $\phi(x, z)$ is a logical contradiction (which we assume it is not).

Note that for a particular profile of basic relations $(P_k)_k$, where each P_k is a binary relation on a set X which satisfies $T(P_k)$, the formula ψ may be valid even if ψ is not a tautology. That is because the failure of ψ to be a tatology may be at a certain configuration of the truth values of the set of variable $\{xP_ky, yP_kz, xP_kz\}_k$ which is not realized in that profile of basic relations. The requirement that ψ is a *tautology* has its 'full force' regarding a particular profile only if the *profile* of basic relations is 'rich enough' so that any way in which the relations $P_1, \ldots, P_K$ can relate to a triple of alternatives is satisfied by some triple of elements in X. (This point relates to the connection between multi-profile and single-profile theorems in social choice theory; see Parks (1976), Pollak (1979) and Rubinstein (1984)).

Claim. Assume that such $T(P_k)$ is nondeterministic and let ϕ be a formula that defines a transitive relation. Then, there is a set $\kappa^* \subseteq \{1, \ldots, K\}$, and a vector of coefficients $\{\delta_k\}_{k \in \kappa^*}$, so that $\phi(x, y) \leftrightarrow \bigwedge_{k \in \kappa^*} \delta_k xP_ky$ is a tautology (and, for every $k \in \kappa^*$, the formula $[\delta_k xP_ky \wedge \delta_k yP_kz] \rightarrow \delta_k xP_kz$ is logically implied by $T(P_k)$).

The Claim states that any definable transitive relation is definable by a simple formula, which is a conjunction of atomic propositions or their negations. To demonstrate the conclusions from the Claim consider, for example, the case when the basic binary relations are both *transitive* ($xP_yy \wedge yP_kz \rightarrow xP_kz$ appears in $T(P_k)$) and *negatively transitive* ($\neg xP_yy \wedge \neg yP_kz \rightarrow \neg xP_kz$ appears in $T(P_k)$). This covers the case when all P_k are linear orderings. Then, each definable transitive relation is also definable by a formula of the type $\bigwedge_{k \in \kappa^*} \delta_k xP_ky$. This class of binary relations resembles the oligarchic binary relations that are familiar from the social choice theory literature. It follows that the only definable transitive and *complete* (either xPy or yPx for every $x \neq y$) binary relations are P_{k^*} or its negation for some k^* (that is, either the dictator or the anti-dictator in social choice terminology).

As another example, consider the construction of a classification system for a set of objects (such as flowers) on the basis of more primitive equivalence relations (such as the number of leaves, color and size). An equivalence relation is nondeterministic ($\neg xP_y y \wedge \neg yP_k z$ does not imply either $xP_k z$ or $\neg xP_k z$); thus, by the above claim every transitive relation which is definable by an equivalence relation must also be definable by a formula of the type $\bigwedge_{k\in\kappa^*} xP_k y$ for some set κ^*. In particular, any equivalence relation definable by other equivalence relations must be the conjunction of those equivalence relations.

The proof of the Claim is not difficult; its central point is an argument resembling the Craig Lemma taken from the logic literature (see, for example, Boolos and Jeffrey, 1989). The Craig Lemma states that if a formula of the type $\phi \to \psi$ is a tautology, then there must be a formula, λ, which uses only *those* atomic propositions that appear in *both* ϕ and ψ, so that both $\phi \to \lambda$ and $\lambda \to \psi$ are tautologies.

In our context we analyze the statement that the formula $\psi = [\phi(x, y) \wedge \phi(y, z)/\bigwedge\bigwedge_k T(P_k)] \to \phi(x, z)$ is a tautology. A basic proposition in the calculus of propositions states that any formula $\phi(x, y)$ in the language of propositional calculus with the variables $xP_1 y, \ldots, xP_K y$ is logically equivalent to its disjunctive normal form, which is a disjunction $\bigvee_m \phi_m(x, y)$, where each ϕ_m is a *truth configuration* of $xP_1 y, \ldots, xP_K y$; that is, a formula of the form $\bigwedge_{k=1\ldots K} \delta_k xP_k y$, where $\delta_k \in \{-1, +1\}$.

For any ϕ_m and ϕ_n in this conjunction it must be that $[\phi_m(x, y) \wedge \phi_n(y, z)/\bigwedge\bigwedge_k T(P_k)] \to \phi(x, z)$ is also a tautology. The key point of the proof is that for this implication to be a tautology it must be that there is a conjunction of facts about $\{P_k(x, z)\}$ which implies $\phi(x, z)$ and is implied by $\phi_m(x, y) \wedge \phi_n(y, z)/\bigwedge\bigwedge_k T(P_k)$.

Proof of Claim 1.1. A set κ is *decisive* if there is a vector, $\{\delta_k\}_{k\in\kappa}$, such that $\bigwedge_{k\in\kappa} \delta_k xP_k y \to \phi(x, y)$ is a tautology. Of course, $\{1, \ldots, K\}$ is a decisive set. We will see now that there is a decisive set that is minimal, namely, a subset of any decisive set. Let $\kappa(1)$ and $\kappa(2)$ be two decisive sets so that $\bigwedge_{k\in\kappa(1)} \delta'_k xP_k y$ and $\bigwedge_{k\in\kappa(2)} \delta''_k xP_k y$ imply $\phi(x, y)$. Since ϕ defines a transitive relation, $[\bigwedge_{k\in\kappa(1)} \delta'_k xP_k y] \wedge [\bigwedge_{k\in\kappa(2)} \delta''_k yP_k z] \wedge [\bigwedge_k T(P_k)] \to \phi(x, z)$ is a tautology. Thus, the set of all $k \in \kappa(1) \cap \kappa(2)$ for which $\delta'_k xP_k y \wedge \delta''_k yP_k z$ is a condition in $T(P_k)\}$, is a decisive set as well. It follows that there is a minimal decisive set, κ^*.

Assume that $[\bigwedge_{k\in\kappa^*} \delta'_k xP_k y]$ and $[\bigwedge_{k\in\kappa} \delta''_k xP_k y]$ are two truth configurations of $\{xP_k y\}_{k\in\kappa^*}$, which imply $\phi(x, y)$. Let $k \in \kappa^*$ be such that $\delta'_k \neq \delta''_k$. Since $T(P_k)$ is not deterministic, we have that for some δ' and δ'', there is no δ for which $T(P_k)$ implies $\delta' xP_k y \wedge \delta'' yP_k z \to \delta xP_k z$. But, since κ^* is minimal, $[\bigwedge_{k\in\kappa^*} \delta'_k xP_k y] \wedge [\bigwedge_{k\in\kappa^*} \delta''_k yP_k z] \wedge [\bigwedge_k T(P_k)] \to \phi(x, z)$ is not a tautology. □

3. Concluding comments

Going back to the classical consuming bundle space with K commodities, the lexicographic preferences are those specified by an order of the K commodities, $i(l), \ldots, i(K)$, so that the bundle $(a_1, \ldots, a_K)$ is preferred on the bundle $(b_1, \ldots, b_K)$ if, for some k^*, $a_{i(k)} = b_{i(k)}$ for all $k < k^*$ and $a_{i(k^*)} > b_{i(k^*)}$. That is, the K commodities are examined one by one, according to some fixed order of priorities, up to the first commodity for which the comparison of the commodity's quantities is decisive. The analysis presented in Section 2 implies that the lexicographic preferences are the only increasing preference relations that are definable in a language with atomic propositions of the type '$x_k \geq y_k$', interpreted as 'the quantity of the kth commodity in the bundle x is at least as high as that in bundle y'. Thus, while economists almost always limit the scope of their analyses and exclude lexicographic preferences, those excluded preferences are the only preferences that pass the test of definability as described in this paper.

Is the exclusion of lexicographic preferences significant? Would we obtain different results in 'market models' in which lexicographic preferences will be allowed than when we would be using models without those preferences? Consider, for example, a four-commodity exchange market where each consumer i owns exclusively one unit of commodity i (commodity i might be, for example, 'individual i's personal attention') and let the four agents' preferences be lexicographic with priority orders of (3, 4, 2, 1) (3, 4, 1, 2), (1, 2, 4, 3) and (1, 2, 3, 4). The vector (1, 0, 1, 0) is a competitive price vector in this exchange market which leads to the exchange between agents 1 and 3. However, there is no competitive equilibrium that allows also the desirable exchange between 2 and 4. The efficient allocation that results from the trade between 1 and 3, and 2 and 4, is not an outcome of any competitive equilibrium. Obtaining this efficient allocation requires different trading institutions.

To conclude, the ultimate goal of embedding the language used by a decision maker into an economic model is, of course, to derive interesting economic consequences. The aim of this paper, however, was much more modest – it was merely to draw the reader's attention to the idea that the 'definability' assumption is 'natural' as well as 'analytically attractive'.

References

Boolos, G.S., Jeffrey, R.C., 1989. Computability and Logic. Cambridge University Press, New York.

Crossley, J.N. et al., 1972. What Is Mathematical Logic?. Oxford University Press, Oxford.

Parks, R.P., 1976. An impossibility theorem for fixed preferences: A dictatorial Bergson–Samuelson welfare function. Review of Economic Studies 43, 447–450.

Pollak, C.R., 1979. Bergson- Samuelson social welfare functions and the theory of social choice. The Quarterly Journal of Economics 93, 73–90.

Rubinstein, A., 1978. Definable preference relations – Three examples. Research memorandum 31. The Center of Research in Mathematical Economics and Game Theory, The Hebrew University, Jerusalem.

Rubinstein, A., 1984. The single profile analogues to multi-profile theorems: Mathematical logic's approach. International Economic Review 25, 719–730.