DEFINEABLE PREFERENCES: ANOTHER EXAMPLE*

Searching for a Boyfriend in a Foreign Town

Ariel Rubinstein
School of Economics, Tel Aviv University and
Department of Economics, Princeton University
rariel@post.tau.ac.il, homepage: http://www.princeton.edu/~ariel

Abstract The paper argues for the formal investigation of conditions under which that preference relations are definable in various simple languages. An example of such an investigation is given. It is shown that if a formula in the "pure language with equality" induces a preference relation on the set of T-tuples of objects in a set (independently of the number of elements in the set), then it must be that there is a sequence of preferences P_n on the set of partitions of $\{1,\dots,T\}$ such that $(x_1,\dots,x_T)$ is preferred to $(y_1,\dots,y_T)$ if and only if the set of objects is of size n and $E(x_1,\dots,x_T)P_nE(y_1,\dots,y_T)$ where $E(x_1,\dots,x_T)$ is the partition of $\{1,\dots,T\}$ where i and j are in the same cell if $x_i = x_j$. Furthermore, for all large enough n the relations P_n are identical.

1. Introduction

The starting point of this paper is a view that decision makers often verbalize their considerations before making a decision. It follows that the "language" a decision maker uses to verbalize his preferences restricts his set of preferences. Thus, interesting restrictions on the richness of the decision maker's language can yield interesting restrictions on the set of an economic agent's admissible preferences.

Before proceeding to the main investigation some background is required. An economic agent in the standard economic model possesses a preference relation defined on a set of relevant consequences. These preferences provide the basis for the systematic description of his behavior and for welfare anal-

*Much of the material in this article is based on Rubinstein (1978), an early unpublished paper of mine. This paper is an "unidentical" twin to Rubinstein (1998a). The material in the two articles appears as Chapter 4 in Rubinstein (2000).

P. Gärdenfors, J. Woleński and K. Kijania-Placek (eds.), In the Scope of Logic, Methodology and Philosophy of Science, Vol. I, 235-243.

Based upon Chapter 4 of Ariel Rubinstein, *Economics and Language*, Cambridge, 2000.
Reprinted with kind permission of Cambridge University Press.

ysis. We usually assume that an economic agent is "rational" in the sense that his choice is derived from an optimization given his preferences. Since 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 only impose conditions such as monotonicity, continuity and quasi-convexity. This restriction does not exclude a preference relation defined on the space of bundles R_2^+ which states that the bundle (x_1,x_2) is evaluated by a (utility) function $(\log(x_1+1))x_2$ whereas it does exclude the lexicographic preferences which state that the bundle (x_1,x_2) is preferred to (y_1,y_2) if $x_1 > y_1$ or $(x_1 = y_1$ and $x_2 \geqslant y_2)$ and which are defined by simple logical operations.

What are the general considerations that motivate us to include or exclude preferences from the scope of analysis? One consideration is that some preferences may better explain empirical data. Although I am not an empirical economist, I am doubtful that this is a factor in the choice of restrictions imposed on preferences in the economic theory literature. Another consideration is "analytical convenience". This is a legitimate consideration but one which must be treated with the necessary caution. Another consideration relates to "bounded rationality". It may be argued that some preferences are more plausible than others since they can be derived from plausible procedures of choice. Finding such derivations is, of course, one of the main objectives of bounded rationality models (for a discussion of this point see Chapter 2 in Rubinstein (1998b)).

This analysis, however, will consider a different issue: the ability to describe the preferences in a decision maker's language. I assume that when a decision maker is involved in an intentional choice, he describes his considerations, whether to himself or to agents who operate on his behalf, using his daily language. (For a discussion of this point in the psychological literature see, for example, Shelly and Bryan (1964).) Thus, "My first priority is to obtain as many guns as possible and only after that do I worry about increasing the quantity of food" is a natural description of a preference relation which fits lexicographic preferences. "I will spend 35% of my income on food and 65% on guns" is a natural description of a rule of behavior (assuming linear budget constraints, this rule is consistent with maximizing the utility function $(\text{food})^{0.35}(\text{guns})^{0.65}$. The function $(\log(x_1+1))x_2$, on the other hand, is a "textbook" utility function that is expressed by a relatively simple mathematical formula, even though there is no rule of behavior stated in everyday language that corresponds to this utility function.

Note that when a decision maker is a collective, and decision-makers in economics are in fact often families, groups or organizations, the assumption that preferences are definable makes even more sense since in this case a decision rule must be stated in words in order to be communicated among the individuals in the collective, both during the deliberation and implementation stages.

The aim of this research project is to demonstrate that the requirement that preferences be definable can be formally analyzed. In the following, I will analyze an example which illustrates the connection between a decision maker's language and the set of definable preferences. In Rubinstein (1998a), I presented an example in which a preference is assumed to be formed from a profile of more primitive binary relations. This result was closely related to the theory of Social Choice. These two examples may serve as the first step in a much more ambitious research program to study the interaction between economic agents with "language" as a constraint on behavior, institutions, communication, etc.

2. Searching for a boyfriend in a strange town

Consider the "tale" of a girl who arrives in a strange town and wishes to find a boyfriend. She obtains information on each candidate in the form of a list of his dates during the past T days. The number T is fixed. The girl has yet to meet either the boys or the girls in town (i.e., they are considered "unknowns"). All she can glean from any two given names on the list is whether they are identical or not. The model can also be interpreted in the context of an entrepreneur comparing candidates for a job, each of whom provides a list of his past employers about whom the entrepreneur has no information.

Following are several principles which the girl might use when comparing two candidates.

(1) Prefer boy A if he dated Alice longer than B did.

(2) Prefer the boy who has dated all the girls in town.

(3) Prefer a boy if all the girls he dated, from the second date on, had previously been dated by the other boy.

(4) If the two boys dated the same girl on the first date, prefer the one who dated her longer.

(5) Prefer the boy who made the most "switches".

(6) Prefer the boy who has dated more girls.

Prior to the formal definitions, let us discuss the definability of these principles. Obviously, these principles are definable in some sense since I have just used them. However, definability depends on language and here I am interested in the decision maker who is able to use only a limited language such that:

(i) The decision maker cannot refer to a boy by his name,

(ii) She can only refer to a girl by the term "the girl he dated at time t" and

(iii) The only comparison the decision maker can make between "the girl he dated at time t" and "the girl he (or someone else) dated at time s" is whether they are the same girl or not.

Principle (1) requires the decision maker to be able to refer explicitly to the name "Alice". Principle (2) can be stated in the decision maker's limited language using only the equality relation although it requires the use of a quantifier. Principle (3, 4, 5, 6) can be stated in this simple language. However principle (3) does not provide a definition of a preference since (for the case of $T = 2$, for example) it implies that the dating history (a,b) is preferred to (b,c), which is preferred to (c,a), which is preferred to (a,b). It is easy to see that principle (4) does not define a preference either. Only principles $(5,6)$ are definable in the limited language and induce preference relations.

The assumption that the decision maker uses a *limited language* to express her preferences can have various motivations. It may be interpreted naively as a reflection of the limited language that is available to her. However, even if the language available to her is rich, it might be the case that she cannot fully use it since the available relevant information she possesses is limited (recall that she is a newcomer to town). And even if she is able to use rich language and information is readily available, she may choose, due to complexity considerations, to express her preferences using a formula which is stated in a limited language.

3. The formal model

We can now move on to the formal analysis. The mathematical tools I will be using are from standard Mathematical Logic (Boolos and Jeffrey (1989) and Crossly (1990) are good introductory books).

We characterize the definable binary relations in the "pure language with equality". This is the language of the calculus of predicates that includes symbols for equality only. In this language, an *atomic formula* is of the type $z_1 = z_2$, where z_1 and z_2 are variable names (such as $x_7 = x_3$ or $x = x_2$). A *formula* is a string of symbols constructed according to the following inductive rules: All atomic formulae are formulae; if φ and ψ are formulae, then $(\neg\varphi$, $\varphi \wedge \psi$ and $\varphi \vee \psi$, $\varphi \rightarrow \psi$ and $\varphi \leftrightarrow \psi$ are also formulae; and if x is a free variable (does not appear with a quantifier) in the formula φ, then $\exists x\varphi(x)$ and $\forall x\varphi(x)$ are formulae as well.

A *model* in this simple language is simply a set G, interpreted here as a set of girls. The validity of a formula φ with the free variables $z_1,\dots,z_K$ in a model G (when we substitute each z_k with an element $a_k \in G$) is defined by the standard truth tables and is denoted by $G \models \varphi(a_1,\dots,a_K)$.

We are interested in preferences on dating profiles of length T that are defined by a formula φ with free variables $x_1,\dots,x_T,y_1,\dots,y_T$. Given a set G, the comparison between two boys who have the dating profiles $(a_1,\dots,a_T)$ and $(b_1,\dots,b_T)$ is defined by $G \models \varphi(a_1,\dots,a_T,b_1,\dots,b_T)$. Note that Principle (2) can be expressed by the formula

$$[\forall x \bigvee_{t=1,\dots,T}(x_t = x)] \wedge [\neg\forall x \bigvee_{t=1,\dots,T}(y_t = x)]$$

while principle (3) can be expressed by the formula

$$\bigwedge_{t=2,\dots,T} \bigvee_{m=1,\dots,t-1}(x_t = y_m).$$

In the rest of this section we are interested in formulae that induce binary relations that are transitive and asymmetric under all possible circumstances, i.e. in all models.

Definition: We say that a formula φ with $2T$ free variables induces a transitive and asymmetric relation ordering if the formulae
$\forall x_1,\dots,x_T,y_1,\dots,y_T,z_1,\dots,z_T[\varphi(x_1,\dots,x_T,y_1,\dots,y_T)\wedge \varphi(y_1,\dots,y_T,z_1,\dots,z_T) \rightarrow \varphi(x_1,\dots,x_T,z_1,\dots,z_T)]$
and $\forall x_1,\dots,x_T,y_1,\dots,y_T[\varphi(x_1,\dots,x_T,y_1,\dots,y_T) \rightarrow \neg\varphi(y_1,\dots,y_T,x_1,\dots,x_T)]$
are tautologies, i.e. they are satisfied in all models.

Note that the requirement that the formula be a tautology requires that the definition be applied to all possible worlds. The definition should induce a preference even when for each possible a list of girls in length T there is a boy with this dating experience. It should also be "consistent" regardless of the number of girls in town.

4. Analysis

Thus, for any set G, the comparison of any two dating profiles of a fixed length T (i) does not depend on any comparison of the girls dated by the two boys (such as $x_3 = y_7$), (ii) depends only on the comparison between the dating stability structures of the two boys for any given size of G and (iii) is constant for sufficiently large sets.

The analysis of the definability condition, consists of three stages:

First step: We can assume that the formula φ is written *without quantifiers.*

The "pure language with equality" has the property (see, for example, Robinson (1963)) that for every model G (that is, for any set), every formula has a logically equivalent formula containing the same set of free variables with no quantifiers. To illustrate this point, consider the formula

$$\varphi(x_1,\ldots,x_T,y_1,\ldots,y_T)=[\forall x \bigvee_{t=1,\ldots,T}(x_t=x)]\wedge[\neg\forall x\bigvee_{t=1,\ldots,T}(y_t=x)],$$

which states that a boy who has dated all the girls in town is preferred to one who has not. Although the quantifier "for all" is used here, the validity of $\varphi(a_1,\ldots,a_T,b_1,\ldots,b_T)$ in a model G, depends only on the number of elements in G and the number of elements in each of the two vectors. If G includes $L<T$ elements, then, when substituting $(a_1,\ldots,a_T)$ for $(x_1,\ldots,x_T)$, the validity of the formula $[\forall x\bigvee_{t=1,\ldots,T}(x_t=x)]$ is the same as that of a formula stating that in the vector $(x_1,\ldots,x_T)$ there are L different elements, which can be written without quantifiers. (Take the disjunction of conjunctions, each of which corresponds to a partition of $\{1,\ldots,T\}$ into L non-empty sets $\{e_1,\ldots,e_L\}$ and state that $x_i=x_j$ for any i and j that are in the same e_k and $\neg x_i=x_j$, if i and j are in distinct cells of the partition.)

The fact that for any model G, φ has an equivalent formula with no quantifiers implies that for any given model G, φ is equivalent to a disjunction of configurations of the variables $\{x_1,\ldots,x_T,y_1,\ldots,y_T\}$, where a *configuration* of $\{x_1,\ldots,x_T,y_1,\ldots,y_T\}$ is the "description of which variables are equal." (In other words, it is a conjunction of formulae where, for each pair of variables z_1 and z_2 in $\{x_1,\ldots,x_T,y_1,\ldots,y_T\}$, either $z_1=z_2$ or $\neg z_1=z_2$ appears in the conjunction with the following constraint: If $z_1=z_2$ and $z_2=z_3$ appear in the conjunction, $z_1=z_3$ is a conjunct as well.)

Second step: Characterization of the preference relations given a set G

For any given cardinality of G, the only preferences that are definable (in this simple language) are those that are induced by preferences on the "dating stability profile" of each candidate. The term "dating stability profile" refers to a partition of $\{1,\ldots,T\}$, in which i and j are in the same element of the partition if and only if the boy has dated the same girl at time i and time j. In other words, given a vector $(a_1,\ldots,a_T)$, define $E(a_1,\ldots,a_T)$ to be the partition of $\{1,\ldots,T\}$ in which i and j are in the same partition if and only if $a_i=a_j$.

For example, if a boy is fickle and dates a girl for only one period, the corresponding dating structure is the partition $\{\{1\},\ldots,\{T\}\}$; a "faithful" boy is characterized by the dating structure $\{\{1,2,\ldots,T\}\}$.

Claim: If φ defines an ordering in a model G, then for any two dating profiles with $E(a_1,\ldots,a_T) = E(b_1,..,b_T)$, it is not true that $G \models \varphi(a_1,\ldots,a_T,b_1,\ldots,b_T)$; in other words, the decision maker is indifferent between the two profiles $(a_1,\ldots,a_T)$ and $(b_1,\ldots,b_T)$.

Proof: We will show that for any dating structure, if two dating profiles have this structure, the defined preference cannot prefer one to the other. For notational simplicity, we consider the case in which the boy dates T distinct girls.

For any configuration $\psi(x_1,\ldots,x_T,y_1,\ldots,y_T)$ in the disjunctive normal form of φ, where $x_i \neq x_j$ and $y_i \neq y_j$ for all $i \neq j$, denote

$$N(\psi) = \{t \mid \text{there is an } s \text{ such that } x_t = y_s \text{ appears in } \psi\}.$$

Let ψ^* be such a configuration with the lowest number of elements in $N(\psi)$. It is impossible that $N(\psi^*) = 0$, since ψ^* would then be the conjunction of all formulae of the type $\neg z_1 = z_2$ for all $z_1 \neq z_2 \in \{x_1,\ldots,x_K,y_1,\ldots,y_K\}$ and the asymmetry tautology would not hold.

It follows that

$$\forall x_1,\ldots,x_T,y_1,\ldots,y_T[\psi^*(x_1,\ldots,x_T,y_1,\ldots,y_T) \wedge \psi^*(y_1,\ldots,y_T,z_1,\ldots,z_T) \rightarrow \varphi(x_1,\ldots,x_T,z_1,\ldots,z_T)]$$

is a tautology. Let ψ be a configuration which satisfies that for any t and r either $x_t = y_r$ or $\neg x_t = y_r$ appears within the configuration and $x_t = y_r$ appears in the conjunction if and only if there is an s such that both $x_t = y_s$ and $x_s = y_r$ appear in ψ^*. Then ψ must be a configuration in the disjunctive normal form of $\varphi(x_1,\ldots,x_T,y_1,\ldots,y_T)$. $N(\psi)$ is a subset of $N(\psi^*)$ and thus, by the minimality of $N(\psi^*)$ it must be the case that $N(\psi^*) = N(\psi)$. Denote by σ the permutation on the set $N(\psi^*)$, defined by $\sigma(i) = j$ if $x_i = y_j$ appears in ψ^*. There must be an integer n such that σ^n is the identical permutation. The formula φ is the disjunction of conjunctions and one of these must be the conjunction of $x_i = y_i$ for $i \in N(\psi^*)$ and the negation of all other equalities, contradicting the asymmetry tautology.

Third Step: For large enough sets "the same rule applies"

We are still left with the possible dependency of the defined preference on the cardinality of the set G: for any integer n there is a formula

$$\varphi_n(x_1,\ldots,x_T,y_1,\ldots,y_T)$$

with no quantifiers such that

$$G \models \forall x_1,\ldots,x_T,y_1,\ldots,y_T[\varphi_n(x_1,\ldots,x_T,y_1,\ldots,y_T) \leftrightarrow \varphi(y_1,\ldots,y_T,x_1,\ldots,x_T)]$$

for all models G with cardinality n. In other words, the decision maker's preference relation on dating profiles may depend on the number of elements in G. This brings me to the **third** and last point: By the compactness argument there is some n^* such that if G's cardinality is at least n^*, $\varphi(x_1,\ldots,x_T,y_1,\ldots,y_T)$ is equivalent to $\varphi_{n^*}(x_1,\ldots,x_T,y_1,\ldots,y_T)$.

Summarizing:

Proposition: If $\varphi(x_1,\ldots,x_T,y_1,\ldots,y_T)$ is a formula which induces a transitive and asymmetric binary relation, then for any n, there is an ordering $>_n$ on the set of "dating structures" of $\{1,\ldots,T\}$ such that if the size of G is n,
$(a_1,\ldots,a_T)$ relates to $(b_1,\ldots,b_T)$ iff $E(a_1,\ldots,a_T) >_n E(b_1,..,b_T)$ and there exists a number n^* such that all $>_n$ are identical for n greater than some n^*.

5. Conclusion

In this paper I have presented an example of an application of a definability constraint on the set of admissible *preference relations*. One could think of similar investigations of other objects such as the choice problems facing the decision maker, or the set of available strategies in a game. Imposing constraints on the class of situations, the individuals' preference relations and set of strategies may yield interesting results especially in interactive situations such as games. This would be a very challenging project. The aim of this paper, however, was only to persuade the reader that the subject can be discussed within formal models and that the definability assumption on preferences is not only natural but also analytically tractable.

References

Boolos, G.S., Jeffrey, R.C.: (1989), *Computability and Logic*, Cambridge University Press, Boston, NY and London.

Crossley, J.N., Stillwell, J.C., Brickhill, C.J., Williams, N.H., Ash, C.J.: (1990), *What Is Mathematical Logic?*, Dover Publications.

Robinson, A.: (1963), *Introduction to Model Theory and to the Meta Mathematics of Algebra*, North Holland, Dordrecht.

Rubinstein, A.: (1978), 'Definable Preference Relations – Three Examples', Center of Research in Mathematical Economics and Game Theory, R.M. 31, the Hebrew University, Jerusalem.

Rubinstein, A.: (1998a), 'Definable Preferences: An Example', *European Economic Review* **42**, pp. 553–560.

Rubinstein, A.: (1998b), *Modeling Bounded Rationality*, MIT, Cambridge.

Rubinstein, A.: (2000), *Economics and Language*, Cambridge University Press, Cambridge.

Shelly, M.Y., Bryan, G.L.: (1964), 'Judgments and the Language of Decisions', in Shelly, M.Y., Bryan, G.L. (eds.), *Human Judgment and Optimality*, John Wiley, New York.